The University of Hull

Developing and validating a scale to study mentors' behaviour in nursing education

being a thesis submitted for the degree of

Doctor of philosophy

in the University of Hull

by

Chen Yanhua

MA Sichuan University

July 2015

Abstract

Aim

To study the conceptualization of mentors' behaviour and develop and validate a scale based on that to assess students' expectation and measure mentors' actual performance.

Background

In the field of clinical nursing education in China, mentors are struggling with student mentoring as no national guidelines exist and proper training is unavailable and nursing students are suffering from a low quality of learning and negative experiences.

Design

A mixed methods exploratory sequential design.

Methods

At the development and validation stage eleven steps were taken; mixed methodology was used, including focus group and cross-sectional survey with a large sample from China (n=669); in data analysis both classical test theory (exploratory and confirmatory factor analysis) and item response theory (Mokken scale analysis) were conducted.

Results

Mentorship in nursing was conceptualized as a model with three correlated factors, i.e. professional development, facilitating learning and psychosocial support, which was guided by the theoretical framework generated through 46 studies and supported by exploratory factor analysis, confirmatory factor analysis and Mokken scale analysis.

Conclusion

This new scale based on this conceptulization is reliable, valid and scalable, which is supported by a wide range of psychometrics. It has shown good content validity according to review by nine mentor experts from the UK, stability over time, homogeneity in content, differentiability between extreme groups and hierarchical properties of mentors' behaviour in importance and reliability and moderate precision of ordering students' expectation. Therefore the mentorship in clinical nursing education is conceptualized as a three–correlated factor model and it is hierarchical in importance. This scale could find utility in nurinsg education in China.

Acknowledgement

This study has been completed in two and half years. I recall the days passed, the joy of searching and discovering knowledge, the anxiety of uncertainty, the hardship of overcoming all sorts of barriers, the ambition and responsibility of improving nursing education in China revived in me. I'd like to extend my hearted-gratitude to my supervisors' constructive guidance, my family support and nursing students' and mentors' cooperation in China.

Professor Roger Watson, my principal supervisor has opened a window for me to see a new world. His commitment to the supervision role moved me deeply, inspired me and challenged me to achieve excellence. His wisdom and knowledge helped me resolve the puzzles I met. His tirelessness in answering my endless questions improved my statistical skills and enlightened me. His professionalism demonstrated to me how to be an academia with integrity. His attitude towards nursing and me changed my viewpoint about nursing profession and myself and made me committed to enhance the image of nursing and become more confident as a human being and a nurse. His patience let me grow and his guidance directed me stepping out of pitfalls in the study and life. The time being with him comes to an end, but his influence and our friendship will last for ever.

My second supervisor, Dr Andrea Hilton has tried her best to help me in the two and half years. She gave me precious ideas and provoked me to think differently and more critically. I really appreciate her guidance and correction to my writing.

Throughout the time living and studying abroad, my son, Chengxi Li accompanied me, brought me happiness, encouragement and hope every day. Without his company I

could not bear the pressure of the study and the solitude of the life here; his tender hug, warm little hands, soft face, bright smile, sweet calling of mummy soothed me greatly. My husband, Bin Li, working back in China, his understanding, support, generosity and trust gives me the possibility and foundation to pursue my dream.

The last but not the least people made great contribution to this study are my dear students and mentors who participated in the focus groups and survey in China and the mentor experts in the UK. They provided precious data about mentorship and encouraged me to make changes to mentorship in China based on this study.

Table of contents

List of tables

List of figures

List of appendices

List of Abbreviations

Abbreviations	Name
AMAQ	Alleman Mentoring Activities Questionnaire
AMOS	Analysis of Moment Structures
ANOVA	Analysis of Variation
AISP	Automated item selection procedure
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CLES+ NT	Clinical learning environment and supervision scale + nursing teacher
CMIN/DF	Chi-square/ degree of freedom
CTCI	Clinical teachers characteristic instrument
CTT	Classical testing theory
CVR	Content validity rating
DMM	Double monotonicity model
EBP	Evidence-based-practice
EFA	Exploratory factor analysis
FL	Facilitates learning
GFI	Goodness of fit index
HEI	Higher education institutions
CVI	Index of content validity
ICC	Intra-class Correlation Coefficient

ICR	Internal consistency reliability
I-CVI	Item content validity index
IIO	Invariant item ordering
IRC	Item response curve
IRT	Item response theory
KMO	Kaiser-Meyer-Olkin
MCSS	Manchester Clinical Supervision Scale
MFS	Mentoring Functions Scale
MSA	Mokken Scale analysis
MHM	Monotone homogeneity model
NCTEI	The Nursing Clinical Teacher Effectiveness Inventory
NMC	Nursing and Midwifery Council
PAF	Principal axis factoring
PD	Professional development
PS	Psychosocial support
RMSEA	Root mean square error of approximation
S-CVI	Scale content validity index
SEM	Structural equation model
SRMR	Standardised root mean squared residual
TACT	Target, Action, Context and Time
TCM	Traditional Chinese Medicine
VIF	Variance Inflation Factor

Chapter 1. Introduction

This chapter outlines the research project. It includes background information on nursing education in China and mentorship research across disciplines, leading to the significance of this research project and defining its concepts for clear understanding and a foundation of the study.

1.1 Background

1.1.1 Nursing in China

Historically, there was only Traditional Chinese Medicine (TCM) in China before western medicine arrived at the beginning of the 19th century; also there was no nursing profession and servants or family members provided general care to patients. Later western hospitals were opened in China and foreign nurses came. The first formal nurse training program was established in Guangdong province in 1888. A degree nursing program was opened in 1920 in Peking Union Medical College with funding from the Rockefeller Group (Wong and Zhao, 2012; Gao et al., 2012), but this program was suspended in 1952 and only secondary nursing programs were provided for quickly preparing nurses to meet urgent public health needs and requirements. Unfortunately, in the ten years of the Cultural Revolution (1967-1977), all programs were closed (Gao et al., 2012; Wong and Zhao, 2012), until 1980 higher nursing education programs were reopened and the nursing profession has developed since then.

China now has over two million nurses but the ratio of nurses to 1000 people is still low (2.05:1000) and the ratio of nurses to beds is approximately 0.4: 1 (Guo, 2011); usually one nurse takes care of 6-8 patients (You et al., 2013), even as many as 30 (Eddins et al.,

2011). All this suggests that the nurse shortage is severe in China when compared to the UK, USA and other menbers of Organization for Economic Co-operation and Development (You et al., 2013).

Nurses there mainly work in hospitals, as community nursing and home nursing are immature. In hospitals where doctors dominate the system, nurses lack autonomy and essentially follow the prescriptions and orders of doctors, working passively in a subservient role, or performing as doctors' hands and legs (Eddins et al., 2011). Therefore nurses have low social status, poor professional image, poor working environment, low pay and high burnout (You et al., 2013). In addition, the tension between health providers and patients, caused mainly by low government investment into public health, makes nursing an unattractive profession in China.

One internal issue in nursing is that nurse career development is neither clear nor inspiring compared to that of other developed countries. Usually nurses do the same job from graduation to retirement, and graduates from any program do the same work, therefore only a few students choose nursing as their major. Most of them were assigned to a nursing faculty, particularly degree students under the contemporary higher education system in China (Eddins et al., 2011). Nowadays clinical nursing specialist programs are growing fast, which gives nurses more of a career plan and aspiration.

1.1.2 Nursing education in China

Nursing was originally a discipline under clinical medical science, so the curricula were influenced by medicine, manifesting a disease-centred curriculum, for example, in its division into surgical, medical, maternal and paediatric nursing subjects and in each subjects the contents were organized according to diseases in each human body system

(Eddins et al., 2011) and doctors acted as nursing teachers in nursing schools (Wong and Zhao, 2012). However, after more than five years of effort by nursing academia, from 2011, nursing has become a top order discipline like clinical medicine, which is beneficial for the independent development of nursing science.

There is a full spectrum of nursing education programs from diploma to doctoral level now; diploma programs still dominate this field, and undergraduate degree programs are growing, currently sharing 8% of total programs (Gao et al., 2012). There are 216 undergraduate programs in 2012 (You et al., 2015), 65 master programs (Li and Han, 2010) and 10 PhD programs (Wong and Zhao, 2012; Gao et al, 2012) which have prepared qualified nursing teachers, so the 'doctor nursing teachers' have been replaced gradually by real nursing teachers, but at PhD level, doctors are still supervising nursing students because of the shortage of qualified nursing supervisors, which may cause the low professional morale and high turnover rate.

To study to be a nurse there are several options in China, such as degree, associate degree, 3-year diploma and 5-year diploma program, in addition all the programs are generic. To be enrolled as degree and associate degree students, you have to finish 12 years of general education and pass the National University Entrance Examination. Students apply to or are selected by a university based on their scores in the exam (Eddins et al., 2011). With regard to diploma programs, before going to secondary nursing school, which may be independent or affiliated to a hospital, students need to finish nine years general education and participate in secondary examination, which is organised by provincial educational institutes. For 3-years diploma programs, academic requirement is lower compared with 5-year programs; 5-year programs can issue a certificate to a graduate, which is similar to an associate degree, and the number of

programs is limited (Eddins et al., 2011). Between diploma, associate degree and degree program there are link programs (such as self-learning and provincial level test program and distant learning program) which allow students to apply higher levels study after finishing their basic study.

There is national curriculum guidance for each program (Eddins et al., 2011). Degree programs last 4 years and associate degree programs have three years study, and the other two diploma programs have three and five years of study, respectively. All the curricula covers basic life sciences, such as anatomy, physiology, chemistry, bacterial and biology, nursing modules such as internal medical nursing, surgical nursing, obstetric and genecology nursing, and modules about human science, for example ethics. The differences between programs are that the depth and scope vary in each course; for degree program it targets to equip nursing students with comprehensive clinical nursing, teaching and research ability, therefore the academic width and depth requirement for them is the highest and extra modules as nursing education and research are provided (Wong and Zhao, 2012). Diploma program students are trained to do more practical work in nursing school, so theoretical requirement is not stressed as much as in degree programs (Eddins et al., 2011). To be a nurse, students from all the programs need to take provincial level registration exam.

When graduates are employed mainly by hospitals, they are not used as they are designed in nursing school (Eddins et al., 2011). No matter which program a graduate comes from, he/she does the same job as they all have been trained as general nurses, also due to poor clinical nursing leadership and management and lack of career development. However the promotion rate and pace are different, which means you can be paid different salaries. Normally degree graduates need less clinical experience to be promoted to a higher level than diplomates. Degree students have a much higher chance

of working in large hospitals in cities which offer higher salaries, more resources and opportunities such as more continuing professional development (Eddins et al., 2011). On the contrary, diploma students are more likely to be employed by small hospitals in rural areas or private hospitals or clinics, which offer less salary and unstable working opportunities and fewer resources (Eddins et al., 2011; Wei and Yang, 2015).

To alleviate the nurse shortage, expanding the quantity of nursing training programs is the main strategy that the Chinese nursing regulatory and statutory body has taken. After over 30 years of progress in nursing education, the number of nurses has increased to 2.05 million at the end of 2010 and will increase to 2.86 million by 2015 and 4.45 million by 2020; the number of nurses per 1000 people rose from 1.52 in 2010 to 2.05 in 2013 (You et al., 2015)

1.1.3 Clinical nursing education in China

With such rapid development and limited numbers of nursing teachers, obviously great pressure has been put on the quality of teaching. First and foremost, clinical learning should never be ignored; it is the ideal place for knowledge and practice integration, skill training, competency and confidence building and professional socialisation. But clinical nursing education in China is never far from problems.

Staff nurses act as teachers on wards without extra pay and they also lack adequate training to perform properly as educators to support and assess nursing students in the vast majority of situations (Eddins et al., 2011). Due to nurse shortage in hospitals, the clinical teachers simultaneously act as staff nurses and face a dilemma between providing care to the patients and training the students, which is also identified in other countries (Myrick and Barrett, 1994; Edmond, 2001; Myall et al., 2008). Undoubtedly,

under this condition, nurses performing the role of clinical teaching are neither confident nor competent to fulfil their roles, which brings little benefit to nursing student's clinical learning (Eddins et al., 2011).

Nearly all nursing students go to hospital in the final year for about consecutive 10-12 months of clinical practice, which is different from the parallel arrangement (in each term, there is theoretical learning and clinical placement learning) in other countries and this is said to be hospital and school-centred, rather than student-centred. When students are on wards, some have an assigned nurse to teach them, some have not; under this situation, students just follow shifts; but on each shift they may have different nurses to work with. There is no national level or provincial level benchmark or standard to guide both students' and mentors' behaviour in clinical learning and teaching, as there is in the UK, USA or Canada.

When nursing students study in clinical placement, there are no strict rules to guarantee their supernumerary status, but in nursing human resource management they are not accounted for as staff. In reality, due to severe nurse shortages in hospitals, students are prone to be treated as a human resource rather than learners, and most often, the placement learning is work-led rather than education-driven， particularly a large proportion of basic nursing procedures are done by them (Eddins et al., 2011). In turn, students' professional identity acquisition, interests in nursing, professional competency development are impaired; their enthusiasm for being a nurse is undermined. Actions must be taken to change the situation: for instance, students' supernumerary status must be maintained and proper mentorship should be provided to meet students' clinical learning needs.

Being a nurse and clinical nursing teacher over 15 years, I have seen the struggle of nursing students and nurses acting as teachers on wards. This situation should be investigated, and the teaching quality, patient's safety and students' well-being should be considered integratedly. After visiting the UK in 2010, I was deeply impressed by mentorship there, particularly the ideas of supporting nursing students for their personal and professional development, which prompted me to explore this more thoroughly to help establish a proper mentorship system, mentor training program and objective assessment of mentorship in China. This study particularly intends to understand what Chinese nursing students want from their clinical mentors, what proper behaviours a mentor should have, and to find a reliable and valid measurement to assess these objectively. It is significant to guide mentors' behaviour, improve the clinical nursing education quality, uphold professional reputation and patients' safety. If a Chinese-context-based, psychometric sound measurement tool of mentorship is located or developed, this tool can be used to assess students' expectation from mentors and inform the match between students and mentors which will benefit mentors and students mutually; this tool can be a guideline to shape mentors' training program; it can be served to guide mentors' behaviour and shape their self-development; it can also be applied by higher education institutes to assess mentors' real performance to inform selection of good mentors and identify mentors' training needs; if this tool frequently applied in clinical nursing education it can make a better mentoring culture and boost a better clinical learning environment, and enhance teaching quality and learning experience in China

1.2 Mentorship

1.2.1 Introduction

Mentor is a concept originating from Greek mythology with more than 3000 years history. It is generally described as a one-to-one relationship between a more experienced and senior person (mentor) and a younger or less experienced person (mentee) in work and study environment, and mentors nurture, guide, coach and assist mentees to realise their personal and professional development. The concept of mentor has more than 50 definitions (Crisp and Cruz, 2009) for researchers in different fields have defined it from their own viewpoint, which echoes Jacobi's observation (1991 p. 506) that 'although many researchers have attempted to provide concise definitions of mentorship or mentor, definitional diversity continues to characterize the literature.'

In America, mentorship flourished after the work of Levenson et al (1978) in business and organization. It has been used as a tool to nurture new leaders, new staff, to raise morale and reduce turn over. It is also 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. Further more mentorship is extensively employed in education for teaching college students and reducing the drop-out rate; in doctoral student education, enhancing research productivity; and nurturing new teaching staff and leaders. Now it has spread all over the world, in all sorts of areas, such as psychology, medical and the nursing field.

Research before 2000, focused mainly on the beneficial part of mentoring, i.e. psychosocial support and career development for the mentee (Allen et al., 2004; Noe, 1988), later benefits for mentors and organization were also studied. Recently the

negative side of mentoring such as neglect, mismatch, bullying, and jealousy has been studied (Eby et al,2008; Ensher & Murphy, 2011). Quantitative research with cross-sectional design has dominated the literature (Allen et al., 2004; Allen, 2008).

Later in this chapter, theoretical underpinning of mentoring, mentoring type, function, measurement, and mentorship in nursing will be discussed.

1.2.2 Theoretical underpinnings of mentoring

The theoretical underpinnings to mentoring mainly mentioned are social learning theory and social exchange theory. Social learning theory shows that people can learn from observing other people's behaviour and the outcomes. In addition, learners can acquire knowledge and skill without 'trial and error' (Bandura and McClelland, 1977). When a mentee perceives that a mentor is a good role model with accomplishment and success, he/she would be more willing to learn and mimic the activity and behaviour of the mentor. Social exchange theory claims that if people feel benefits and rewards from a relationship, they continue to associate with that person and invest into that relationship (Cropanzano and Mitchell, 2005). In the education field, adult learning theory and experience learning theory are also related to the mentoring process (Gopee, 2011).

1.2.3 Types of mentoring

There are two main types of mentorship: formal and informal mentorship. Now more and more organizations are employing formal mentorship to improve growth and development of junior members in organizations and to enhance staff retention and job satisfaction (Allen et al., 2006). Formal mentorship means the mentor and mentee relationship is formally assigned or chosen in an organization and both mentors and

mentees will be trained for proper behaviour in maintaining this relationship. Informal mentorship is just formed by two people's mutual liking, identification and attraction (Kram and Isabella, 1985). Studies show that informal mentorship is more effective than formal, for mismatch of two people or 'forced relationship' in formal mentorship will cause conflicts, mutual lack of recognition, no participation and contribution to this relationship (Allen et al., 2006). Now formal mentoring programs imitate the nature and process of informal ones and have become more effective.

With the development of mentorship, new types of mentorship have emerged, for instance, peer mentoring, e-mentoring or distance mentoring, group mentoring. Peer mentoring refers to mentors and mentees having the same level in an organization whereby they mentor each other (Bryant et al, 2015). For example, students in higher year groups mentoring those in a younger age group; patients with more experience in a certain disease mentoring those with less experience. E-mentoring or distance mentoring means that when mentors and mentees live in places geographically far away, mentoring can take place using email, teleconference or telephone instead of face-to-face contact (Qing et al, 2010). Team or group mentoring implies that a mentor mentors a group of mentees over a period of time. The mentoring relationship (model) has developed from a top-down/classical/hierarchy of mentee benefit to a reciprocal/peer-mentoring relationship with mentor and mentee sharing mutual benefit and contribution (Jones and Brown, 2011).

1.2.4 Roles/functions of mentors

Kram (1983) identified mentors' functions as follows: career development and psychosocial support, including nine roles, as shown in Table 1.1. This mentoring

theoretical framework is the most cited by researchers. Later, the two-function model is separated into three functions: career development; psychosocial support; and role modelling (Scandura, 1992). This three-function model is also widely accepted (Weng et al., 2010; Allen et al., 2006).

Table 1.1 Kram's mentoring model

Career Function	**Psychosocial Function**
Sponsorship	Role modelling
Exposure-and-visibility	Acceptance-and-confirmation
Coaching	Counselling
Challenging assignments	Protection
	Friendship

1.2.5 Measurement of mentoring effectiveness

Many empirical studies in business and industries show that mentoring benefits mentees, mentors and organizations. For mentees, salary level and promotion rate, are used as objective, tangible variables to measure the career development function of mentorship, while the subjective variables such as job satisfaction, job commitment, and organization commitment are applied (Dreher and Ash, 1990; Kram and Isabella, 1985; Kram, 1983; Allen and Eby, 2003). As to mentors, more generality, sharing of knowledge and skill are considered internal benefits and rewards, in addition, professional development is taken as an external benefit (Dreher and Ash, 1990; Kram and Isabella, 1985; Kram, 1983; Allen and Eby, 2003). With regard to organizations, high morale, productivity and low staff turnover are identified as the benefits of mentoring (Allen and Eby, 2003).

Mentoring as the independent variable causing the dependant variables (salary, promotion rate, job satisfaction outcomes) to change has been stated as above. As a mediate variable, the quality of mentoring behaviour and/or the mentoring relationship itself is fundamental and has been carefully studied (Allen and Eby, 2003; Allen et al., 2006; Dreher and Ash, 1990).This enables a deeper understanding of mentorship, such as what function of mentorship is missing or weak and how to improve it, which is an advance over just showing the difference between the mentored and non-mentored group.

1.2.6 Mentorship in nursing

Mentorship has been adopted in many nursing fields, such as clinical teaching of nursing students; newly qualified nurse transition; staff nurse support and development; nursing leader development in a clinical setting and in academic nursing; masters and PhD students teaching; new teacher orientation and support in teaching and research 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. At early stage academic nursing attempted to define concepts of mentor and mentoring and to clarify the roles and functions of mentors without reaching consensus (Myall et al., 2008). Later, researchers focus 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).

1.2.7 Mentorship in clinical nursing education

In North America

In the USA and Canada, the clinical teaching model went from apprenticeship in the early stage to a more academic model. As nursing education moved into universities or colleges in the 1970s, the teachers in nursing schools took the role of teaching students in the placement with a teacher-student ratio from 1:10 to 1:20 (Myrick, 1988, Udlis, 2008). This 'nursing teachers teaching in a clinical model', was challenged by the budget in higher education institutes and the competence outcome of nursing graduates, so the lecturer teaching model shifted to a model of clinical staff training student nurses, which is called 'preceptorship' by North Americans (Myrick, 1988, Udlis, 2008).

In the UK

In the UK, nursing education experienced a similar transition in clinical teaching with different motivation. Because of Project 2000, the locus of nursing teaching was transferred to universities for preparing knowledgeable nurses. However this university didactic teaching model showed that the newly qualified nurses were not fit for practice (Andrews and Wallis, 1999), so *Fitness for Practice* was launched and 'mentors' (nurses) were assigned to each student in clinical learning time to safeguard the fitness for practice since then. Sweden, Norway and Finland, have this model as well; some call it 'supervision' (Saarikoski and Leino-Kilpi, 2002, Saarikoski et al., 2007, Landmark et al., 2003).

In China

In China, nurses mainly act as teachers to teach and train students in the final year of clinical learning; some students are paired to nurses on a one-to-one base, some are not. The one-to- one relationship between clinical staff and nursing students was named the 'one-to-one clinical teaching model' (一对一临床带教模式). As the nursing world

enters into the mentoring phase, the use of mentorship started in the 1990s, then Chinese nursing educators embraced the concept of mentorship from the end of the 1990s, and named it the 'Dao Shi Model' (导师制) (Wang et al., 2011; Liu et al., 2012; Ma et al., 2007). The nurses mentoring students are called 'Dao Shi'(导师); 'Dao' (导) conveys the meaning of guiding, while 'Shi' (师) means teacher; 'Dao Shi' (导师) contains the meaning that a teacher should impart knowledge and skill, guide students' careers and enlighten them in the puzzles of life, work and study. The Dao Shi Model has emerged in clinical teaching, school teaching and new staff transition in the field of nursing.

All of these models aim at nursing students' or new staff's personal and professional support and development, with an intensive, individualised, daily based and continuous relationship in the real nursing world.

1.2.8 Mentors' roles and functions in nursing

As the concept of mentor (mentorship) is used inconsistently in different countries to describe differing variations in relationships between experienced nurses and nursing students/new graduates/staff nurses, so the roles and functions should not be the same. Even in the same field, with the evolution of the profession, the roles and functions vary correspondingly.

In North America

In North America, some mentors' roles often cited as effective are based on Darling's three-function model of mentorship (Darling, 1984): inspirer, investor, and supporter function, including 14 roles: 'model, envisioner, energiser, investor, supporter，

challenger, standard-prodder, teacher, feedback-giver, eye-opener, door-opener，idea-bouncer, problem-solver, and career counsellor'.

In the UK

In the UK, *the Standards to Support Learning and Assessment in Practice* of Nursing and Midwifery Council (2008) defined the eight main roles that mentors must have to support nursing students:

- Establishing effective working relationships,
- Facilitation of learning,
- Assessment and accountability,
- Evaluation of learning,
- Create an environment for learning,
- Context of practice,
- Evidence-based practice,
- Leadership.

1.2.9 Measurement of mentorship in nursing

Effectiveness of mentorship in nursing

Mentorship has increased in popularity over recent years and its effectiveness has also been investigated. Some mentoring programs have shown that mentorship has increased retention of new nurses, job satisfaction, familiarizition with the working environment

and leadership skills (Fox, 2010; Komaratat and Oumtanee, 2009; Hamilton et al., 1989). Mentorship can provide better adapt ability (Ridley et al., 1995), role conception (Dobbs, 1988), nursing performance and role socialization (Jones, 2000; Scales et al., 1993) to nursing students than teaching by traditional school nurse teachers on a ward. It is also beneficial to mentors as they may feel more job satisfaction and self-esteem by sharing knowledge with young nurses and nursing students and they have the opportunity to learn from mentees (Dibert and Goldenberg, 1995; Usher et al., 1999; Hyrkäs and Shoemaker, 2007). Mentoring programs can be effective for some particular purposes such as improvement of evidence based practice, research knowledge and skills (Byrne and Keefe, 2002; Wallen et al., 2010).

Influential factors

Just being with a staff nurse does not guarantee that mentoring and learning take place, some 'toxic mentor' may even block study (Darling, 1986). Gray and Smith (2000) also identified the characteristics of bad mentors, such as disliking jobs, over-protecting students, lack of knowledge, intimidating students and being less friendly. In the one-to-one relationship, learning activity and efficacy is mainly influenced by the relationship (Andrews and Wallis, 1999). Saarikoski et al (2007) identified the mentoring relationship as the most important factor in the clinical learning environment, and this is further confirmed by other researchers in Norway, Sweden and the Netherlands (Johansson et al., 2010; Henriksen et al., 2012; De Witte et al., 2011). Therefore, the effectiveness of mentorship needs assessment (Sawatzky and Enns, 2009).

Common barriers preventing mentors from providing proper mentoring include lack of training and preparation; lack of time; lack of support from ward managers, peer nurses and teaching staff from nursing school; and role ambiguity and conflict (Dibert and Goldenberg, 1995; Usher et al., 1999; Hyrkäs and Shoemaker, 2007). Training programs can usually improve mentors' attitude toward students, teaching ability and knowledge and assessment competency (Fox, 2010; Komaratat and Oumtanee, 2009).

Measurement tools

Intending to develop a reliable tool to assess the mentorship in the nursing academic and research field, Berk et al conducted a comprehensive literature review and did not find proper tools to measure mentorship quality in the nursing field (Berk et al, 2005). Not surprisingly, nurse educators and clinical researchers used tools from other fields such as organisation and education (Altuntas, 2012; Weng et al., 2010; Aponte; 2007, Gwyn, 2011; Richard, 1995; Jones, 1997) or self-designed primitive scales to measure mentorship. As about 10 years passed, the situation may have changed, therefore a systematic search needs to be carried out to identify tools, appraising their qualities objectively.

1.2.10 Summary of mentorship

Mentorship is universally used in wide spread professions and disciplines, and it is beneficial to mentees, mentors and organizations in business and organization fields. It was imported into the nursing field and shows effectiveness and benefits to new nurses, new nursing faculties, nursing students as mentees and senior nurses and nursing teachers as mentors. Clinical nursing education in China needs substantial and instant improvement in better understanding about students' expectation from mentors and

better preparation of mentors for their roles. It is necessary and valuable to study mentors' behaviours and functions and make sure good mentoring take place to benefit mentors and mentees in nursing in China. To do this, the first issue is to define mentorship properly and make objective measures of mentorship, so the next section deals with concepts definition and differentiation. Searching and critique of assessment tools of mentorship will be presented in Chapter 2.

1.3 Definition of terms

1.3.1 Introduction

As said above, different countries use different terms to refer to the same relationship. Hence a clear definition is imperative for further study**.** There are many supporting roles in placement for nursing students learning, like 'preceptor', 'clinical supervisor', 'mentor' , 'facilitator', 'personal tutor', 'clinical educator' , 'link teacher'. Among them, the three terms 'mentor' (mentorship, mentoring, mentored), 'preceptor' (preceptorship, preceptoring, preceptered), and 'supervisor' (supervision, supervisory, supervising, supervised) are found often to be used interchangeably (Jokelainen et al., 2011; Andrews and Wallis, 1999; Cooper and Palmer, 2000), which is complicated and confusing. Therefore their definitions and conceptualizations are examined and clarified first.

1.3.2 A historical point

To come to a thorough understanding of the three terms: 'mentor', 'supervisor', and 'preceptor', a historical method may give some insights. In April, 2013, the following search terms: mentor*, precept* supervis* and nurs*, were used in PubMed, which has

a longest history of collecting data from medicine, nursing and allied health subjects, without date and language limiters, to find the earliest papers with these terms and the trend of their use in the nursing field. The results are shown in Tables 1.2 and 1.3.

Table 1. 2 The number of papers with the term mentor*, precept*and supervis*

Terms	Before1979	1980-89	1990-99	2000-09	2010-04/2013
Mentor* and nurs*	8	118	741	1811	894
Precept* and nurs*	71	265	530	1036	468
Supervis* and nurs*	2074	1881	4049	3254	1445

Table 1.2 shows that ‘preceptor’ and ‘supervisor’ exist before the term ‘mentor’ in the nursing field. ‘Mentor’ is gaining momentum with larger increments in each decade than ‘preceptor’. ‘Supervision’ has dramatically larger numbers of articles in each decade than the other two terms, which implies that it is an overarching term, with broader application than ‘mentor’ and ‘preceptor’ in the field of management for securing nursing care quality, in psychological/mental health nursing for interpersonal skill development; in clinical support for staff nurses in professional development; and, in clinical teaching for nursing students support, on a one-to-one, day-to-day basis or session basis (Yegdich and Cushing, 1998). Particularly in the last decade of the 20th century, the number of papers using supervis* surged by more than 2000. One probable explanation is that nursing staff clinical supervision gained momentum after the study (Butterworth et al., 1998; Butterworth and Faugier, 1992) and it was formalised by the Department of Health (1993) in the UK and the United Kingdom Central Council for

Nursing, Midwifery and Health Visiting (UKCC, 1996) announced that every nurse should get access to clinical supervision.

Some papers using the three terms were selected and are displayed in Table 1.3 to show the meaning and use of the terms in early days. The contents were extracted from the abstracts or full texts, therefore some articles, even with a longer history, whose abstracts or full texts were inaccessible are not included here.

The earliest paper pertinent to nursing using 'mentorship' was found in 1960s, which had little relationship to nursing as implied by the title, 'parents and mentor' (no author listed), later article located in 1978, and 1979 were about physicians who acted as mentors of nurse practitioners and the mentorship impact on cost cutting and boosting of patient care quality in hospitals in USA ((Hohman, 1979; Spencer and Winkels, 1978), but these findings disagree with some nursing research articles which claim that mentorship began from early 1980s in nursing (Andrew & Wallis, 1999; Jokelainen et al, 2011).

The first paper using the term 'preceptor' was published in the USA in 1964 with the title 'the nursing instructor as model and preceptor' (Stein, 1964), but no abstract was found. The three papers listed in Table 1.3 were related to using preceptorship to prepare nurse practitioners, which is consistent with the depiction of Myrick (1988) who stated that preceptorship surfaced after the nursing practitioner programs in the 1960s. Later it was used to deal with nursing students' reality shock and reduced education budget in nursing.

This study aims to find concepts related to nursing education and training rather than to management, so papers using the term 'supervis*' pertinent to nursing management

were not selected. The earliest supervising paper can be traced back as early as 1937, relating to supervisors' role in clinical nursing professional instruction and support to staff nurses in public health in the USA shown from the full article (Earp, 1937) and the other two papers confirmed the role of public health nurse supervisor (Heisler, 1940; Walsh and McNeil, 1951). This role may differ from that of supervisors in psychotherapy and in nursing later. Yegdich and Cushing (1998) stated that clinical supervision originated from psychoanalysis, in which a senior psychotherapist acted as the supervisor supervising a young one or a student to promote self-reflection and therapy development through case review. Gradually it was imported into the field of mental health nursing, which has the highest level of clinical supervision engagement among all clinical nursing specialties (Bishop, 1998; Sloan and Grant, 2012) now. The earliest paper about mental health nursing supervision entitled 'Mental health nursing, skills in supervision' was published in 1987. And the research in Europe (Lyth, 2000) reflected the influence of psychotherapy on clinical supervision and its flourishing in the 1990s.

Table 1.3 Example papers published early in PubMed with the three terms

Term used	Reference	Content	Country
Mentor*	Hohman, 1979	Mentorship was used in a hospital to solve shortage of nurses, rising costs and the lack of staff developing programs.	USA
	Spencer and Winkels, 1978	Physician mentors supervised nurse endoscopists for one year, demonstrating efficiency of nurse endoscopists.	USA
	Vance, 1982	Cultivating mentoring relationship among nurses was advocated to establish a supportive environment.	USA
Precept*	Linn, 1975	Preceptorship program was used to train family nurse practitioners.	USA
	Davidson et al., 1975	Six months of preceptorship with a paediatrician was conducted to prepare paediatric nurse practitioners.	USA
	Geolot et al., 1977	Physicians acted as preceptors to train emergency nursing practitioners.	USA
Supervis*	Earp, 1937	The public health nurse supervisor role was introduced and differentiated with administrative role in public health area.	USA
	Heister, 1940	A training program of public health nurse supervisor was introduced.	USA
	Walsh and McNeil,1951	An advanced training (supervision) program for nursing students who wanted to be a public health nurse or supervisor was reported.	USA

Summary of historical review of concepts analysis

This historical study shed light on the origin of the three terms and their implications. 'Mentorship' is the youngest term in the nursing field. It was directly imported from business and organization (Hohman, 1979) aiming at staff development and organization benefit of nurse retention. 'Preceptorship' originated from health science field (the title of the oldest paper about preceptorship was found several decades earlier in medicine than in nursing), used to help the adaptation and development of nursing students and graduates (Ferguson and Hauf, 1973). 'Supervision' has the longest history and broadest meaning and the oldest papers (Earp, 1937; Heisler, 1940). It reflects the managerial implication and educational function of supervisors to facilitate public health nurses to adapt to their roles and to provide psychosocial support. In addition, it was influenced by psychoanalysis and has flourished since the 1990s. All of the roles were taken by more experienced and knowledgeable nurses or doctors, intending to support new or young nurses. 'Supervision' and 'mentorship' are more staff-based, while 'preceptorship' more likely implies teaching students. These similarities may cause the confusion in use today, but the subtle difference gives a clue that further clarification is needed.

1.3.3 Definitions of the three terms in different countries

The conceptualization and use of the three terms differ geographically and the different concepts in different countries are explored and presented in Table 1.4, mainly in the USA, Canada and the UK. Particularly, clinical supervision was conceptualised as a relationship of nursing staff to nursing staff both in North America and the UK, whereas some European countries such as Sweden described it as a relationship between nurses

to students in clinical learning. Hence the Swedish definition is included to present a whole picture of the 'supervision' concept.

Table 1.4 Definitions of the concepts: mentoring, preceptoring, and clinical supervision

Mentoring	**USA** Mentoring is a reciprocal and collaborative learning relationship between two individuals with mutual goals and shared accountability for the success of the relationship. The mentor is the guide, expert and role model who helps develop a new or less experienced mentee (Jacubik, 2012 p.9). **Canada** Mentoring involves a voluntary, mutually beneficial and usually long-term professional relationship. In this relationship, one person is an experienced and knowledgeable leader (mentor) who supports the maturation of a less –experienced person with leadership potential (mentee) (Canadian Nursing Association, 2004 p.18). **UK** A mentor is a registrant who has met the outcomes of stage 2 and who facilitates learning, and supervises and assesses students in a practice setting (Nursing and Midewifery Council, 2008 p.57).
Preceptorship	**USA** A one-to-one relationship between a staff RN and a nursing student during an intense, time-limited clinical experience, with the support of nursing faculty to facilitate student learning and provide evaluation of course objectives (Udlis, 2008 p.20). **Canada** A formal, one-to-one relationship of pre-determined length, between an experienced nurse (preceptor) and a novice (preceptee) designed to assist the novice in successfully adjusting to and performing a

	new role (Canadian Nursing Association, 2004 p.13). **UK** A period of structured transition for the newly registered practitioner during which he or she will be supported by a preceptor, to develop their confidence as an autonomous professional, refine skills, values and behaviours and to continue on their journey of life-long learning (Great Braitain. Department of Health, 2010 p.11).
Supervision	**North America** Clinical supervision is an opportunity for a more experienced nurse to monitor, educate, and support a less experienced nurse in how they perform clinical skills (Cutcliffe & Lowe, 2005 p.486). **UK** Clinical supervision is a support mechanism for practising professionals within which they can share clinical, organization, developmental and emotional experiences with another professional in a secure confidential environment in order to enhance knowledge and skills. This process will lead to an increased awareness of concepts including accountability and reflective practice (Lyth, 2000 p.729). **Sweden** The concept of supervision is used as an overarching term. The term supervisor refers to a person who guides, supports and assesses the student and is responsible for the intended learning outcomes within clinical education (Johansson, 2010 p.2087).

1.3.4 Differentiation of the three terms ‘mentorship’, ‘preceptorship’ and ‘clinical supervision’

Differentiation of the three terms across countries

Mentorship

In the USA and Canada, the meaning of mentors and the role it signifies have very traditional notions, which highlight a duration of two to four years, or an indefinite period of time (Flynn, 1997 p.14) and the informal relationship between mentors and mentees. The intimacy of the relationship is also one of the facets highlighted when people try to distinguish the concept of mentor and preceptor. So mentorship is not used frequently in the pre-registration education compared with preceptorship in the USA and Canada.

But recently this has changed somewhat: more mentorship has been formally established for new teachers in nursing schools and clinical staff nurses’ support and professional development, particularly after the Magnet Recognition Program (Jakubik, 2008; 2012) for the research findings supported the assertion that ‘mentoring relationships have benefits for staff nurses regardless of the relationships, length and regardless of whether they are formal or informal’ (Jakubik, 2008 p.279).

For Canadian nursing, mentorship can be used to describe the short and long term relationship between the nursing student and the registered nurse (Wilson et al., 2010; Sawatzky and Enns, 2009). Therefore, mentorship is used in clinical teaching for nursing students in the USA, and Canada now (Andrusyszyn et al., 2007).

‘Mentor’ entered the UK from North America and mentorship is firmly established to mean the relationship between students and clinical nurses. However it is not clear why

UK nursing educators use ‘mentorship’ instead of ‘preceptorship’ to describe nurse to nursing student clinical teaching other than to describe staff nurse supporting staff nurse. One explanation may be that preceptorship was mainly used with senior nursing students in North America, while the UK has a clinical mentoring system throughout three years learning, which is longer than that of North America (Gray and Smith, 2000).

Preceptorship

In the USA, Canada, and Australia, the term ‘preceptorship’ implies a short period of time when a clinical staff nurse, as teacher and coach, helps nursing students to acquire skills and knowledge (Mills et al., 2005). The preceptorship programs are also used to help the smooth transition of newly qualified nurse in the UK. Both the relationships of teaching nursing students and helping the new graduates last for several weeks to several months, which are obviously shorter than that of mentorship in North America. This relationship is normally assigned by the management personnel and academic staff.

Clinical supervision

Clinical supervision in this research is not concerned with management, which is different from ‘supervision’ as used in clinical governance. Clinical supervision is mainly used to develop skills in psychiatric/mental health nurses, and has a long history in USA, Australia, New Zealand and the UK (Bishop, 1998; Macdonald, 2002). It is also used in many other nursing fields such as community nursing and general wards. Clinical supervision in the UK has more emphasis on the holistic help, support and developing function, while in the North America, supervisors are considered to be experts to monitor, educate, and support less skilled staff in clinical skills (Cutcliffe and

Lowe, 2005). It also been used to describe the nurse to student clinical teaching relationship in Norway, Sweden and Finland (Landmark et al., 2003). In this situation clinical supervision sometimes means a one-to-one, co-operative relationship, or it can mean session meeting and discussion leading by supervisors who may be clinical staff or faculty staff (Lindquist et al., 2012; Berglund et al., 2012).

Use of the three terms in North America, the UK, Australia and the other countries

North America

In the USA and Canada preceptorship for students and newly qualified nurses is established in most situations (Myrick and Yonge, 2002); mentors for staff nurses are becoming more popular (Jakubik, 2008; 2012). But 'mentorship' is also used in nursing students' clinical teaching as a long and informal program (Andrusyszyn et al., 2007). 'Supervision' is seldom used to imply such a supporting and developing relationship between registered nurses and students but is mainly used to indicate the relationship between registered nurses and staff in the mental health nursing field (Cutcliffe and Lowe, 2005).

In the UK

The three concepts are quite clear in the UK, where the three roles and concepts are clarified by the statutory body Nursing and Midwifery Council (NMC), which ensures that each concept has a precise definition. They use 'mentors' to support nursing students' clinical learning, and 'preceptors' to help the transition of newly qualified students, while the role of supervisor is developed to support the staff nurse, improve the wellbeing of the bedside nurse and also for the quality of patient care. From 2003

every nursing student has a named mentor to teach, support, and assess them. Additionally, the NMC has advocated that every new qualified nurse should have a preceptorship program for smooth transition from four to six months (Bond and Holland, 2011). The three terms are arranged in a continuum in nursing from students to new graduates, for clinical supervision ending with staff. All these three roles create a comprehensive support system.

Australia

In Australia, preceptor and clinical supervision are popular and freely used, but mentorship is not getting the chance to develop according to McCloughen et al (2006).The concept of preceptorship is imported from the USA. They adopt the UK model for clinical supervision.

Other countries

In other countries, the selection of terms is usually influenced by the UK, or the USA, and their own culture and tradition. For example, some European countries traditionally use clinical supervisors for student nurses clinical education. Under the influence of the Bologna Declaration on Higher Education (BDHE) for joint accreditation in the European Union in nursing education, they are more likely to choose ‘mentor’ when they develop a common training program for mentors of nursing students (Fulton et al., 2007).

In summary, the three terms are applied in clinical nursing across the world as shown in Table 1.5 and Figure 1.1. Mentorship can be used to describe long- and short-term relationships between clinical nurses and nursing students/new graduates/staff nurses. Preceptorship can be applied to describe short-term relationship between staff nurse and

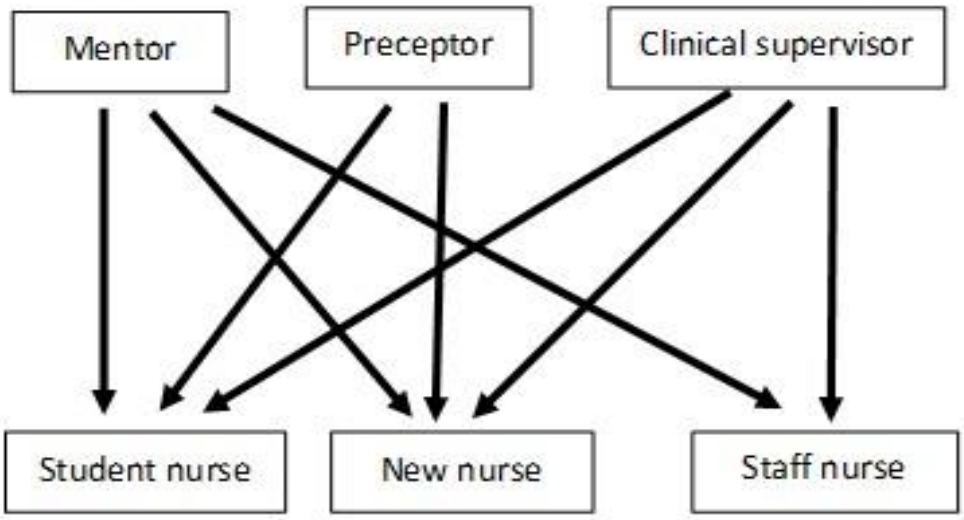
students/new graduates. Supervision means staff-to-staff continuous supportive relationship, one-to-one staff nurse to nursing student short-term relationship, session-based discussion to integrate the theory and practice gap for students, and overseeing and directing managerial strategy. Focusing on the relationship between nurses and nursing students, the three terms are used: preceptor, mentor and supervisor in different countries as shown in Figure 1.1. In nursing education, USA will continue to employ preceptorship as the main strategy, while the UK will lead with mentorship. With respect to nursing staff development, the USA will continue to apply mentorship, while in the UK, clinical supervision will have more weight, imposing more influence on the European countries in both nursing education and staff development. There is no sign of a consensus in the near future between the USA and UK. So the confusion of the concepts and roles of mentoring will continue.

Table 1.5. Concepts analysis in different countries

Term	Geography	Prevalence	Short-term	Long-term	Nurse-to-student	Nurse-to-new staff	Nurse-to-nurse
Mentorship	UK	Mandatory for students	√		√		
	USA	Popular for staff		√	√	√	√
	Canada	Not clear		√	√	√	√
	Australia	Seldom		√			√
	Some European countries	Seldom	√		√		
Preceptorship	UK	Increasing for new nurse	√			√	
	USA	Popular for students	√		√	√	
	Canada	Popular for students	√		√	√	
	Australia	Popular for students	√		√		
	Some European countries	Seldom	√			√	
Clinical Supervision	UK	Popular for staff nurses		√			√
	USA	Popular in psychiatric nursing		√			√
	Canada	Increasing use		√			√
	Australia	Increasing use		√			√
	Some European countries	Popular for staff nurse	√		√		

*some European countries includes: Sweden, Finland, Dutch, Norway.

Figure 1.1 Terms used to imply different relationshps in nursing over the world



1.3.5 Concepts and the operational definition for this study

'Mentor' is chosen as the concept for this research. The rationales for such selection are as follows. First mentoring relationships may be formal or informal and short- or long-term (Sawatzky and Enns, 2009). Meanwhile, 'mentor' will become a more popular term and hotter research topic compared with 'preceptor' as shown in Table 1.2. Additionally, this research is based in the UK and taking Chinese culture and the definition of 'Daoshi' into consideration, 'mentor' will be clearer for communication. But any articles pertinent to the supporting role and has a one-to-one, face-to-face , day-to-day relationship in the clinical nursing arena, using the term 'preceptor', 'mentor' or 'supervisor' will be reviewed.

This research operationally defines mentorship as: a relationship between a registered nurse and a pre-registration nursing student; registered nurses facilitate students' learning and promote personal and professional development of students on a one-to-one, face-to-face, working together base. The rationale for this definition is: first

to clarify registered nurse as a mentor instead of nursing teachers who are employed by nursing schools teach nursing students in clinical placement; second, mentees are pre-registration students who have no clinical experience rather than specialist or other advanced nursing students; third the mentoring relationship is focused on one-to-one, not group mentoring as these two types of mentorship have much difference; fourth face–to –face mentoring is stressed as opposed to e-mentoring or distance mentoring.

1.4 Outline of this thesis

This thesis has five chapters together. Chapter 1 has introduced background information on such things as nursing and nursing education in China; the concepts of mentoring, its theoretical underpinning and function types; mentoring in nursing field; the reason for this study; analysis and differentiation of the three most often used concepts about students mentoring i.e. 'supervision', 'mentorship', and 'preceptorship' across geographical regions. After clarification of concepts, three systematic literature reviews were conducted as shown in Chapter 2 to identify any applicable mentoring behaviour scale in the non-nursing field, nursing field and the nursing field in China. No scale was recognized as suitable for measuring mentors' behaviour in clinical nursing education, so the findings from literature review recommends that a specific scale is needed. Then a temporary mentorship theoretical framework with three dimensions was generated and is proposed as the base of a new scale development. Chapter 3 describes methodologies used in the new scale development and validation process such as online focus groups for item confirmation and generation, an expert panel for content validity and a survey for test-retest reliability and construct validity. Chapter 4 reports all the results of scale development and validation, such as mentorship conceptualization, its hierarchical property and psychometrics including stability over time, discriminant validity in

extreme groups and scalability of ordering student expectation and the hierarchy in students' expectation. Chapter 5 interprets the meaning of these findings and the results are compared internationally and across disciplines. The implications for nursing education on the practical and theoretical sides are also discussed. The limitations of this thesis and scopes of future research are discussed as well.

Chapter 2. Literature review

2.1 Introduction

2.1.1 Structure of the literature review

This literature review consists of four parts:

- Part one aims to identify and evaluate the mentoring scales from non-nursing fields;
- Part two aims at identifying scales assessing mentors' behaviour in the nursing field;
- Part three explores a theoretical framework for nursing students' mentoring, and
- The final part presents evidence from China.

A preliminary search, which set out to obtain basic knowledge and understanding of the research area before systematic reviews, showed that nursing researchers selected assessment tools from the business and education fields to measure mentors' behaviour and that they used different concepts and terms like 'preceptor', 'mentor' and 'supervisor' to describe mentorship as discussed in Chapter 1, so literature reviews were carried out separately and systematically in non-nursing fields and nursing fields, using different terms in an English database. Mentorship in China is relatively new in business, education and health science, which may not impact on nursing much, therefore only articles on mentoring in the nursing field were searched in the Chinese database.

As the main objective of the literature review is to identify and appraise mentoring scales, the scales identified will be analysed in terms of measurement purpose,

theoretical framework/conceptualization, psychometric evidence, and the extent of use by researchers and practitioners. Particularly, psychometric properties are vital for scientific measurement, which shows how accurate a measurement can be, so the theories of psychometrics need to be addressed first, which is presented below.

2.1.2 Psychometric properties of scales

In classical test theory, to judge a measurement, it is imperative to know its reliability and validity. Both of them are conceptually complicated and have multiple facets, when a trait to be measured is invisible such as attitude, behaviour and quality of life.

Reliability

Reliability means to what extent the measurement of a scale is reproducible (Streiner and Norman, 2008). Philosophically, it also includes the meaning of exploring the true value (which is never known) of an object under measurement; 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). Mathematically and practically, the three aspects of reliability - test-retest reliability, internal consistency, and inter-rater reliability - are often 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

Test-retest reliability is applied to explore consistency of a measurement over time, in a group of subjects, and the process is to administrate the same scale to the same group of participants twice, with some interval from two hours to 14 days (Streiner and Norman, 2008 p.182). The key issues when considering the time interval are to avoid memory

and maturation effects as they will bias the test-retest reliability. The Pearson Correlation was the statistical strategy most often used to evaluate the test-retest reliability (Streiner and Norman, 2008), but now the intra-class correlation coefficient (ICC) is more popular and suitable for its accuracy and flexibility in estimation of consistency and absolute agreement of a measurement, which will be discussed further bellow. For a scale, the criterion of test-retest reliability judgement depends on the purpose. For research, 0.7 and above is acceptable (group-based and no consequence); with regard to clinical application, 0.9 and above can be reasonable (as this is individual-based and can have consequences) (Hogan and Cannon, 2003; Streiner and Norman, 2008). As to item level criteria, more details will be discussed in Chapter 3. Items or scales showing low test-retest reliability may imply a problem in understanding, which suggests that actions, such as rewording, are in need.

Internal consistency reliability

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. There are three methods mainly used to calculate internal consistency reliability (ICR): the split-half, Kuder-Richardson formulae, and Cronbach's alpha (α). Split- half reliability divides a scale into halves and investigates the correlation between the two halves; and the results may vary as the dividing methods can be different, such as odd-even number dividing and first-second half dividing. The Kuder-Richardson formulas are suitable for dichotomous variables, but in reality polytomous variables are more commonly used, while Cronbach's alpha can be used by scales with any type of

variables. The number of items can have an impact on internal consistency reliability, generally speaking, the longer a scale is, the higher the reliability will be (Streiner and Norman, 2008). Since internal consistency is based on a single test, it is easier to carry out than other reliability tests, but the interpretation of the results should be done with cautious (Streiner and Norman, 2008).

Inter-rater reliability

Inter-rater agreement or inter-scorer reliability tests different raters' deviation from 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. Intra-class Correlation Coefficient (ICC) is the most popular method to compute it.

ICC has many models (Shrout and Fleiss, 1979; McGraw and Wong, 1996), considering four factors: one-way (used in nested design and where there is no ordering among nests) or two-way model (there is systematic variance of measurement at different observations of each subject); fixed or random level (raters are chosen randomly or not; aiming at generalization or not); single or average measurement; consistency or absolute agreement (considering whether there is systematic variance among raters or not). ICC using analysis of variance (ANOVA) is recognised as superior to the Pearson Correlation using a linear model for many aspects: conceptually, the Pearson Correlation tests inter-class correlation (the relationship of two different variables), while intra-class tests the correlation of scores of one variable at different times or rated by different people; ICC has the ability to isolate factors affecting reliability; it is more flexible; simulating the effect on reliability of increasing or decreasing the number of

raters. It can also be a very good substitute for Cohen's Kappa coefficient (Streiner and Norman, 2008).

Reliability is essential for assessment of a scale's quality, which can impact on the validity and decide the maximum of validity (Streiner and Norman, 2008), but it is not sufficient. A scale with reliability tells you the scores are consistent in different situations but it cannot assure you how true the outcomes are and whether it measures the trait you intend to measure. However validity is capable of showing these.

Validity

Validity is the extent to which a tool measures the concept that it purports to measure, but it is considered to be sample- and context-influenced rather than an intrinsic property of an instrument (Streiner and Norman, 2008). 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

Content validity usually includes face validity; neither is in need of formal statistical testing. Face validity means using words and semantic analysis to judge the items' relevance to the concept it is intended to measure and it is the easiest and quickest way to establish validity. However it is far from sufficient. 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. Quantitative methods such as index of content validity (CVI) and content validity rating (CVR) can be used to make it more objective (the details of

types of CVI, CVI calculation, expert selection and criteria of CVI are reported in Chapter 3). The higher the content validity is, the broader the inference researchers can make (Streiner and Norman, 2008). 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

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 is. 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. Criterion validity includes concurrent validity and predictive validity. Concurrent validity testing needs to distribute the new scale and the criterion to the subjects simultaneously, and the correlation coefficient can be calculated using cross-tabulation for categorical variables or Pearson's Correlation for interval variables (Streiner and Norman, 2008). Predictive validity means the ability of a scale to predict the outcome, such as the scores of students in high school predict their performance in university; the standard measurement will be administered at a later time. If the research is exploring a new area without any tool or any 'gold standard' existing, it is impossible to test the criterion validity of a new tool, but it is feasible to establish its construct validity.

Construct validity

A construct or a hypothetical construct refers to an unobservable trait (latent trait) which can just be observed or measured indirectly though the manifested behaviours. When constructing a new construct, people need to prove this new construct is true or better

than existing constructs. This is the process of establishing construct validity. It is different from content and criterion validity testing: construct validity testing is a long process. In the process, different kinds of hypotheses and theories can be constructed and tested simultaneously (Streiner and Norman, 2008). It includes many categories: in this section, convergent and divergent validity, factorial validity, i.e. exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), and discriminant validity with extreme groups are discussed.

Convergent and divergent validity

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, such as depression is not associated with intelligence. One noteworthy point is that the criteria of correlation coefficients are not very clear, such as how high or how low is considered to be enough; some researchers considered that too high may indicate the two tools measure the same construct (Streiner and Norman, 2008); but one thing is sure: the correlation coefficient of divergent validity should be lower than that of convergent validity.

Factorial validity

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. If items show loading

coefficients over 0.4 on a factor, they converge to this factor; at the same time they need to be divergent from other factors (the cross-loading coefficient should be lower than 0.4) (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. More theories and strategies of factor analysis are discussed in Chapter 3.

Discriminant validity with extreme groups

Extreme groups are applied to test the ability of a tool to differentiate between two groups of people with and without a certain trait, like mentored and non-mentored groups. Usually a t-test is needed to compare the means of the two extreme groups to prove the differential ability. It is the minimum differentiating ability that a tool should have. In reality a tool should differentiate between levels of a trait.

All the above psychometric theory is based on classical test theory. 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. 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); the theories and advantages of item response theory will be discussed in Chapter 3.

2.2 Literature review objectives

The objectives of the literature review are:

- To identify tools for measuring mentors' behaviour
- To analyse the tools critically
- To explore a theoretical framework of mentoring nursing students in clinical education.

2.3 Review of mentoring scales from non-nursing fields

The first part of the literature review aims to identify if there was any mentoring scale in non-nursing fields applicable to measure mentors' behaviour in the clinical nursing education field. Search strategies, criteria and systematic review process are shown in Table 2.1 and Fig 2.1

2.3.1 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

2.3.2 Search strategies

The search terms and strategies used were: Mentor* N3 (behav* or skill* or role* or activit* or function* or relation*) and (scale* or tool* or instrument* or questionnaire* or inventor*) not nurs*

These search strategies and terms were decided after consulting an expert in the University of Hull. Time limiter was set from January 1980 to August 2013, for mentorship has gathered momentum from the 1980s; language and age group limiters were applied. Truncation was used and the reference lists were also inspected for a more comprehensive search.

2.3.3 Inclusion criteria

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

2.3.4 Exclusion criteria

- Studies not about mentoring
- 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

Table 2.1 Criteria and search strategies used in literature review

Criterion 1: Limiters	Published Date: 01/01/1980-31/08/2013; 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 inventor*) not nurs*
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 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)

2.3.5 Results

The process of systematic review is presented in Figure 2.1. The results are shown in Table 2.2, including 26 papers related to 20 mentoring scales: eight scales from the business field, 11 from the education field and one from medicine and allied health field.

Figure 2.1 Flow diagram of the systematic review in non-nursing fields

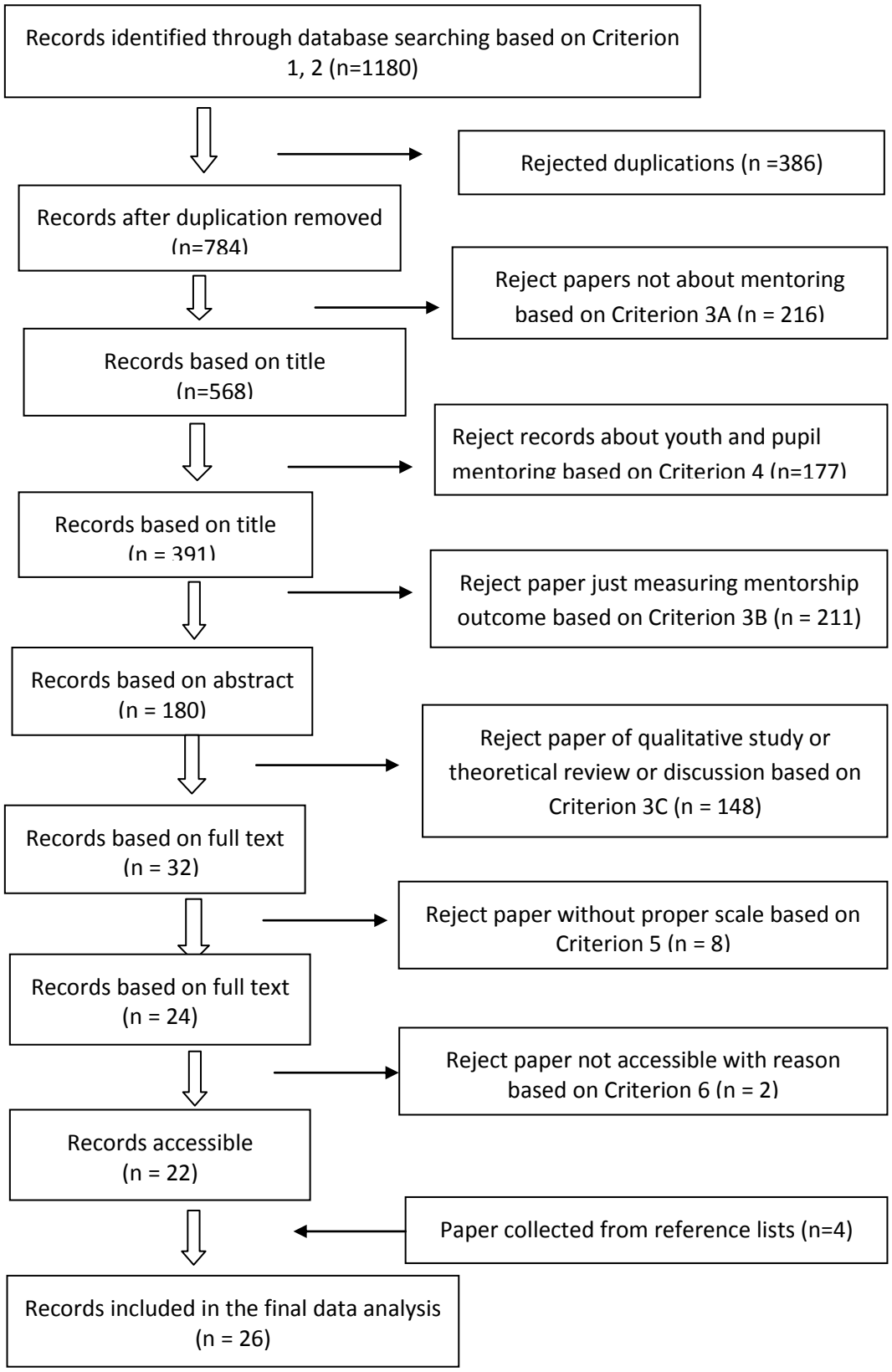


Table 2.2. Scales identified in non-nursing field

Reference (authors)	Scale name and number of items	Subscales (conceptualization/ content)	Psychometric properties	Comments (theoretical framework, target, application)
Busch, 1985	Mentoring instrument: mentors' perception (69)	Two subscales: behaviour and relationship	EFA	No claimed theoretical framework; showing factorial validity; aiming at measuring postgraduate students mentorship; no further use was located.
Schockett and Haring-Hidore , 1985	Mentoring function16	Two factors: career development and psychosocial support	EFA	Based on the two-function model of mentoring, with some psychometric evidence; used in business and industry.
Alleman, 1987, Alleman and Clarke, 2002	Alleman Mentoring Activities Questionnaire (AMAQ, 72)	Nine subscales: teach the job, provide challenge, teaching policies, career help, protect, sponsor, career counselling, friendship and demonstrated trust	ICR, IRA, CRIT, DISC, CONC	No claimed theoretical framework; showing variety of psychometric evidence; used widely by large international corporations, nurses, academics, administrative staff and graduate students.
Noe, 1988	Mentoring Functions Scale (MFS, 29)	Two functions: psychosocial function and career function	EFA CONT, ICR	Based on the two-function model of mentoring; showing some psychometric evidence; widely used in business and industry, and also in nursing staff mentoring.
Ragins and McFarlin, 1990	Mentor Role Instrument (MRI, 33)	Two factors: career and psychosocial function	CFA ICR	Based on the two-function model of mentoring; showing some psychometric evidence; widely used in the field of business and industry and in nursing staff mentoring
Wilde and Schau, 1991	Mentoring instrument: mentees perception (65)	Four factors: psychological and professional mutual support, comprehensiveness, mentee profession	PCA	No claimed theoretical framework; showing a little psychometric evidence; aiming at measuring postgraduate students mentoring; no

		development, research together		further use.
Sands and Parson, 1991	Ideal mentoring function (29)	Four factors: friend and support, career guide, information, and intellectual guide	ICR, CRIT, CONT	Using Ericson's Adult development theory as theoretical underpinning; showing some psychometric evidence; aimed at measuring teaching staff mentorship; used by nursing teaching staff.
Scandura, 1992; Scandura and Ragins, 1993; Pellegrini and Scandura, 2005; Hu et al., 2011	Mentoring function Scale (MFS, 9)	Three subscales: psychosocial function, career function and role modelling	MGCFA, CFA, EFA	Based on the two-function model of mentoring, a three-function model is established and confirmed; showing continuous psychometric evidence; short length; used widely by large international corporations, nurses, academics, administrative staff and graduate students
Pollock, 1995	Mentoring functions (19)	Two factors: career and psychosocial function	CFA ICR	Based on the two-function model of mentoring; showing some factorial validity evidence; used in business and industry
Cohen, 1995	Principles of Adult Mentoring Inventory (PAMI, 55)	Six behavioural functions: relationship emphasis, information emphasis, facilitative focus, confrontation focus, mentor model, and student vision	ICR	'The principles of mentoring function' is cited widely, with just internal consistency reliability; used by doctoral students for dissertation purpose.
Rose, 2003; 2005	Ideal Mentor Scale (IMS, 34)	Three factors: integrity, guidance, and relationship	CONT, CONV, ICR, EFA, CFA,	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.

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	Based on the two-function model of mentoring; showing some psychometric evidence; newly developed.
Hudson et al., 2005	Mentoring for effective primary science teaching (MEPST,45)	Five factors: personal attributes, system requirement, pedagogical knowledge modelling, feedback	CFA	No claimed theoretical underpinning; based on literature review, the new scale was constructed to measure science teachers' mentorship; showing no further use.
Eby et al., 2008	Negative Mentoring experience scale (NMES, 36)	Three factors: Performance problem, interpersonal problem, destructive relational patterns	CFA, CONT, CRIT, CONV, DISC	Social exchange theory is the theoretical underpinning; showing wide range of psychometric evidence; newly developed and validated; aiming at measuring negative mentoring experience.
Crisp, 2009; Crisp and Cruz, 2010	College Student Mentoring Scale (CSMS, 25)	Four factors: psychological and emotional support, degree and career support, academic subject knowledge support, and role model	ICR, CFA, MGCFA	Based on literature review; showing some psychometric evidence; aiming at measuring college students' mentorship; newly developed and no further use.
Pamuk and Thompson, 2009	Technology mentor benefits instrument (28)	Three factors: technical benefit items, academic benefit items/profession pedagogical benefit	EFA, ICR, CONT,	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.
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	EFA, ICR	Social exchange theory is the theoretical underpinning; showing some psychometric evidence; newly developed and validated; aiming at measuring mentoring relationship challenge in business.

Koc, 2011	Mentor Teacher Role Inventory (MTRI 49)	Eight factors: providing support on teaching, orientation to the school/classroom, providing moral support, providing feedback on lesson planning and teaching performance, guidance about resources for teaching, evaluation, providing feedback on observation forms, self-preparation for the mentor role	EFA, ICR	No claimed theoretical framework was identified. It is a newly developed scale to measure teacher's mentorship, with no further test nor use.
Harris, 2013	Perception of Mentoring relationships survey (PMRS, 24)	Three subscales: benefits of mentoring, mentor's role and mentee's role	CONT, EFA, ICR	Social learning theory is the theoretical underpinning; with some psychometric evidence; aiming at measuring college students mentoring; newly developed.
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	No claimed theoretical framework was identified; measures researcher mentors' competency in medicine; newly developed.

ICR: internal consistency reliability
IRA: inter- rater agreement,
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

2.3.6 Discussion

Theoretical framework/conceptualization

In the field of business and organization, mentorship is recognized as a human resource development tool, which was conceptualised as two domains (career development and psychosocial support) (Kram, 1983; Kram and Isabella, 1985), and this structure is supported by five scales (Pollock, 1995; Ragins and McFarlin, 1990; Noe, 1988; Busch, 1985; Schockett and Haring-Hidore, 1985). Later, the two-function model was split into a three-function structure (career, psychosocial and role modelling function), and it 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 (staff development).

In the education field, 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 the education field mentorship is conceptualized differently as it is used in varying situations, such as mentoring of staff, mentoring of college students and mentoring of PhD students and that mentorship is relatively new and has not drawn enough attention to develop and validate a strong measurement tool or to establish a ubiquitous theoretical framework.

Reliability and validity

The most frequently used measure of reliability in the 26 papers is internal consistency reliability due to its quick and easy nature; one scale (Alleman and Clarke, 2002) reported inter-rater reliability. None presented test-retest reliability; this may imply that mentorship assessment is at the stage of construct understanding and exploring, while the precision and consistency of mentorship measurement has not been so acute in the business field, and more efforts are needed (Allen et al., 2008).

With regard to validity, factorial validity is investigated more frequently than others; convergent and divergent validity are also explored (Rose, 2003; Alleman and Clarke, 2002; Eby et al., 2008). Measuring equivalence/invariance is tested using multi-group confirmatory factor analysis by Hu et al (2010), which 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 is used by two scales (Alleman and Clarke, 2002; Eby et al., 2008), which may suggest 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, and so on.

Extent of use

The Alleman Mentoring Activities Questionnaire (AMAQ) is a widely and commercially used scale in business, education and nursing (Lee and Carmen Montiel, 2011; Kavoosi et al., 1995), having proper instruction on administration and scoring, while the Mentoring Functions Scale (Scandura, 1992; Scandura and Ragins, 1993; Pellegrini and Scandura, 2005; Hu et al., 2011) is becoming more popular within the business and other fields, which may be due to the short length and stable three-dimensional structure, while the absence of copy right may be another reason. The two-dimensional mentoring scales (Dreher and Ash, 1990; Noe, 1988; Ragins and McFarlin, 1990) are also widely used in different fields with up to 1000 citations (Allen et al., 2008). All these scales are used in nursing field 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). Among them, AMAQ may be the most widely used in the nursing field (Richard, 1995; Jones, 1997; Aponte, 2007; Kavoosi et al., 1995).

In the education field, more new scales were developed recently (Harris, 2013; Koc, 2011), but further study is needed, while in the health allied area, medical educators and researchers began to develop their own mentorship scales (Fleming et al., 2013). It shows a specializing process of mentorship conceptualization and measurement. Very few of them were used in nursing field except one scale (Sands et al., 1991) which was used to assess faculty mentoring in nursing school (Frandsen, 2003).

This reflects the fact that mentorship originated from business and is obtaining public acknowledgement across disciplines. Nursing professionals recognized the importance of applying mentorship and the significance of precise measurement of its effectiveness. They chose some assessment tools from business or education to do that due to lack of their own scales. But the assessment was confined to staff nurses' mentorship or nursing teachers' mentoring, as no single study using these scales to measure pre-registered students' mentorship in the field of clinical learning was identified. It may imply a conceptualization difference between student mentoring and staff mentoring in nursing.

2.3.7 Summary

Mentorship measurement was pioneered by the business discipline with universally accepted theoretical framework, i.e. career function and psychosocial function, and the trend of scale development is becoming deeper and more comprehensive: from focusing on the positive side of mentorship shifting to negative mentoring experiences and challenges (Eby et al., 2008; Ensher and Murphy, 2011). In the education field, the measurement is heading to a more specialized process. For example, researchers in different subjects are developing their own scales, as mentorship takes place in different contexts and the conceptualization varies (Pamuk and Thompson, 2009; Koc, 2011; Harris, 2013). 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. Scales from the business and education fields are used by nursing researchers to measure mentorship of staff nurses or teaching staff, but none have been used to assess the mentoring of nursing students. This implies that no scales existing in the non-nursing fields are suitable to measure nursing students' mentorship. So the search for scales needs to be carried out in nursing database, which is reported in the next section.

2.4 Review of mentoring scales from nursing field

The second part of the literature review aims to identify if there is any suitable mentoring scale in the nursing field which can be used to measure the mentoring of students' in clinical education. The search strategies, terms and the criteria of inclusion and exclusion are presented in Table 2.3 and Fig 2.2

2.4.1 Search strategies

The search terms and strategies used were (Precept* or supervis* or mentor*) N3 (behav* or skill? or role? or activit? or function* or relation*) AND (scale? or tool? or instrument* or questionnaire? or inventor? or effec* or quality? or experienc* or percept* or measur* or assess* or evaluat*) AND nurs*

Time, language and age group limiters were used, shown in Table 2.3. Reference list was inspected and relevant papers were included.

2.4.2 Databases

- CINHAL
- Medline
- PsycINFO
- ERIC
- Academic Search Premier

2.4.3 Inclusion criteria

A. Articles about the development and validation of mentoring scales

B. Subjects are staff nurses and nursing students

C. Functions, roles, behaviours, skills and activities of mentors, preceptors and supervisors

D. Articles should be research based.

2.4.4 Exclusion criteria

A. Papers are not related to mentorship.

B. Mentorship about patients, midwives, medical students， nursing assistants, academic staff, or any other groups of people

C. Studies about specific competence of mentor/preceptor/supervisor required (teaching, ethic, leadership, assessment)

D. Internet mentoring program, group mentoring and peer mentoring

E. Comments or other non-research based articles

F. Full text not accessible

Table 2.3 Criteria and search strategies used in literature review

Criterion 1: limiters	Published Date: 01/01/1980-31/08/2013; Medline: English Language; Human; Age Related: All Adult: 19+ years; Cinhal: Language: English; English Language; Human; Eric: Educational Level: Higher Education; Language: English; PsycINFO: English; Age Groups: Adulthood (18 yrs & older) Academic Search Premier: Language: English
Criterion 2: Terms / concepts / keywords	(Precept* or supervis* or mentor*) N3 (behav* or skill? or role? or activit? or function* or relation*) AND (scale? or tool? or instrument* or questionnaire? or inventor? or effec* or quality? or experienc* or percept* or measur* or assess* or evaluat*) AND nurs*
Criterion 3: Content	Quantitative or qualitative studies about face-to-face, one-to-one nursing students mentoring and papers related to mentoring scales development and validation are included, while three kinds of papers listed below are excluded: A. Studies not about mentorship but leadership, nursing role, patients care and so on B. E-mentoring, group mentoring/supervision, peer mentoring C. Studies about particular functions of mentorship such as turn-over decreasing, job satisfaction, stress reduction, EBP and teaching
Criterion 4: Subject	Pre-registration nursing students and nurse mentors are included while studies about mentoring other participants such as patients, nurse, faculty staff are excluded.
Criterion 5: Publication review	Scientific studies are included, while comments are excluded
Criterion 6: Accessibility	Likelihood of availability, time and budget constraint

2.4.5 Results

From the five databases, 46 papers were included (20 articles reporting 13 scales; 26 papers using quantitative and/or qualitative methods to explore the mentoring function/roles and mentors' behaviours in nursing student's clinical study) following the criteria. In this part, only 20 papers about scale development and validation are reviewed (Table 2.4), while all the 46 papers (see the references in Figure 2.3 and Appendix 3) are reported in the next section of mentorship theoretical framework exploration. The results and systematic review process are presented below in Table 2.4 and Figure 2.2.

Figure 2.2 Flow diagram of the systematic review in nursing field

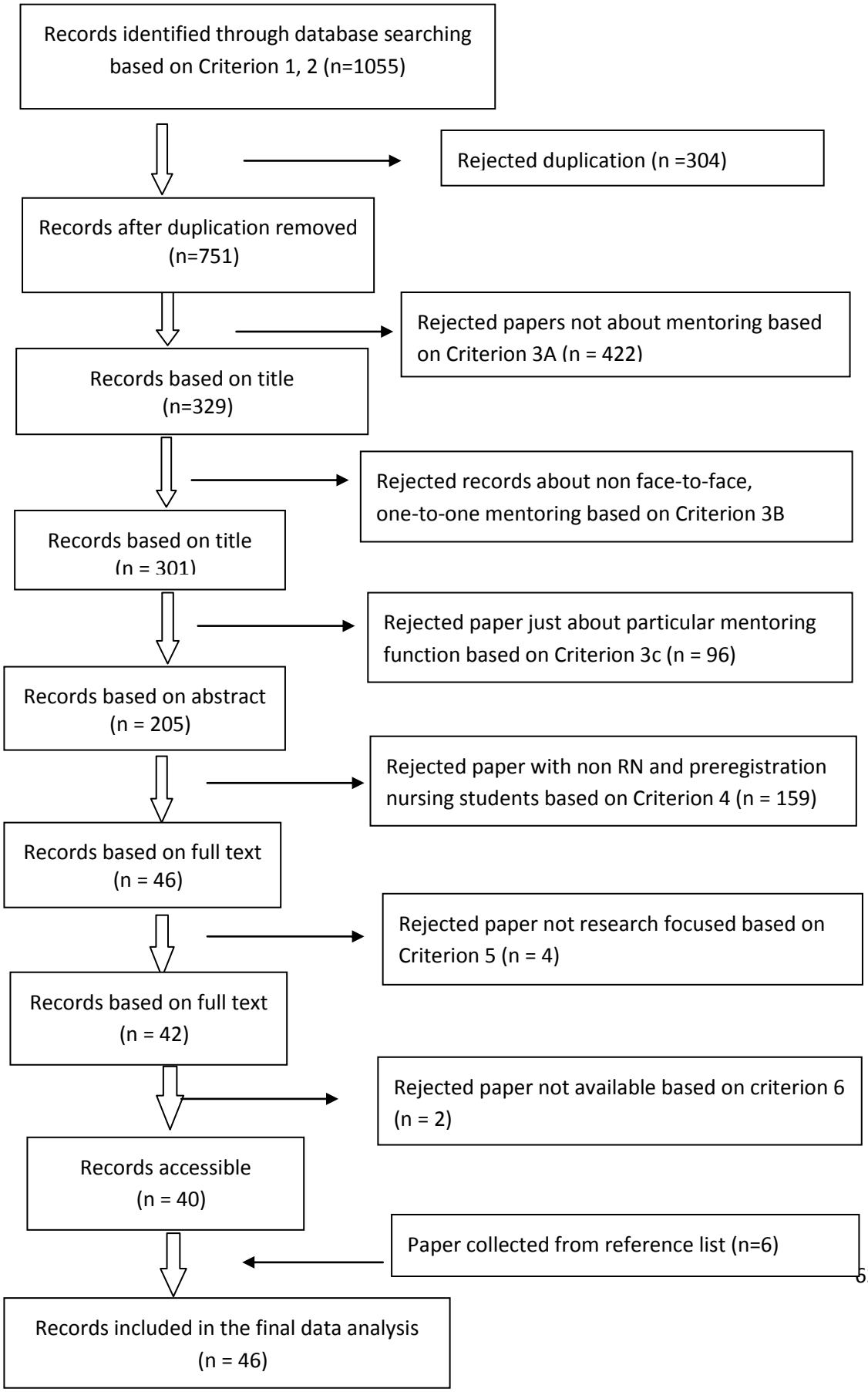


Table 2.4. Mentoring scales in the nursing field

Reference (authors)	Title of Scale and number of items	Subscale/ concept-ualization/content	Psychometric properties	Comments (theoretical framework, target and use)
Darling, 1984	Measure of mentoring potential (MMP, 14)	Three functions: inspirer, investor and supporter	No report	Self-developed three principal functions of a mentor served as the theoretical framework; no psychometric evidence was shown; aiming to assess staff nurses' mentorship; seldom used by other studies although it is the first mentoring scale cited frequently it is.
Knox and Mogan, 1985	The Nursing Clinical Teacher Effectiveness Inventory (NCTEI, 47)	Five subscales: teaching ability, nursing competence, evaluation, interpersonal relations, personality	ICR, TestR, face and CONT	No claimed theoretical framework; showing some psychometric evidence; aimed at measuring clinical nursing teachers' capability; used internationally.
Dibert and Goldenberg, 1995	Preceptorship benefit and reward scale (PPBR, 14), Support scale (PPS, 17), Commitment to preceptor role (CPR,10)	Three subscales: benefit and reward, support, and commitment	ICR	Based on Kanter's (1977) theory of empowerment structures (information, resources, opportunities, and support); showing little psychometric evidence; aimed at measuring mentors' reward, support and commitment to student nurses' preceptoring; used internationally.
Hall, 1997; 1998	Scale about Mentoring relationship (40)	Eight subscales: encouragement, sponsorship, teach the job, teach the informal, role modelling, exposure and visibility, counselling, protection	Claimed but not reported	Based on two-function mentoring theory in the business field; showing no psychometric evidence; aimed at measuring staff nurses' mentoring relationship; no further use.

Chow and Suen, 2001; Suen and Chow, 2001	Scale of students mentoring (33)	Five subscales: befriending, guiding, advising, counselling and assisting	Face and CONT	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.
Saarikoski and Leino-Kilpi, 2002; Saarikoski et al., 2005	Clinical learning environment and supervision scale (CLES) + nursing teacher (NT) (27; 34)	Five subscales: pedagogical atmosphere, supervisory relationship, premises of nursing care on ward, leadership style of ward manager, role of nursing teacher	EFA, CFA, CONC, Test-R, ICR	Based on literature review; showing a wide range of psychometric evidence; aiming at measuring the whole nursing students' clinical learning environment; tested and used internationally.
Winstanley and White, 2003; 2011, Severinsson and Sand, 2010	Manchester Clinical Supervision Scale (MCSS, 36, 26)	Seven subscales: trust/rapport, supervisor advice/support, improved care and skills, personal issues, importance/value of (CS), finding time, reflection	ICR, EFA, CFA, RAS	Based on Proctor's (1986) theoretical framework of three supervision functions: normative, restorative and formative; showing a wide range of and advanced psychometrical evidence; aiming at measuring staff nurses' mentoring benefits and effectiveness; used internationally by 12 countries.
Berk et al., 2005	Mentorship Effectiveness Scale (12)	One	No reported psychometric evidence	No claimed theoretical framework; showing no psychometric evidence; aimed at measuring nursing teaching staff mentorship; used by nursing and medical science educators.
Lee et al., 2009	Satisfaction of preceptor's teaching behaviour (20)	One	ICR, CONT (CVI)	No claimed theoretical frame work; showing little psychometric evidence; aimed at measuring the behaviour of new staff nurses' mentor; no further use.

Hallin and Danielson, 2010	Scale about preceptor performance (42)	Nine subscales: students' responsibility, support from university, assistant nurse, ward manager, colleagues, perception and action as preceptor, feedback, preparation	EFA,CFA, ICR, face and CONT	Based on Kanter's theory of empowerment structures; showing some psychometric evidence; aimed at measuring' mentors' reward, support and experience and preparation of student nurses; newly developed and used locally.
Weng et al., 2010	Mentoring quality scale (9)	Three subscales: Career, psychosocial function, role modelling	CFA, CONV, DISC	Adapted from Scandura's (1994) mentoring function scale; showing some psychometric evidence; aiming at measuring staff nurses' mentoring; newly adapted, no further use.
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	Based on the mutual benefits theory (Zey, 1992) in the business field; showing some psychometric evidence; aimed at measuring staff nurses' mentoring; newly developed and used by the author herself later.
Kristofferzon et al., 2013; Löfmark et al., 2012	Nursing Clinical Facilitator Questionnaire (NCFQ, 24)	Not clear	ICR	No claimed theoretical frame work; showing little psychometric evidence; aiming at measuring the behaviour of students' mentor; newly developed.

ICR: internal consistency reliability
TestR: Test-retest reliability
CONC: concurrent validity
CONT: content validity
EFA: exploratory factor analysis
CFA: confirmatory factor analysis
RAS: Rasch modelling
Face: face validity

2.4.6 Discussion

Theoretical framework

In North America

In the USA, the mutual benefits theory from the business field (Zey, 1991) was used as a framework by Jakubik (2008). It views the mentoring relationship as a strategy beneficial to organization, mentee and mentor, involving mutual investments and benefits from and to the three parties. Mentors invest in mentees at four levels of mentoring: teaching, supporting, providing organizational intervention, and sponsoring.

The theory of structural power in organizations (Kanter, 1977) was also used by nursing researchers to develop scales to assess the support for mentors (Hallin and Danielson, 2010; Dibert and Goldenberg, 1995). This theory proposes that accessing empowerment structures (information, resources, opportunities, and support) leads to self-efficacy, motivation, commitment, and job satisfaction.

These two theories from the business and organization field are useful to guide mentoring instrument development in the nursing field, but they raise concerns about different aspects of mentorship such as benefits and rewarding, which do not focus on mentors' behaviour and functions towards mentee. Particularly, Kanter's theory is more popular and used as a theoretical framework of mentors' benefit and support rather than the benefits for nursing students (Hallin and Danielson, 2010; Dibert and Goldenberg, 1995). This implies that they are not suitable for students' mentorship.

In the UK

In the UK, Proctor's (1986) three-dimension theoretical framework is a widely accepted mentoring (supervision) theory. The Manchester Clinical Supervision Scale (Winstanley

and White, 2003; 2011) is based on it. The three-supervision-function theory is presented below (Proctor, 1986 p. 21–34):

- Normative: to promote and comply with policies and procedures, development of standards, and contribution to clinical audit
- Restorative: to enable practitioners to better understand and manage the emotional burden of practice
- Formative: to develop knowledge and clinical skills

But this three-function model describes staff nurses' mentoring experience rather than students' mentoring experience, therefore it is used to construct staff nurse mentoring scales rather than students mentoring instrument. For example, Chow and Suen (2001) did not apply it as a theoretical framework for tool development but used the five roles of mentors defined by English National Board for Nursing, Midwifery and Health Visiting: befriending, guiding, assisting, counselling and advising (ENB, 1988). No theoretical framework was reported in some nursing students mentoring instruments (Knox and Mogan, 1985; Löfmark et al., 2012; Lee et al., 2009), which may cause a construct ambiguity theoretically, as theory is necessary for developing a sound instrument (Streiner and Norman, 2008).

As discussed above, some widely accepted theoretical frameworks of nursing mentorship were identified, such as Proctor's and Kanter's theories. But they focus on different aspects of mentorship, not showing fitness for students mentoring. These proved that there is an absence of a sound nursing students' mentorship theory.

Reliability and validity

Eight out of the 13 scales did not report enough psychometric properties; often, they reported internal consistency reliability and/or described the content validity,

particularly the following six tools (Hall, 1998; Darling, 1984; Berk et al., 2005; Lee et al., 2009; Suen and Chow, 2001; Löfmark et al., 2012). But two scales were exceptional: the Clinical Learning Environment and Supervision Scale (CLES) + nursing teacher (NT) and the Manchester Clinical Supervision Scale (MCSS).

The CLES +NT has comprehensive psychometric evidence, including face validity, content validity, construct validity, concurrent validity, test-retest reliability and internal consistency reliability, using different samples in deferent countries (Saarikoski and Leino-Kilpi, 2002; Saarikoski et al., 2005; Saarikoski et al., 2008). The MCSS has presented advanced statistic strategy, i.e. Rasch Measurement Model.

This suggests that, to a large extent, psychometric testing is not carried out scientifically and sufficiently in the development and validation of instruments in the nursing mentorship field, showing relative weakness compared with the non-nursing fields in general. Some researchers who are dedicated to scale development and validation put psychometric testing and objective measurement at the centre, using robust statistics and international samples, which implies a direction of scientific and accurate measurement across cultures in nursing mentoring field.

Extent of use

The MCSS has been used by more than 90 research programs in the field of nursing and midwifery in 12 countries (Buus and Gonge, 2013; Winstanley and White, 2011; Severinsson and Sand, 2010). This indicates the acceptance of the scale and demonstrates that the importance of nursing staff mentoring is increasingly and broadly recognised. 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.

The CLES+NT has been translated into Dutch, Italian, Norwegian, Swedish and Germany (De Witte et al., 2011; Bos et al., 2012; Henriksen et al., 2012; Bergjan and Hertel, 2013) and has been used in European countries by many nursing researchers. This shows that clinical teaching of nursing students and measurement of clinical learning environment draw more attention internationally and nursing research is approaching more international cooperation and understanding.

The Nursing Clinical Teacher Effectiveness Inventory (NCTEI) has been often used by international nursing researchers (Kotzabassaki et al., 1997; Benor and Leviyof, 1997; Allison-Jones and Hirt, 2004), although the psychometric evidence just includes internal consistency reliability and test-retest reliability. This may reveal that the clinical nursing teacher role is used internationally, and there is a lack of proper knowledge of measurement instrument selection.

The Measure of Mentoring Potential (MMP) has been frequently cited but seldom used to test mentoring function and behaviour (Andrews and Wallis, 1999) for the 'unclear information on sample selection, interview data analysis; lack of justification for findings' (Gray and Smith, 2000 p.1543); there has been no report of reliability and validity; in addition, the concepts of investor, door-opener, eye opener, need further definition.

Critique on the four most relevant or robust mentoring scales

All the 13 scales were examined based on psychometrics, targeted population, and extent of use (see, Table 2.4). The following two scales are identified as the most outstanding based on their psychometrics and extent of use:

- The Manchester Clinical Supervision Scale (MCSS) (Winstanley and White, 2003; 2011, Severinsson and Sand, 2010)

- The Clinical learning environment and supervision scale (CLES) + nursing teacher (NT) (Saarikoski and Leino-Kilpi, 2002; Saarikoski et al., 2005).

Two scales are identified as the most relevant to nursing students' mentoring based on the samples used in development and validation:

- Suen & Chow's (2001) students mentoring scale
- The Nursing Clinical Teacher Effectiveness Inventory (NCTEI) (Knox and Mogan, 1985).

One robust scale, the MCSS, has a clear and sound theoretical framework, detailed development and validation process, showing rigorous testing of reliability and validity, wide acceptance and use for more than ten years(Winstanley and White, 2003; 2011, Severinsson and Sand, 2010) . But it is not suitable to measure students' mentors' behaviour for the following reasons. First, the stimulus to develop the scale was to measure the effectiveness of clinical supervision between staff nurses in placement to enhance the quality of nursing care and peer support, based on interviewing nurses accepting and providing clinical supervision and mentoring. Although Winstanley and White (2003) declared that this tool could be used both by staff nurses and student nurses, one study which recruited nursing students to test it (Severinsson and Sand, 2010) showed that the scores from the three subscales 4, 5, 6 were very low. This may due to the unsuitability of the scale. Second, staff mentoring function focuses on career and nursing care quality from a clinical governance perspective, while student mentoring emphasises learning and professional transition and development from an educational point of view (Hardyman and Hickey, 2001). Third, mentoring for staff lasts for long periods while student mentoring usually continues a few weeks on each placement. Fourth, nursing students mentoring relationship is based on day-to-day work

and a one-to-one relationship, while mentoring staff is usually session-based or group-based. Fifth, the 13 items in subscale 4, 5, 6: personal issues, importance/value of clinical supervision (CS) and finding time for supervision, such as 'I find CS sessions time consuming', 'CS is unnecessary for experienced/established staff', are considered to be irrelevant to nursing students clinical mentoring for they do not measure the behaviour or role of mentors and also for students mentoring is mandatory and day-to-day and one-to-one based. However, trust/rapport, supervisor advice/support, improved care and skills, reflection can be very relevant to mentoring nursing students.

Another scale with rigorous psychometric properties and wide acceptance is the CLES+NT(Saarikoski and Leino-Kilpi, 2002; Saarikoski et al., 2005), but it is still not suitable for nursing students' clinical mentoring measurement. One reason is that it focuses on the clinical learning environment, has just eight items related to general mentoring relationship, which is a small part of the mentors' behaviour. The other reason is that with regard to the subscale of nursing teacher (NT), it offers little guidance to mentors as nurse teachers are employed by an educational institution and act as linking roles between mentors and nursing students in the clinical teaching process.

The third scale (Chow and Suen, 2001; Suen and Chow, 2001) has contents very relevant to mentoring students, including 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.

Finally, the Nursing Clinical Teacher Effectiveness Inventory (Knox and Mogan, 1985) has the longest history, assessing 47 behaviours of clinical teachers, and it is accepted and used widely. Clinical teachers have different roles from mentors; they are nurses employed by nursing schools, teaching groups of students in full time in clinical placement (Knox and Mogan, 1985), while mentors work with students on a day-to-day base, employed by hospitals, working as full-time nurses. The conceptualization of their roles and functions should be different: since teaching and nursing competence are critical for clinical teachers, so teaching skills are more specific and emphasised，including items such as explaining clearly and emphasising what is important (Knox and Mogan, 1985), while since mentoring is conceptualized as facilitating learning and professional development (Jokelainen et al., 2011; 2013; Andrews et al., 2006; van Eps et al., 2006), the role modelling function will be more influential, and relationship will be more important on the one-to-one basis. In addition, the psychometric evidence is not sufficient, lacking construct validity and criteria validity. So this scale is not recognised as a suitable tool to measure mentors' behaviour in the clinical students teaching field. Undoubtedly, there are many behaviours that both clinical teacher and mentors can share.

2.4.7 Summary

Thirteen scales in the nursing field were identified. These scales measure different aspects of mentorship in different situation; some measure mentors' support, reward and benefits, some measure mentors' behaviour in clinical nursing education field; some

measure clinical supervisor or mentors benefits and behaviour among clinical nurses or faculties; some measure nurse teachers' (employed by school) behaviour and function toward nursing students. Correspondently the theoretical frameworks and conceptualizations are different, and no agreement has been reached about mentorship in clinical education field. In addition, psychometric evidence is weaker compared to scales used in the business field. No ideal scale was found to measure mentors' behaviour in the clinical nursing education field after critically assessing the scales found in this review. Therefore a new specific scale is needed. To develop a new scale, a theoretical framework is required first, so the next section will try to establish one.

2.5 Exploring a theoretical framework for mentors' behaviour in clinical nursing education

In the literature review Section 2.4, mentoring tools in nursing were searched and appraised, which showed no instrument that was applicable to assess mentors' behaviour in clinical nursing education and implied that a new and specific tool was needed. Therefore, in this section a theoretical framework was generated to be the foundation for a new scale development

2.5.1 Data

The papers (n=46) used to construct the framework included the 20 papers related to 13 instruments (see Table 2.4) analysed in former section of mentoring scale searching and appraising in nursing and 26 other mentoring studies were also included, which were selected based on the literature review process, see Figure 2.2. All the references were listed in Appendix 3 and Figure 2.3.

2.5.2 Analysis method

Content analysis was used to analyse the data. It is a commonly used way to analyse written data, and data from interviews and observation. It is mainly divided into two subcategories: inductive content analysis and deductive content analysis. Inductive analysis is used when the theoretical framework is unclear or fragmentary, while deductive analysis is applied to retest existing theories (Hsieh and Shannon, 2005; Graneheim and Lundman, 2004). As to this study, there is no single theoretical framework of clinical student nurse mentoring universally established nor accepted, therefore inductive content analysis would be more sensible. After repeated reading of the articles and items in the scales, familiarization and whole sense have been obtained, and codes were noted based on each article, then similar codes were grouped together to

generate subcategories. Finally, similar subcategories were categorized into upper categories.

2.5.3 Results

Based on the 46 articles (see the tefencec in Appendix 3 and Figure 2.3), mentors' behaviour were grouped into three categories, namely, psychosocial support, facilitating learning, and professional development (Figure 2.3, Appendices 1-3). In the psychosocial dimension, two subcategories were included, relationship building, and support and encouragement, which were supported by 37 codes; facilitating learning consisted of three subcategories, with 37 codes: assessment and feedback giving, teaching and guiding, and planning and learning objectives. The professional development dimension covered two subcategories: role modelling and professional socialization and clinical competency building, with 30 codes. All of the codes were listed, followed by their numbered reference in the brackets and all of the 46 references were presented.

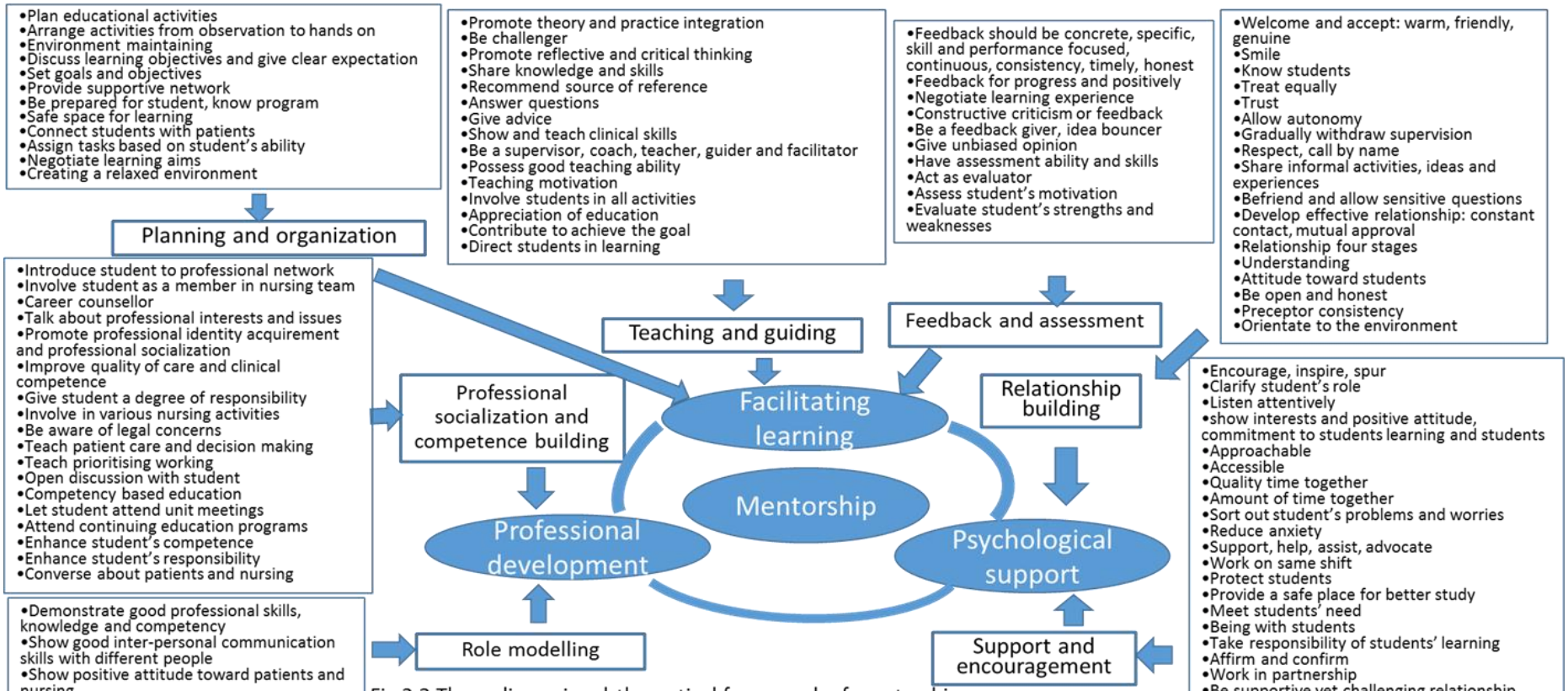


Fig 2.3 Three-dimensional theoretical framework of mentorship

1. Andrews and Chilton, 2000
2. Hall, 1997
3. Hall, 1998
4. Berk et al, 2005
5. Heffernan et al., 2009
6. Spouse, 2001
7. Cahill, 1996
8. Suen & Chow, 2001
9. Chow & Suen, 2001
10. Darling, 1984
11. Dibert& Goldenberg, 1995
12. Charleston & Happell, 2006
13. Elcigil and Sari, 2008
14. Gray and Smith, 2000
15. Smedley et al 2010
16. Charleston&Happell,2005
17. Hallin and Danielson, 2009
18. Luhanga et al, 2010
19. Cele et al., 2002
20. Jakubik, 2012
21. Jokelainen et al, 2013,
22. Knox and Mogan , 1985
23. Charleston and Goodwin, 2004
24. Löfmark et al, 2012
25. Lee et al, 2009
26. Andrews & Wallis 1999
27. Meno et al, 2003
28. Jokelainen et al, 2011
29. Myall et al, 2008,
30. Neary, 2000
31-32. Saarikoski et al, 2002; 2005
33. Bos et, al 2012
34. Kim,2007
35. Warne et al, 2010
36. Weng et al, 2010
37. Lo, 2002
38. Webb and Shakespeare, 2008
39-40. Winstanley and White, 2003; 2011
41. Severinsson and Sand, 2010
42, Price, 2006
43. Mårtensson et al,2012
44. Öhrling and Hallberg, 2000a
45. Öhrling and Hallberg, 2000b
46. Öhrling and Hallberg, 2000c

2.5.4 Discussion

Psychosocial support

Clinical learning is stressful for nursing students, for example the shock at the reality of the nursing world, facing death and illness, relationship with mentors about evaluation, learning requirements and so on (Watson et al., 2008; 2013). So psychosocial support is imperative and in this category, the following subcategories are included:

- Establishment of relationships
- Support and encouragement

The one-to-one teaching strategy is based on relationship: establishing and maintaining an effective mentoring relationship is recognized as an essential factor in clinical learning. The Manchester Clinical Supervision Scale (MCSS) and the Clinical learning environment and supervision scale (CLES) + nursing teacher (CLES+NT) both have the relationship dimension and studies show that is the most important factor in their variables (Saarikoski and Leino-Kilpi, 2002; Saarikoski et al., 2008; Winstanley and White, 2003). For nursing students, the clinical learning mentoring relationship lasts for several weeks, and the stages go through knowing students, building trust, and letting go (Haitana and Bland, 2011). And mentors can know students through working together for three to four shifts (Haitana and Bland, 2011). This relationship can be established and maintained through proper behaviour such as showing acceptance and welcoming, knowing and understanding and sharing informal activities such as having a cup of coffee. When students make progress, mentors can withdraw supervision gradually and allow students more autonomy and independent skill practice.

To be supportive, mentors should first treat students as learners, instead of an extra pair of hands; studies show that nursing students are not satisfied by being treated as human

resources on wards (Andrews and Chilton, 2000; Cahill, 1996), so supernumerary status and quality learning time with mentors should be guaranteed. Also mentors should solve problems students meet and help them to adjust to clinical learning and nursing role. Encouragement is one of the most valuable components to enable learning. It is important to instil confidence, eliminate doubt and promote professional growth, for example any progress made by students should be recognized and confirmed and any contributions of students to the nursing team and patient care should be valued by mentors.

Facilitating learning

Learning is the most important task for students, and numerous studies have identified facilitating learning as the main role of nursing students' mentors (Neary, 2000; Jokelainen et al., 2011; 2013; Myall et al., 2008; Altmann, 2006; Atkins and Williams, 1995). The following points are included in this category:

- Planning and organizing
- Teaching and guiding
- Feedback and assessment.

Good planning and organisation is a very critical aspect highlighted by many researchers and nursing students (Suen and Chow, 2001; Löfmark et al., 2012; Jokelainen et al., 2011; Neary, 2000), but some mentors did not recognize this role (Watson, 1999). Learning can take place in variety of activities. Experiential learning theory shows that observing and doing are necessary experiences for meaningful learning (Kolb, 1984). Particularly, during the first day in clinical placement, more activities should be arranged (Jokelainen et al., 2013). Arranging activities, including various degrees of involvement from observation to hands-on according to students'

competence, is highlighted by nursing students, and students value the chance to see new things and try new skills under supervision (Peirce, 1991).

As part of planning and organization, establishing and maintaining a conducive learning environment should be never underestimated. According to CLES + NT (Saarikoski et al., 2008), the clinical learning environment includes the following factors: ward managers' leadership style, pedagogic atmosphere on the ward, nursing premises on the ward, supervision relationship and nursing teachers' role. It looks as though mentors can do little to manage such a complicated environment, but in the mini-scale environment surrounding students, a caring atmosphere and relationship network among health care professionals and nursing school teaching staff can be created and maintained through mentors' effort.

Teaching and guiding are activities that mentors should inherently have (Darling, 1984; Meno et al., 2003; Gray and Smith, 2000). Teaching, coaching, and guiding are the roles that students thought their mentors performed most frequently in clinical teaching and supervision. To be effective in teaching and guiding, mentors should listen carefully and actively, ask questions to encourage students to think critically and deeply or in different ways (Myrick and Yonge, 2002; 2004). Giving explanations of clinical care and asking about the rationale of skills and techniques can promote theory and practice integration, which is a hot topic and hard issue in nursing teaching and learning (Landmark et al., 2003). Reflection is another learning strategy recommended to nursing staff and nursing students. Mentors need to allocate time to promote students' reflective thinking and learning through discussion and asking questions to guide them thinking about what they have done and experienced (Landmark et al., 2003; Bray and Nettleton, 2007; Severinsson and Sand, 2010).

Feedback is a factor emphasized by nearly all researchers and students in mentoring. It is a way of reminding students of their weaknesses and even mistakes, and of encouragement. And this is the very advantage of mentors offering instant feedback in the real nursing world (Myrick and Yonge, 2004; Udlis, 2008). But feedback from mentors is sometimes problematic for not being concrete, specific, consistent, or timely; and sometimes feedback is not used to develop skills and target progress but for fault finding and personality judgement (Cahill, 1996; Elcigil and Sari, 2008).

Assessment is carried out to various extents by mentors in different countries, most of which are formative assessments (Altmann, 2006). With regard to the UK, the situation is unique: mentors and signoff mentors are responsible for the fitness for practice of the nursing students and they are involved in summative assessment. The assessment role has been greatly debated for many years in the UK, and finally, it was universally accepted and established as an NMC standard (Nursing and Midwifery Council, 2008). This is one of the most difficult jobs for mentors due to the fact that objective measurement of students' competence is difficult and challenging for lack of pedagogic training (Moseley and Davies, 2008; Omansky, 2010), and because of conflicts and confusion about the dual roles of being a mentor and an evaluator simultaneously (Bray and Nettleton, 2007). To foster mentors' assessing capability and confidence, preparatory programs for mentors are usually launched.

Professional development

This dimension includes the following subcategories**:**

- Professional socialization and competency building
- Role modelling

In nursing students' clinical mentoring, the function of professional development focuses on nursing students' professional socialization and nursing care competence improvement. Professional socialization theory claims that professional socialization is a complex process by which a person acquires knowledge, skills, and sense of occupational identity (Cohen, 1981). This process involves internalization of the values and norms of a profession.

The main impact on professional socialization is from clinical learning experience (McKenna et al., 2010; Severinsson and Sand, 2010), which can affect selection of nursing career and re-affirm prior career decisions. During the critical transition phase from student to nurse, mentoring activities have a pivotal role to help students form a positive professional identity. The behaviours of mentors should focus on imparting nursing knowledge, skills training, fostering competence and formulizing professional identity (Myall et al., 2008).

Mentors should try to promote students to think as nurses and to do as nurses by day-to-day work on the same shift. Treating students as nurses and team members, and involving them in various kinds of nursing activities are highly appreciated by students (Jokelainen et al., 2011; 2013; Peirce, 1991). All nursing activities, such as care planning and organization, prioritising, problem solving, decision-making, inter-professional interaction, nurse-patient interaction, need to involve students. Proper tasks should be assigned; some responsibility and more autonomy can be given to them to nurture their confidence and competence.

Mentors as role models play a principal role for students' professionalization and learning (Atkins and Williams, 1995; Watson, 1999). As social learning theory indicates, when a learner observes certain behaviour and the reward for it, they will adopt the behaviour. But sometime, students cannot judge the situation: if constantly exposed to a

bad role model, bad behaviours will also be adopted, therefore good role modelling should be guaranteed (Darling, 1986). On the same work shifts as students, mentors should mind their attitudes and behaviours, and demonstrate professional integrity. This shows students what a good nurse should be like and should do.

Mentors also need to demonstrate evidence-based-practice (EBP) to students, and promote the acquisition of EBP behaviour in nursing students and the adoption of the best evidence to enhance patient care. This is emphasised by the NMC in the UK, but some mentors are quite reluctant to show any interest and capability in nursing research and theory. In reality there are many obstacles to implementing evidence-based practice for nurses and nursing students, such as lack of evidence-based practice knowledge, skill, evidence searching and appraisal competency, leadership support and positive attitude toward EBP (Sandström et al., 2011). So in students' mentoring literature, EBP is NOT frequently mentioned, accepted nor conducted (Smith-Strøm et al., 2012). But as a direction and guidance, it should be included in defining mentoring behaviour, which is beneficial to mentors themselves, patients and the nursing profession in the future.

2.5.5 Summary

Based on the literature review, a mentoring theoretical framework was built, using the three dimensions of psychosocial support, facilitating learning and professional development. This will serve as a foundation for developing a new scale to measure mentors' behaviour in clinical education. However, the relationships between the three dimensions are not explored, while the validation procedures may do that. All these findings were rooted in English language databases, demonstrating more western nurses' perception of mentorship. Therefore to understand the context in China, a

literature review was conducted in the Chinese databases, which is reported in the next section.

2.6 Evidence from China

2.6.1 Introduction

To understand mentorship in the nursing field in China, Chinese literature was searched. The search strategies, terms andthe criteria of inclusion and exclusion are presented in Table 2.5 and Fig 2.4.

2.6.2 Search strategies

Search strategies and terms (nursing OR nurse) AND (mentor OR supervisor OR preceptor OR clinical teacher OR clinical educator) were used. Reference list was inspected and relevant papers were included.

2.6.3Chinese Databases

- China Biology Medicine disc, (CBMdisc)
- Biology Medicine disc, (CBMdisc)
- VIP China Biology Medicine

2.6.4 Inclusion criteria

A. Articles about the development and validation of mentoring scales

B. Subjects are staff nurses and nursing students

C.Articles should show proper items and dimensions.

2.6.5 Exclusion criteria

A. Papers are not related to mentorship.

B. Mentorship about patients, midwives, medical students， nursing assistants, academic staff, or any other groups of people

C. Studies about specific competence of mentor/preceptor/supervisor required (teaching, ethic, leadership, assessment)

D. Internet mentoring program, group mentoring and peer mentoring

E. Studies about scale development showing no proper items or dimensions

F. Full text not accessible

Table 2.5 Criteria and search strategies used in literature review

Criterion 1: limiters	Published Date: 01/01/1980-31/08/2013
Criterion 2: Terms / concepts / keywords	(Nursing OR nurse) AND (mentor OR supervisor OR preceptor OR clinical teacher OR clinical educator)
Criterion 3: Content	Papers related to mentoring scales development and validation were included, while three kinds of papers listed below are excluded: A. Studies not about mentorship but leadership, nursing role, patients care and so on B. E-mentoring, group mentoring/supervision, peer mentoring C. Studies about particular functions of mentorship such as turn-over decreasing, job satisfaction, stress reduction, EBP and teaching
Criterion 4: Subject	Pre-registration nursing students and nurse mentors are included while studies about mentoring other participants such as patients, nurse, faculty staff are excluded.
Criterion 5: Scale review	Studies about scale development showing no proper items or dimensionswere excluded
Criterion 6: Accessibility	Likelihood of availability, time and budget constraint

2.6.6 Results

Six articles related to five questionnaires were located (Hou et al., 2005; 2011; Xu and Kang, 2010; Ma and Tan, 2010; Ye et al., 2013; Hu et al., 2011). The results and systematic review process are presented below in Table 2.6 and Figure 2.4.

Figure 2.4 Flow diagram of the systematic review in Chinese database

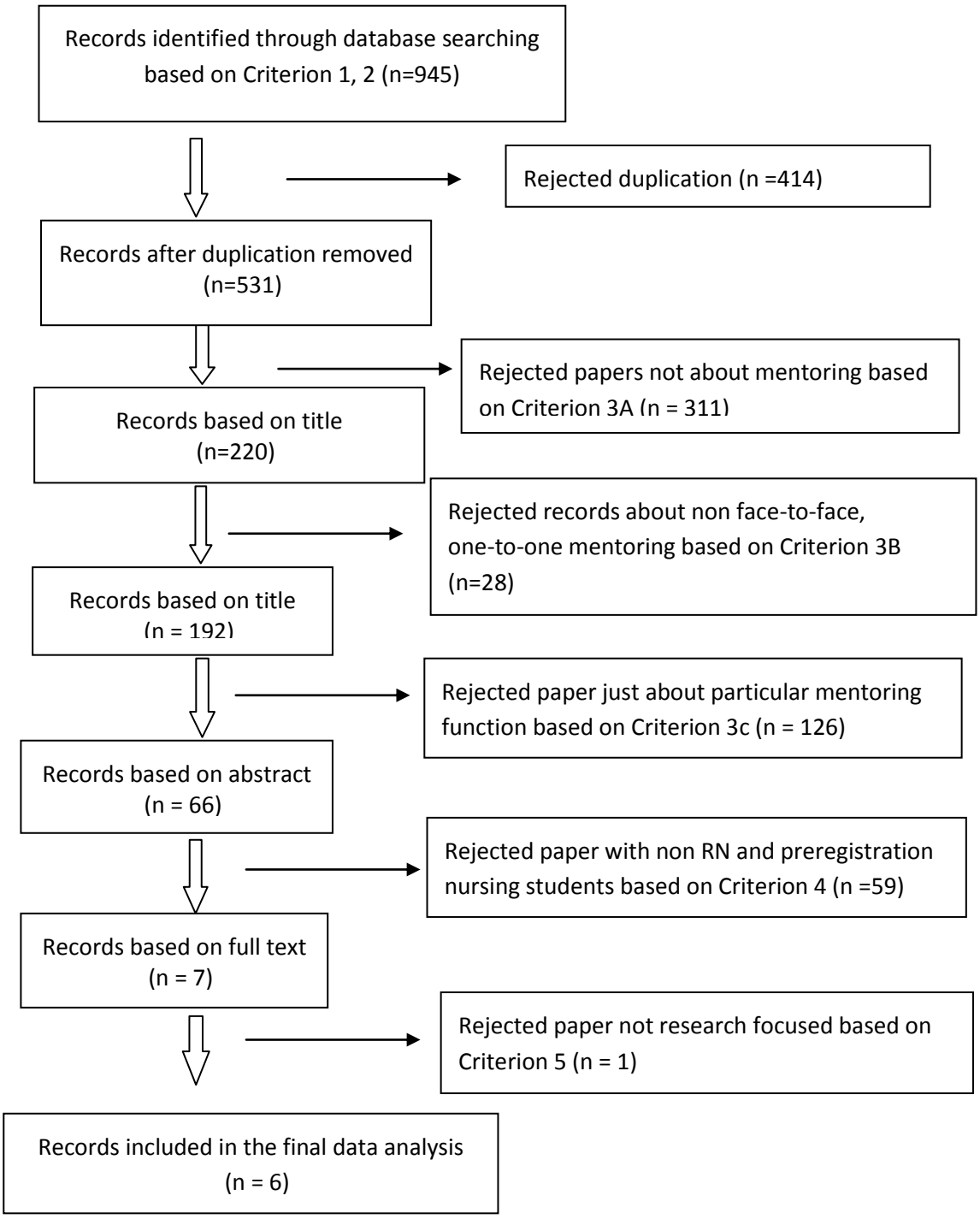


Table 2.6. Mentoring scales in nursing in China

Reference (authors)	Title of Scale and number of items	Subscale/ conceptualization/content	Psychometric properties	Comments (theoretical framework, target and use)
Hou et al. (2005; 2011)	Clinical Teachers' Core Competency Inventory (26)	Four factors: leadership competency, problem-solving, teaching competency, and clinical practice competency.	ICR, EFA	No claimed theoretical framework;showing some psychometric evidence; aimed at measuring mentors' core competence; no further test or application
Xu and Kang, 2010	Scale of students mentoring (9)	Four factors: nursing attitudes toward patients, role modelling, communication competency, nursing competency (nursing knowledge and skills) and teaching ability	No reported psychometric properties	No claimed theoretical framework; showing no psychometric evidence; aimed at measuring mentors' core competence; no further test or application.
Ma and Tan, 2010	Instrument of mentors' ability (26)	Sixfactors: teaching, leadership, innovative, mental health support, communication and research ability	Content validity, EFA, and ICR	No claimed theoretical framework; showing some psychometric evidence; aimed at measuring mentors' core competence; no further test or application
Hu et al., 2011	Nursing Students Clinical Learning Environment Scale (30)	Six subscales: teaching ability; mentor's qualification; interpersonal relationship; ward working; atmosphere, organization and support of learning; and learning opportunity	ICR, EFA and CFA	No claimed theoretical framework; showing some psychometric evidence; aimed at measuring mentors' core competence; no further test or application
Ye et al., 2013	Mentors' competency scale (32)	Eight subscales: teaching competency, teaching qualification, practice competency, nursing research competency, communication, critical thinking, organization and management, and personality,.	Showing content validity and ICR	No claimed theoretical framework; showing some psychometric evidence; aimed at measuring mentors' core competence; no further test or application
ICR: internal consistency reliability EFA: exploratory factor analysis CFA: confirmatory factor analysis				

2.6.7 Discussion and critiques of mentoring scales from Chinese database

Five scales reported by six articles about nursing students mentoring were identified. Staff mentoring has been tried in a new nursing leader development program (Wang et al., 2011), and a new nurse development program (Huang and Wu, 1999) was opened. No scales of measuring staff mentorship are identified yet. All the scales developed in China are new compared to those from overseas, which may imply that mentorship is younger in nursing in China. None of them is considered as suitable to measure mentors' behaviour in China for variety of reasons presented bellow.

Theoretical framework and content

Among the self-developed tools, the items were usually generated from literature review; no theoretical frameworks were presented (Hou et al. 2005; 2011; Ma and Tan, 2010;Xu and Kang, 2010;Hu et al., 2011; Ye et al., 2013), which may imply that the concept under development and test is not clear; this is a drawback compared to the trend of theory guiding scale development (Streiner and Norman, 2008); no empirical methods were employed, such as interview and observation, which lacks context base or background information, in turn it may weaken the content validity of the new scale under development. Some studies stated the researchers' personal theoretical analysis and the items were generated from their minds (Hou et al. 2005; 2011). In addition, the content of the scales focused more on the mentor's teaching skills and competency, research competency rather than the actual behaviour of facilitating students' learning (Hou et al. 2005; 2011; Ma and Tan, 2010; Xu and Kang, 2010; Ye et al., 2013). Other roles and functions for instance psychosocial support were not reflected sufficiently from these scales and the one-to-one relationship between mentors and students was not

yet studied systematically. All these show that these scales do not represent the proper concept or cover the full and relevant content of mentor's behaviour.

Psychometrics

Psychometric evidence was not sufficient. Some did not report any psychometric evidence (Xu and Kang, 2010); no advanced test theory was applied; all of them showed no continuous validation and psychometric building up process. This weakens user's confidence, which may explain to some extent why there is no further use of any of the tools in China.

Extent of use

None of the five scales showed further use except the authors' own development and validation. This may due to the weakness discussed above, for example, no theoretical framework, no valid content, and no enough psychometric evidence. In addition, these tools are relatively new; they have not drew much attention from nursing educators or researchers in China or mentors' behaviour and its measurement has not been an hot topic due to nursing education regulatory body has not put much emphasis and resources on the improvement of the quality of clinical nursing education.

In summary, due to the weakness of all the five identified scales in Chinese literature, for example, no theoretical framework, no valid content, no enough psychometric evidence, no any further use, none of them is considered ideal scale to measure mentors' behaviour in China.

2.7 Conclusion

In the first part of this literature review chapter, some useful scales were identified to assess mentors' function and behaviour in the business and education field. Some of these were also used in the nursing field to measure staff nurse and staff nursing teacher mentorship in clinical settings and nursing schools, but none was used in nursing student mentoring for invalid content and conceptualization. In the second part, the systematic literature review did not generate a very applicable scale to measure registered nurse mentors' behaviour when they mentor nursing students, for lack of psychometric properties, or because they were measuring different concepts. From the theoretical exploration in Part Three, three themes of nursing student mentoring emerged: psychosocial support, facilitating learning, and professional development. This is different from the theories in business and general college student mentoring. Chinese databases were searched: no applicable scales have yet been found. Therefore to develop a scale to measure mentors' behaviour became the task of this study and a proposed mentorship theoretical framework has been established as a start. In the following chapters (Chapter 3 and 4) the new scale developing and validating process and results are reported.

2.8 Literature review updating

All the three systematic literature reviews were updated from 1 September 2013 to 30 June 2015 and no new scales about student mentoring in nursing or non-nursing fields were identified.

Chapter 3. Methodology

3.1 Research aims and questions

This research aims to study the conceptualization of mentors' behaviour in the field of clinical nursing education in China and develop and validate a scale based on that conceptualization to measure students' expectations of mentorship and mentors' actual behaviour.

The research questions are:

- What is the conceptualisation of mentors' behaviour in the field of clinical nursing education?
- Is this new scale based on the conceptualization psychometrically sound?

3.2 Ethical considerations

Ethical approval was granted by the research ethics committee of the Faculty of Health and Social care at the University of Hull, after which permission for data collection was obtained from two universities and three hospitals in China (Appendices 4-9). The main ethical considerations are presented below.

First, for online focus group discussion, before data collection, a research information sheet (Appendix 10 in Chinese and Appendix 11 in English) and cover letter (Appendix 12 in Chinese and Appendix 13 in English) were sent out in which the confidentiality of all participants was assured and an assurance given that pseudonyms would be used instead of their real names. To all students or mentors showing interest in this study, informed consent forms (Appendix 14 in Chinese and Appendix 15 in English) were sent; then these were signed and returned. In the data collection process, a chat room was established just for the focus group which only participants with QQ (a Chinese

online messaging service) ID and a password could access, so privacy was maintained. The conversation online was copied into a word document which was saved on a personal computer and the 'conversation' was then deleted from the chat room.

For the expert panel and online survey, no personal data were collected. Participants received the research information sheet and cover letter, a consent statement (Appendices 16-17) that included terms of participation (e.g. participation is voluntary, data cannot be withdrawn later as anonymous, participation will be considered as consent). For the face-to-face survey, this change in data collection was submitted to the chair of the Research Ethics Committee and permission was obtained.

All data were stored in a personal computer with password protection and also in a USB with encryption. Only supervisors and the researcher could access them.

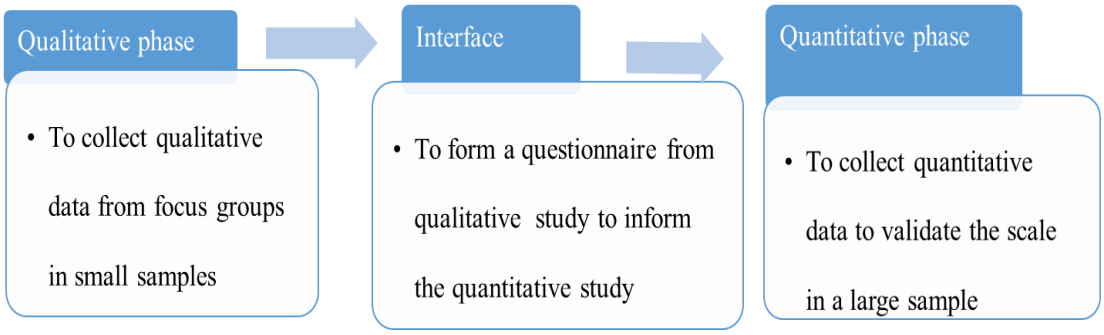
3.3 Methodology

3.3.1 Introduction

Methodology is the general approach taken to address a research question and it is influenced by paradigm and linked to concrete methods. In this research, mixed methodology (combining focus groups in scale development and a survey in scale validation) was used to answer the two research questions with an exploratory sequential design (Creswell and Planoclark, 2011), shown in Figure 3.1, which is explained later. The paradigms in this study are plural; post-positivism and constructivism are compatible; an understanding of students' experience, perception and conceptualization of mentorship in China could be constructed through deep interaction with students and mentors there; while positivism can guide the study to explore if the

themes or perceptions in a small sample can be generalized to a large population and if the tool is suitable to measure students' expectations and mentors' behaviour.

Figure 3.1 Exploratory sequential design

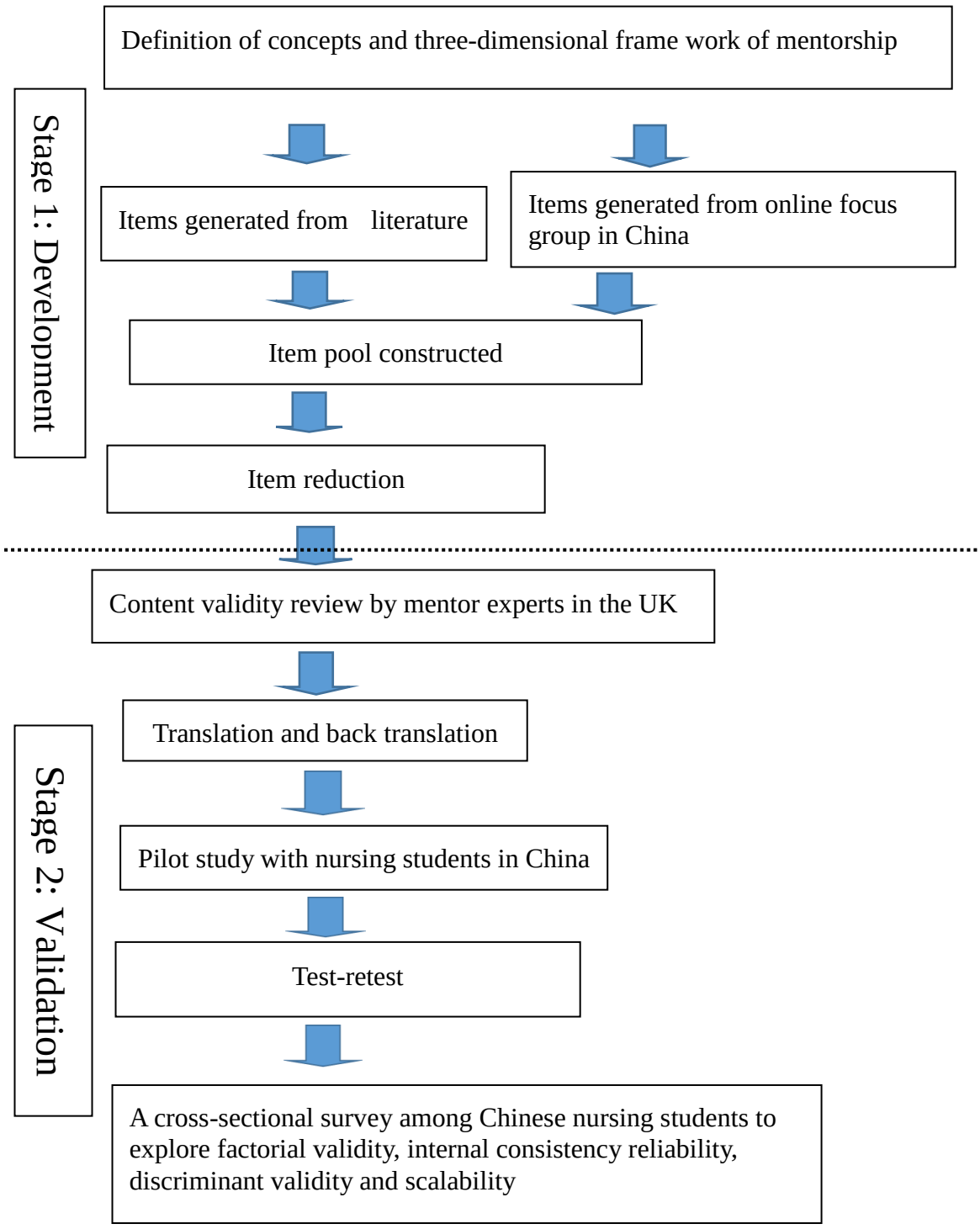


Mixed methods design emerged in the late 1980s and flourished later, as it contains the advantages and offsets weakness of quantitative and qualitative methodology, showing deep understanding of research questions and wide generalization of research results. To apply mixed methodology, four things need to be decided: the level of interaction between the quantitative and qualitative studies (interactive or independent); the priority of the two parts (which part gets more weight, quantitative or qualitative, or whether they should be equal); the timing (sequential or concurrent); and where and how to mix, i.e. the interface, which can take place during interpretation, data analysis, data collection and at the level of design (Creswell and Clark, 2011). With respect to this study, the quantitative and qualitative studies were interactive, as the qualitative data from focus groups were analysed, the results forming a questionnaire which informed quantitative data collection in the later survey for the questionnaire validation. The priority was given to quantitative study, for the ultimate aim of the study was to test if

the new scale is psychometrically sound to assess mentors' behaviour and students' expectations, applying different statistical strategies, instead of deep understanding of students' experiences and perceptions. The timing was sequential, qualitative first and quantitative later, as the results from the qualitative study guided the quantitative data collection. The interaction took place in data collection; after collection and analysing qualitative data from the focus groups, a questionnaire was developed and applied to collect quantitative data in the survey.

With regard to specific methods, nursing student and mentor focus groups (synchronous online focus group and asynchronous online focus groups) in China, expert panel review from the UK, and cross-sectional surveys (including online survey and hard copy survey) were conducted to collect data; as to data analysis methods, both quantitative methods (statistical strategies) and qualitative methods (thematic analysis) were conducted; measurement test theories, including classic test theory and item response theory were used to investigate the psychometrics of the new scale such as test-retest reliability, internal consistency reliability, content validity, factorial validity, discriminant validity and scalability. Descriptive and inferential statistical strategies, such as exploratory factor analysis, confirmatory factor analysis and Mokken scale analysis, were used to analyse quantitative data. Statistical software, such as SPSS22.0, AMOS 22.0 and R were employed. All the specific methods and steps of the new scale developing and validating are shown in Figure 3.2 and will be discussed later.

Figure 3.2 Flow chart of scale development and validation process



3.3.2 Item pool construction through literature review

Items of a new scale can come from many sources, for instance, literature review, existing scales, individual interview, focus group and expert opinion. For this study items were generated mainly from two sources: literature and focus groups.

The mentoring scales in the business, education and the nursing field can provide relevant items and some scales have shown rigorous psychometric properties, so it is justifiable to select some items from existing scales (Streiner and Norman, 2008). But this study excluded any scales from the business or education fields, as those mentoring scales are based on different conceptualization, which is manifested from the critical analysis of literature above (Chapter 2). At the same time, if there are some relevant and suitable items, they should have been imported into nursing scales already, such as shown in Weng et al's research (2010).

The selection of items was shaped by the purported three-dimensional mentoring framework (Chapter 2), which is open to change according to the real context in China. And this implies that other empirical procedures such as focus group and expert panel are necessary to provide more content validity.

3.3.3 Item pool construction through focus group

Introduction

It is not considered appropriate to apply the preliminary item pool mainly based on western countries' literature to assess mentors' behaviour in China, for nursing students and mentors there may conceptualize mentorship differently to some extent. Thus focus group discussion, a popular and rigorous strategy, was conducted among nursing students and mentors respectively to improve face and content validity in China.

Online text-based focus group

Focus groups are commonly used to collect data for qualitative research through exploring participants' experiences, perspectives, and values (Krueger and Casey, 2009; Morgan, 1997; Morgan et al., 1997), and they are also widely used to generate items of scales and test items for clarity, relevance and coverage in business, marketing, social science, and medical science (Rose, 2003; Streiner and Norman, 2008; Vogt et al., 2004). The rationale for using focus group discussion is that it may give a scale more content validity. But for feasibility considerations, only online focus groups were used in this study..

The online focus groups were text-based as video or camera is not always available to or used by students. The other reason is that bandwidth can be a potential problem as video and image information can exert a great burden on unstable information transmission among a group of people. To conduct text-based online focus groups, there are two forms of discussion: synchronous and asynchronous discussion. Synchronous discussion requires good typing speed and careful organizing of people but it can provide instant interaction and fast response, while asynchronous discussion allows participants more freedom and convenience and can provide deep reflective points but it usually lasts a long time, sometimes several months. Both types of discussion were used in this research depending on the situation of data saturation and the participant's real life.

The online focus group is supported by a communication software QQ, which is developed and used by Chinese. It has variety of tools and functions, supporting text, video and audio real-time communication. Such as QQ group (chat room) is often used to communicate within groups and individuals, which is similar to Skype. It can be set

up for a specific purpose. Access to it must be permitted by a gatekeeper. QQ has Qzone tool as well, supporting blogs, and it gives individuasl the authority to decide whom are allowed to read and give comments to his/her blogs. In the university where the researcher worked there was no university email system, QQ groups were used to send public messages, and usually each class has its own QQ group. In this study QQ chat room was used to disseminate the research information, recruit potential respondents for online focus group and online survey and collect data from text-based synchronous focus group discussion. QQ blog was applied to asynchronous focus group discussion.

Group size

The recommended size for a face-to-face focus group is five-thirteen (Matthews and Ross, 2010), or six-twelve (Kenny, 2005), six-eight is more precise and appropriate (Krueger and Casey, 2009; Morgan et al., 1997), but for synchronous online focus group three-five will be more practical (Salmons, 2011), as synchronous discussion is fast, fewer people will be easier to moderate; also a relatively deep and comprehensive discussion is possible. While in a small group (as few as three participants) it will be relatively difficult to encourage discussion, and it will be particularly vulnerable if participants do not turn up, or someone is too shy to share their opinion. The size of asynchronous discussion is not studied adequately to give a precise recommendation. The size of six-eight is acknowledged to be too small (Murray, 1997) to maintain interaction and obtain rich data. With regard to this research, the sizes were over ten.

Number of groups

The number of groups is mainly decided by the data saturation requirement, but usually three to five groups are needed to assure data saturation of a particular topic (Morgan, 1997). It may also influenced by group composition: the more diverse groups are, the

fewer groups are needed. In this study, data from focus groups were analysed after each group interview to check if new themes emerged: if they did, more groups were recruited.

Composition of groups

With regard to this research, both mentors and students were included as they are mentoring providers and consumers, respectively, and they were more likely to have the most authentic and rich experience. But people such as nursing educators in higher education institutions (HEIs) were not included as they might not provide information as important and relevant as mentors and students. In addition, the limited time and resources of this research would not allow that. For the student focus groups, the year they studied, their gender, learning program and clinical learning hospital were considered, while for mentor focus groups, their mentoring experience and the hospital type they were working in were considered, as all these factors may enhance the diversity and richness of data. Hospital type mainly refers to teaching hospital or non-teaching hospital. A teaching hospital affiliated to a medical university usually has a better teaching atmosphere and mentors are more experienced as theoretical and clinical teaching is a part of their work besides clinical job. Non-teaching hospitals not affiliated to a medical university accept students as nursing education is expanding rapidly; the mentors there often have less education experience as they are seldom involved in class and bedside teaching, particularly class teaching. Eleven mentors were from teaching hospitals and nine from non-teaching hospitals in this study.

Data analysis of focus groups

In this study, the data were text-based, therefore they were recorded accurately and no transcription was needed. The level of analysis was set at descriptive level instead of

interpretative or recommendatory as recommended by Vogt et al (2004), for the focus groups aimed to test and generate items, not to get deep understanding of experience or solve problems of mentoring students were facing at that point, or generate any theories. All the analysis was done by hand. One reason was that most of the qualitative analysis softwares do not support Chinese; the other reason was that the qualitative nature of the data would be saved better.

There are many methods used to analyse qualitative research data, such as grounded theory, phenomenology analysis, content analysis, thematic analysis and discourse analysis. Unfortunately there is no specific method for focus groups, except some strategies introduced by Kruegu and Casey (2009), but they did not focus on instrument development (item generation). The current study used thematic analysis. This is a flexible qualitative data analysis method which should be the first step for a qualitative researcher (Braun and Clarke, 2006). It includes six steps. First all the materials must be read several times until familiarity with the information is reached. Then the text is coded line by line, after which themes should be searched for among the codes. Following that, themes will be reviewed, defined and named, and finally a report of analysis will be prepared.

The results of the focus groups were checked to make sure if the themes and items found in literature review were confirmed or needed to be refined or discarded: more importantly, it may provide some new themes beyond the literature. Based on these new themes, new items were generated.

3.3.4 Item reduction and phrasing through supervision team discussion

After the item pool was established, the wording and reduction of items were carried out within the supervision team. Both the literature and suggestions of the focus groups

included more than enough items in case of poor content validity but too many items will give respondents more pressure and cause bias in response, so items were inspected for redundancy.

With regard to wording, there are some debates about positive or negative phrasing of items. Streiner and Norman (2008 p.82) claimed that all of the items should be positively worded and negative wording is harmful to reliability. For this study, positive wording was used to increase content validity and face validity. After this, a series of steps were carried out to investigate the psychometrics.

3.3.5 Psychometrics

Psychometric evidence is key to judging the quality of a scale. In this study rigorous procedures were taken to explore the psychometric properties of this new tool. But the process of psychometrics building is endless, particularly establishing construct validity is a continuous process (Streiner and Norman, 2008). In this study content validity, factorial validity, test-retest reliability, internal consistency reliability and discriminant validity were examined. All methods of data collection and data analysis in the psychometric exploration will be discussed in the following sections.

With regard to validity, there is no gold standard to measure the behaviour of nursing students' mentors currently (based on Chapter 2), criterion validity cannot be established. Convergent and divergent validity testing need other tools but due to the limited time and resource of the current study, they will not be conducted. Neither can the ability of a scale to tell the change after intervention be measured.

In the current study, only test-retest and internal consistency reliability were tested to measure reliability of the new scale. As a mentor has usually one student at a time, it is

not applicable to ask two or more students to rate the same mentor to test the inter-rater reliability; even the same mentor mentors several students at different periods of time in one year and long duration can cause memory bias.

Scalability, besides reliability and validity was explored as well. Scalability is defined as 'the extent to which individual items in a scale measure the latent trait that is being measured and do so distinctly from other items in the scale' (Bannigan and Watson, 2009 p.3240). To do so, IRT, specifically Mokken Scale analysis (MSA), was used, for IRT is taken as a more advanced test theory and getting more and more popular in scale development and validation. It can be used to test other properties of a scale such as hierarchical ordering of items and item scalability (Watson et al., 2012; Bannigan and Watson, 2009). The detail of the scalability investigation using MSA will be reported later in this chapter.

3.3.6 Experts panel for content validity review

Introduction

The first step of validation is to test the content validity, which is defined as 'the degree to which an instrument has an appropriate sample of items for the construct being measured' (Polit and Beck, 2004 p.423) and it is the most important parameter for assessing a scale's quality. It mainly deals with two objectives: coverage and relevance (Hogan and Cannon, 2003). In this section types of content validity index (CVI), the rationale for using British nursing mentor experts rather than Chinese experts to review the new scale, calculation methods and criteria of CVI are discussed.

Expert selection and recruitment

The content validity was tested using mentor experts from the UK. Because there is no formal training for mentors in China, there will be a lack of Chinese mentoring experts,

while the UK leads student mentoring, so it is appropriate to choose UK mentoring experts. Selection was based on their experience and expertise as nurse mentors.

The recommended minimum size of an expert panel is 3, but 10 is recommended for a reliable outcome (Hyrkäs et al., 2003; Lynn, 1986; Polit and Beck, 2006). Taking response rate into consideration, 12 experts were invited. They rated each item for relevance to mentors' behaviour using the labels of 'not relevant', 'somewhat relevant', 'quite relevant', 'highly relevant' at four levels (1-4) which can prevent a neutral response from raters; and these are the most commonly used labels (Polit and Beck, 2006).

The 12 experts were invited by a colleague in the University of Hull. They received an email including the research information sheet, cover letter, informed consent statement and the new scale, after they had agreed to participate in this research. Reminders were sent out to manage the response rate.

Measurement of content validity

Content validity can be tested through subjective judgment and can also be quantified using CVI and other methods such as average congruence percentage (ACP); among nurse studies, CVI is the most frequently used method (Lynn, 1986; Polit and Beck, 2006). CVI means the extent of all items rated as relevant by all experts panel. And it can be tested on two levels: item level denoted by the item content validity index (I-CVI) and scale level denoted by the scale content validity index (S-CVI).

The calculation and criteria of CVI are presented below.

The formula for I-CVI is as following:

Formula 1

$$\text{I-CVI} = \frac{\text{Number of raters giving a rating of '3' or '4'}}{\text{Total number of raters}}$$

For fewer than five experts, the value 1 is considered acceptable, while for six raters and more, 0.78 and over is acceptable, otherwise the items should be deleted or reworded (Polit and Beck, 2006).

For S-CVI, there are two types of computing methods. One is universal agreement S-CVI/UA, defined as 'the proportion of items on an instrument that achieved a rating of 3 or 4 by all the content experts' and the other is average agreement S-CVI/Ave, defined as 'average proportion of items rated as 3 or 4 across the various judges': S-CVI/Ave is preferable (Polit and Beck, 2006 p. 492). The acceptable value for S-CVI/UA is 0.4, and for S-CVI/Ave is 0.9 (Polit and Beck, 2006 p. 492).

3.3.7 Test-retest reliability

Test-retest reliability tells the stability of a scale over time which is particularly suitable for enduring traits such as personality. In this research, the scale is measuring students' expectations which is thought to be enduring. Fifty nursing students were recruited, the rationale being that increasing sample size from 50 to 300 will not change the reliability much (Steiner and Norman, 2008), although there is much debate and sample sizes vary widely, from 50 to 1000. One week later the same scales were administered to them again, as this time interval was thought would not cause memory and maturation bias. The intra-class correlation coefficient (ICC) was applied to compute the test-retest reliability, for it has many advantages as described in Chapter 2. The criteria are listed below in Table 3.1 after Landis and Koch (1977).

Table 3.1 Criteria of ICC

Value	Interpretation
<0.20	Poor
0.21-0.40	A fair agreement
0.41 to 0.60	A moderate degree of agreement
0.61 to 0.80	Substantial agreement
0.81 to 1	Almost perfect agreement

3.3.8 Cross-sectional survey

Introduction

In addition to content validity and test-retest reliability, other psychometrics such as construct validity, internal consistency reliability and discriminant validity and scalability were explored using a large sample cross-sectional survey after the pilot study. In this section the sample size consideration, sampling methods, data collection methods and response rate management are discussed.

Sample size

The sample size estimation was mainly based on factor analysis which is the main data analysis strategy for investigating the conceptualization of mentors' behaviour and the factorial validity of the new scale. There are some rules of thumb to judge the impact of sample size on the accuracy of factor analysis, such as the ratio of number of items to participants (ranging from 1:5-10 is good) (Ferguson and Cox, 1993); in this study, there are 47 items, and the sample size should be over 470; the ratio of number of factors to the number of items should be over 1:3-6; and the minimum number should be 200-400, as recommended by Ferguson and Cox (1993); the ratio between subjects and factors should be over 1:6. Basically, the rules indicate that this sample size should be approximately 500 and the actual sample size was 742. More objective methods of calculating sample size adequacy were conducted using Kaiser–Meyer–Olin (KMO) in SPSS 22.0, which is reported in Chapter 4.

Sampling methods

It would be representative to have a randomized nursing student sample from the whole of China as it would allow broader generalizability of this new tool. But for

considerations of feasibility, only convenient sampling was used in the university where the researcher worked in China. This university has over 2000 nursing students from four programs (degree, associate degree, 5-year diploma, and 3-year diploma) learning in different hospitals at their final year of study, which can provide a sufficient sample. Students from all programs were recruited and this may give a fair representation of Chinese nursing students as these programs are commonly provided nationally, and it is also suitable according to the operational definition of nursing students in the following section, the difference of the four programs are presented in table 3.2.

Table 3.2 Differences of four programs

Program	Enrollment requirement	Year of study	Depth and width of curricula	Employment
Degree	12 years basic education and national university entrance examination, higher scores are needed than associate degree	Four years	Widest and deepest content of theoretical curricula are provided, ; 10-12 months clinical study	Employed by tertiary hospitals, most stable jobs
Associate degree	12 years basic education and national university entrance examination	Three years	Less wide and deep content; 10-12 months clinical study	Employed by tertiary or secondary hospitals, stable jobs
3-year diploma	9 years basic education, provincial level test; lower scores are needed than 5-year program	Three years	Lest wide and deep content; 10-12 months clinical study	Employed by small hospitals or clinics, less stable jobs
5-year diploma	9 years basic education, provincial level test,	Five years	Less wide and deep content; 12-15 month clinical study	Employed by tertiary or secondary hospitals, stable jobs

Data collection

The main cross-sectional survey was launched after the pilot study. Students were asked to rate the importance of each behaviour of mentors towards their successful learning, using five options from ‘not important at all’ to ‘quite important’ and to assess their mentors’ actual behaviour with five options from ‘strongly disagree’ to ‘strongly agree’, as this study is testing the new scale’s functions of assessing students’ expectations from mentors, and measuring mentors’ real behaviour, and also because exploratory and confirmatory factor analysis need different data sets ideally. Other information, such as program, age, gender, days of study, mentorship type experienced and satisfaction with mentorship were also collected.

Questionnaires administration and response rate management

Questionnaires were administered to nursing students through the Bristol Online Survey tool because students in clinical learning are usually scattered in different hospitals and even in the same hospital they will work on different shifts, on different wards and live in different places. Obviously, it is hard to summon them together to collect the data in the process of clinical learning. One advantage of online administration is that it saves money and time. At the same time, nursing students are skilled with computers, and the internet is also widely accessible via PC or mobile phone.

One month was given to finish the online questionnaire, after that the online survey was closed and no respondents can access more. When participants intended to fill the questionnaires they need to complete in once, otherwise the information filled in would lose without submission online. Reminders were sent out to increase response rate in this period, In addition, to guarantee the sample size and response rate, alternative survey

methods were considered. In this study face-to-face hard copy survey was used as the online response rate was low.

3.3.9 Exploratory Factor Analysis

The structure of the hypothetical construct of nursing students mentoring was explored using factorial analysis. Factor analysis has been widely used to test the validity of scales, to develop and test theories, and reduce large number of variables to a parsimonious set of factors since Spearman developed this method in 1904 (Thompson, 2004). It includes exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

EFA is intended to find out the common factors of a latent variable (here, mentoring behaviour) under investigation through analysis of a linear correlation matrix and calculation of the correlation of items to factors and the correlation between factors and a construct. It is applied in situations where a theory is not firmly established, and it tries to use empirical scores from samples to generate common factors and find out potential dimensions. Before running EFA, researchers may or may not have any proposed conceptualization in their mind and the data analysis will not be affected by researchers' perspectives, so it is completely data-driven (Thompson, 2004).

In doing EFA, some key points need to be considered, such as methods of extraction, number of factors to extract, methods of rotation, and the reliability of final construct. All these steps are discussed in the following section.

Methods of extraction

To perform EFA, the first thing is to decide which method should be chosen to do factor extraction, factor analysis or principal component analysis (PCA). There are debates

about the two methods. PAC is more frequently used (Watson and Thompson, 2006; Ferguson and Cox, 1993; Gaskin and Happell, 2014), although it is not a method of factor analysis strictly speaking as it aims to do data reduction, assuming the communality of each item is 1.While factor analysis aims to describe a constructin terms of a smaller number of latent factors through exploring the correlations between items in terms of latent factors; only the shared variance among items is analysed.

PCA has more developed methods to determine the number of factors and can handle large number of items more efficiently than factor analysis and it may have more accuracy in decision of number of factors to extract. But along the development, factor analysis has become more and more popular as it overcomes its shortcomings in comparison to PCA (Gaskin and Happell, 2014).

Usually PCA provides identical solutions to those of factor analysis methods when the number of items is over 30 and communalities are greater than 0.7, but when the number of items is less than 20 and any communalities are under 0.4, differences might occur (Field, 2009). In this condition factor analysis can provide more accurate analysis (Widaman, 1993).

Number of factors to extract

The criteria for deciding how many factors should be extracted include: eigenvalue over 1.0, scree plot, parallel analysis and other methods. Often eigenvalue over 1.0 is used when the number of items is less than 20, but when the number of items is over 30, the second frequently used method, scree plot, may provide more precise decisions than the eigenvalue. When neither method can give a precise recommendation, another method,

parallel analysis, can be used, which is more accurate (Ferguson and Cox, 1993; Gaskin and Happell, 2014). Monte Carlo PCA for parallel analysis software was considered in this study when necessary. The procedure is to use a simulation to generate eigenvalues based on the number of items and the number of respondents, and these eigenvalues are compared with the actual eigenvalues. Where corresponding real eigenvalues are greater than simulated ones that indicates the possible presence of a factor.

Methods of rotation

Rotation of factors aims to get a more parsimonious and interpretable solution than the initial factor extraction. Based on the relationship between factors, there are two categories of rotation methods, orthogonal rotation (not related) and oblique rotation (related). In this study there are no solid established theories about mentorship in the field of clinical nursing education, and the relationships between factors are not clear yet, so both orthogonal rotation and oblique rotation were tried.

Internal consistency reliability of the new construct

After the structure has been identified through EFA, the reliability of the structure needs to be checked, then internal consistency reliability was calculated to investigate the homogeneity of all items in this new scale. Item-scale α, and item-subscale α were investigated and the criterion of α is set as 0.7.

3.3.10 Confirmatory factor analysis

After EFA, putative factors of nursing students mentoring were identified, but the factors identified need further testing to explore if this structure can be confirmed in the other data set. This can be realized through structural equation modelling (SEM) using Analysis of Moment Structures (AMOS).

Structural equation modelling can do exploratory and confirmatory analysis. In this study confirmatory analysis was used. The requirement is that researchers must be very clear about the number of factors, the items associated with each factor, and whether the factors are correlated, so it is theory-driven. The recommended procedures for CFA were conducted sequentially, including model specification, model estimation, equivalent model consideration and model stability check (Kline, 2005).

Model fit was estimated based on the model fit indices. Because there is no one ideal criterion that can be used to evaluate model fit precisely, a variety of indices have been developed to estimate model fit from different aspects. Kline (2005) recommended the minimal set of fit indices which should be reported, such as chi-square, comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardised root mean squared residual (SRMR). Model modification may be needed based on the model fit index and hypothesised model. In this study the indices and their criteria are presented in Table 3.3

Chi-square test calculates the similarity between the model and the data set, if the data set fits the model, the chi-square value should be small and the significant value should be over 0.05. But it is influenced by correlation of variables and sample size, which is prone to reject the model as the correlation coefficients increase and the sample size is large, so it is not used often for judgement of model fit. However it is useful in model comparison and model modification (Kline, 2005). CMIN/DF (chi-square/degree of freedom) as an estimation of chi-square considering the influence of degree of freedom is used as well, which should be under 3 (Kline, 2005).

Another very important index which we need to report is the root mean square error of approximation (RMSEA) which takes sample size into account and is less sensitive to

model complexity. About the appropriate value of RMSEA, there is some inconsistency. Browne et al (1993) recommends that RMSEA <0.05 implies close approximate fit and 0.05 to 0.08 shows reasonable fit, as also does Kline (2005). Its 90% confidence interval is usually reported together. The third category is the comparative fit index (CFI), which should be 0.95 and above. The final index is the standardised root mean squared residual (SRMR) which measures the difference between the residuals of the sample covariance matrix and the hypothesised covariance model. The overall difference should ideally be 0, but below 0.08 is considered acceptable (Hu and Bentler, 1999). Other frequently used indices, such as the Goodness of fit index (GFI), should be over 0.9.

Table 3.3 Criteria for model fit in CFA

FIT index	VALUE
CMIN/DF	<3
GFI	>0.9
CFI	>0.9
RMSEA	<0.05
SRMR	<0.08

3.3.11 Mokken scale analysis (MSA)

After the construct of mentors' behaviour was investigated, applying the classical testing theory (CTT), MSA, a non-parametric item response theory (IRT) was used to test the dimensionality and scalability of the new scale as well, and the results from MSA were compared to those of CTT.

CTT and IRT

Classic test theory assumes that: Y=T + e, where Y is the measured score; T is the true score and e is the error. Both true score and error score are unobservable; error is not related to true score; the space in between ordinal variables, e.g. data from a Likert scale is the same, for example the space between 'not important' and 'important' is the same as that between 'very important' and 'extremely important'; and it assumes item equivalence. All of these assumptions are unrealistic (Streiner, 2010). Meanwhile, a

person's position is calculated by adding item scores together, and it is relative to the norm of the sample, so it is sample-dependent. CFA and EFA, as belonging to the family of the CTT, estimate model fit using linear regression which can only find how many factors are in the new scale but cannot investigate which behaviour is more important or more difficult, nor how unique an item would be.

IRT is a mathematical model linking the possibility of response to an item based on personal latent trait to the characteristics of the item (Sijtsma et al., 2008). It treats all the issues in a more realistic manner, acknowledging the difference between items and the difference between spaces, investigating an item's individual performance not dependent on the sample, using other parameters such as difficulty and differentiability instead of true score and error (Streiner, 2010). This overcomes the weakness of CTT. The item response curve (IRC) is the foundation to understand IRT (to avoid confusion between the abbreviations of item characteristic curve and intra-class correlation coefficient, IRC is used in this thesis), as shown in Figure 3.3 (Watson et al., 2012). The horizontal axis represents a latent trait; the vertical axis represents probability (0-1) of response to items; the curve denotes the difficulty of an item by location and difficulty increases towards the right; the slope represents the differentiability of the item: the steeper the slope is, the better differentiability the item has.

IRT can order persons according to their latent traits by means of adding item scores and can order items according to their difficulty or importance on the same scale. It makes assumptions such as unidimensionality and local independence (Watson et al., 2012). IRT includes parametric item response theory (i.e. the Rasch model, one-parameter IRT; there have been two-, three- and four-parameter models hitherto) and non-parametric item response theory (i.e. Mokken scale analysis, MSA). Only MSA was used in this study. The reasons will be explained later.

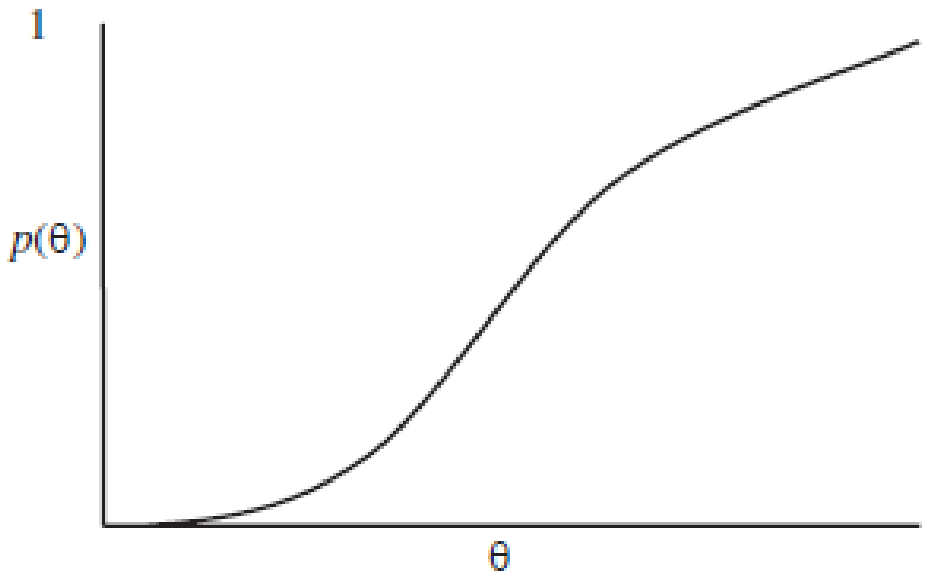


Figure 3.3 Item response curve (Watson et al., 2012)

Guttman scale, MSA and parametric item response theory

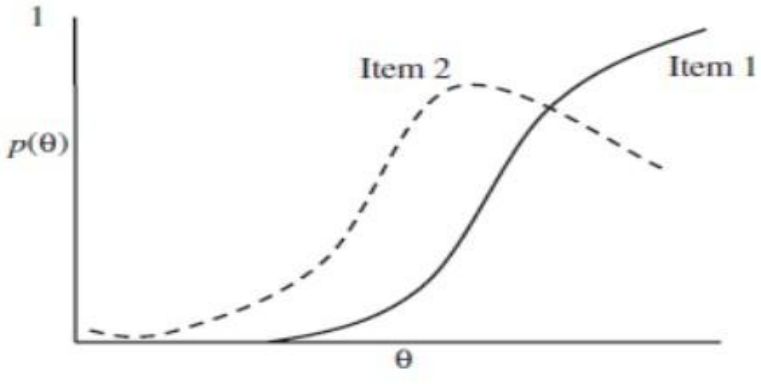
Mokken scale analysis originated from the Guttman scale. The Guttman scale is deterministic: it claims that if respondents respond to certain items positively, they must respond positively to less difficult items (Watson et al., 2012). But for MSA it is stochastic, implying that when a respondent responds positively to one item, the probability of positive response to less difficult items is higher. It also implies that the probability of a respondent with more of a trait responding positively to a given item is higher than that of respondents with less of the trait (Watson et al., 2012). This is more realistic than the Guttman scale.

Parametric item response theory has several models as mentioned before.
One-parameter logistic model (e.g. the Rasch model) investigates items' difficulty and constrains their differentiability, assuming all items have identical slopes. A

two-parameter logistic model releases differentiability (Streiner, 2010), implying the slopes of items can vary. Compared to parametric item response theories like that of Rasch, MSA has no strong assumptions about the data characteristics and no strict requirement on the IRC's ogive shape. So it allows more items to be retained and this will not undermine the reliability of a scale seriously (Stochl et al., 2012; Sijtsma et al., 2008).

Models and assumptions of MSA

Mokken scale analysis can be used to analyse dichotomous items and polytomous items. It has two models, one the monotone homogeneity model (MHM), which means the items' scores increase as the trait increases, with no decreasing (see Figure 3.4 (Watson et al, 2012)). This can order respondents according to their raw accumulated scores. The other model, invariant item ordering (IIO), denotes that all IRCs cannot intersect, which means items can be ordered according to their difficulties and this item ordering is the same for all respondents (Sijtsma and Junker, 1996; Ligtvoet et al., 2010), see Figure 3.5 (Watson et al., 2012), which replaced the double monotonicity model (DMM) applying to dichotomous items only.



.

Figure 3.4 Monotonicity homogeneity model (Watson et al., 2012); Item 1 shows monotonicity; items 2 shows violation of monotonicity by a decreasing curve.

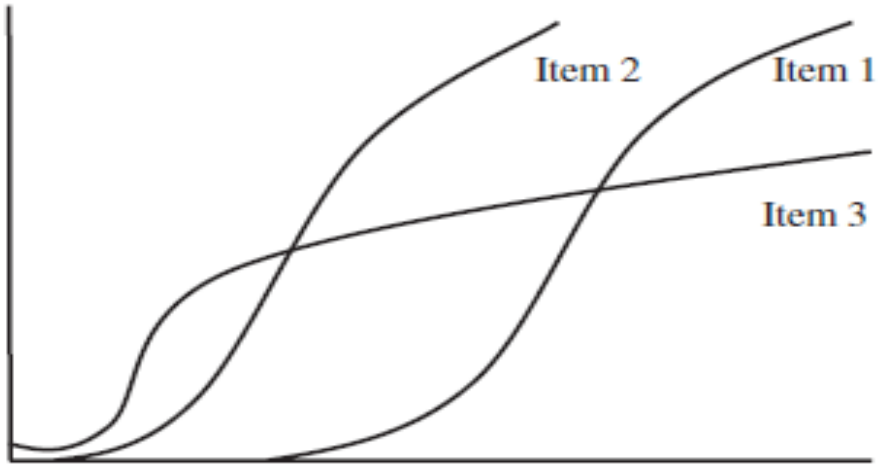


Figure 3.5 Invariant item ordering (IIO, Watson et al., 2012); Item 1 and 2 show IIO; items 3 shows violation of IIO by intersecting with item 1 and 2.

To carry out a Mokken scale analysis, the following assumptions must be checked as they are key for model estimation. For an MHM model three assumptions need to be met. These are unidimensionality, local independence, and monotonicity. For an IIO model, an extra assumption, IRCs interception, should be checked. The models and assumptions are presented in Table 3.4

Unidimensionality of scales means that items in a scale should attempt to measure one construct; those items that do not fit can be identified and deleted, and this can be tested using the automated item selection procedure (AISP). Monotonicity as presented above can be tested using an R procedure: *check. monotonicity,* (R is a free software environment for statistical calculation and graphics. It has been used in this study to carry out Mokken analysis). Local independence means that the respondents' response to one item is not affected by other items but just related to the trait of the respondent; it is usually assumed but not tested, which is not unique to MSA; it manifests as multicollinearity in regression analysis (Field, 2009). Local dependence is usually

caused by item overlapping or ‘chains’ and manifests itself in inflated H_is, much higher than 0.5 (Watson et al., 2014). Scrutinising item content can spot it.

Invariant item ordering (IIO) is more restrictive than monotonicity but more accurate and useful in ordering items according to their difficulty or importance. It can be checked using method: *check.iio*. Items showing high critical values will be removed and scalability coefficient H^T will be checked then.

Table 3.4 Models and assumptions of Mokken scale analysis

Model	MHM	IIO
Assumption and methods	Unidimentionality (method: aisp)	Unidimentionality (method: aisp)
	Local independence	Local independence
	Monotonicity (method: check. monotonicity)	Monotonicity (method: check. monotonicity)
		Non-IRCs intersection (method: check.iio)

Scalability coefficients

Scalability strength can be judged by scalability coefficients such as H_{ij} (item-pair), H_i (item quality) and H_s (total scale quality) for MHM, H^T for IIO. According to Ligtvoet et al. (2010), H_{ij} should be over 0, and the other three coefficients usually should be over 0.3, and the rule of thumb cut-off points are presented in Table 3. 5. The software R can investigate any violation of monotonicity and IIO, and it can also compute scalability coefficients and select items for each subscale.

Table 3.5 Scalability coefficients

Scalability coefficients	Interpretation	Cut-off point
H_{ij} (item-pairs)	Inter-Item correlation should be non-negative to form a Mokken scale.	If MHM holds , H_{ij} >0
H_i (item)	Precision of item discrimination index: shows the strength of the correlation between an item and the latent trait under investigation.	$H_{i,}$ H_s, H^T <0.3, unscalable; 0.3< $H_{i,}$ H_s, H^T <0.4, poor scalability;
H_s (scale)	Hs is a weighted mean of the item coefficients, an index for the precision of ordering person.	0.4< $H_{i,}$ H_s, H^T <0.5, moderate scalability; $H_{i,}$ H_s, H^T >0.5, strong scalability
H^T	Precision of invariant item ordering	

3.3.12 Discriminant validity test using t-test

A valid scale should have the ability to differentiate a trait between extreme groups at least (Streiner and Norman, 2008). In this research, students who experienced one-to-one mentorship can be considered as having proper mentorship, while students who experienced other mentorship such as following shifts (on each shift students following different mentors, no named mentor being responsible for students learning) and group mentoring (a nurse in charge of five-seven students) can be considered as having no mentorship. Theoretically, students who experienced one-to-one mentorship will have more chance to be exposed to mentoring behaviour such as facilitating learning and professional development and psychosocial support, that is to say mentors' behaviour will be rated higher in one-to-one mentorship. If this new tool is valid it should capture the difference between the two groups. And the means of the scores in the two groups were compared using the *t*-test.

3.4 Research steps

3.4.1 Definition of concepts

This scale is focused on mentors' behaviour in the nursing field, which is a complicated latent construct; therefore, the first thing required is to define the concepts around nursing students' mentoring clearly and propose a theoretical framework for mentors' behaviour. This has become a trend of scale development (Streiner and Norman, 2008) and it also functions as a foundation for selection of observable variables (items) to represent the unobservable construct under investigation. 'Mentorship' ('mentor'), 'nursing students' and 'behaviour' are the three key terms for this research project. 'Mentorship' was defined operationally in Chapter 2; in this chapter 'mentor', 'nursing students' and 'behaviour' are defined operationally.

Mentor

For the purposes of this thesis, a mentor is a registered nurse facilitating a student's learning and supporting the student's professional development in clinical placement on a one-to-one, day-to-day basis.

Nursing students

'Nursing students' refers to pre-registered nursing students from any program learning in clinical practice. In this study new graduates were excluded because they have different targets and expectations compared with pre-registered nursing students. Nursing students are learning to be a nurse while graduates are learning the ward culture to be able to adapt better to the new environment and work more efficiently when they have passed registration. In the UK preceptors are assigned to new graduates and it usually last several months. So it is not plausible to combine both populations in this study.

Behaviour

The Online Oxford Dictionary defines behaviour as 'the way in which one acts or conducts oneself, especially towards others' (Oxford Dictionaries, 2015). In this study 'behaviour' was studied using the TACT principles: Target, Action, Context and Time (TACT) (Ajzen, 2002); the Target is nursing students; the Action is mentoring; the Context is in clinical placement; and the Time means being with students.

Also, behaviour can be conscious or unconscious, overt or covert. In order to give raters clear clues to make sound judgments about mentoring behaviour, the behaviour under investigation must be observable. Unobservable actions such as intentions and motivations of behaviour are out of the scope of this study.

3.4.2 Theoretical framework exploring

Based on the systematic literature review and content analysis about nursing students mentoring, a three-dimensional theoretical mentoring framework was proposed, as shown in Figure 2.3 in Chapter 2. It includes facilitating learning, professional development and psychosocial support. Psychosocial support includes establishment of relationships and support and encouragement. Facilitating learning contains planning and organizing, teaching and guiding, plus feedback and assessment. Finally professional development consists of promoting students' professional socialization and role modelling. This served as a basis to select items from the literature review and provided guidelines for focus group interviews.

3.4.3 Item pool construction

First, items were selected from 10 scales (Winstanley & White, 2011, Chow & Suen, 2001 , Löfmark et al, 2012; Hou et al, 2011; Hallin & Danielson, 2010;Saarikoski et al, 2008; Knox & Morgan, 1985; Lee et al, 2009; Berk et al, 2005; Jakubic,2012) identified in Chapter 2, as these scales had some subscales or items closely pertinent to the three-dimensional mentoring framework (see p.75) and/or had good psychometric evidence and/or wide use. In addition, some items were selected from studies about nursing students mentoring and the NMC guideline of supporting learning and assessment (NMC, 2008). The resources of each item are presented in Appendix 19.

After that focus groups were conducted. The rationale for using online focus groups and text-based chatting, focus group composition, group size and data analysis method are reported in the methodology section above, while questions asked in the focus groups are presented in Appendix 18.

In this study synchronous discussion was used in three student groups. As it was challenging for the mentor groups, an asynchronous online focus group was also used later. To conduct synchronous focus group among mentors was problematic due to different reasons. First it was difficult to organise some identical time for all participants to discuss at the same time as they worked on different shifts. Second, even if they managed to participate in the study, they could not concentrate on the discussion as they were distracted by other roles in life, for instance, they answered phone calls, looked after their children or parents. Third, they were relatively slower in typing than nursing students, which made the discussion less efficient and less data were generated. Therefore asynchronous was more suitable as it allowed them more flexibility in time and space.

In the asynchronous discussion, the mentor group included 15 participants, while the student group had 13. Considering the characteristics of students and mentors, the composition of groups and data saturation, six focus groups were conducted, four synchronous groups (three student groups and one mentor group) and two asynchronous groups (one mentor group and one student group). Both male and female students from all programs learning in various hospitals plus mentors from different hospitals with different mentoring experience were included. Based on this, items from literature were confirmed and new items were generated, thus forming an item pool which then needed to be properly phrased and refined.

3.4.4 Item reduction and phrasing

After discussion in the supervision team, it was decided that items should be phrased positively to maintain face validity and motivation to respondents (Steiner and Norman,

2008). As each item should ask about one activity, any item that indicated two or more aspects was separated. And the items thought to be irrelevant or duplicate were dropped.

3.4.5 Experts panel review for content validity

All the invited mentor experts were registered nurses and registered mentors, each with more than 10 years of mentoring experience and with at least a degree education qualification, and all worked locally as mentor supporters. Two weeks after sending them the new scale, a reminder was sent out, for only four experts had given feedback. After that five more experts responded. The content validity index (CVI), including item level and scale level CVI, was calculated according to the review of the nine experts. Although content validity is pivotal and necessary, it is not sufficient alone. More objective and empirical validation procedures like construct validity are needed and that were tested among Chinese nursing students. This English version therefore, had to be translated into Chinese before that.

3.4.6 Translation and back translation

Translation

After the expert review and before this scale went to Chinese nursing students, it was translated properly into Chinese. The items were drafted in English, all the concepts and terms were discussed between the author, a Chinese research student, and her British supervisors to make sure the meaning and concepts are equivalent in both Chinese and English culture.

The scale was translated into Chinese by four bi-linguists from China, one of them being a professor teaching English in one medical university and the other three nurses

and nursing teachers with PhD qualifications. The difference and disparity were discussed to reach a consensus.

Back translation

To check the accuracy of translation, back translation is necessary, which was done by an English professor in the foreign language school in one Chinese university. Then the back translated version was compared with the original English version to see the equivalence. The examination was done by Professor Roger Watson who is a native English speaker, and any problematic items were reworded, translated and checked again.

3.4.7 Pilot study

Before the large scale administration of the questionnaires to nursing students, a small scale pilot study was carried out to evaluate the time needed to complete the questionnaire, the clarity of items, feasibility and response rate of the online survey.

Before piloting, the options for each item were discussed and the decision was made to use a Likert scale, allowing a neutral option. As the scale aims to assess students' expectations from their mentors (importance of mentors' behaviours), the options were: 'quite unimportant, unimportant, not clear, important, and quite important' (1-5), which was thought as ideal – sufficient for respondents to distinguish and good for the reliability of the scale.

This study applied the Bristol Online Survey tool. This is based in the UK, so the accessibility, display and language were checked beforehand. To do this, five friends were invited to open the survey in China. This suggested that there were some problems about accessibility: two opened it on a computer; one opened it with a mobile phone;

the other two could not open it with a mobile phone but then opened it with a computer, where it displayed correctly. Therefore, in the pilot, potential participants were instructed to use computers to complete the questionnaire.

Then research information was advertised in three QQ groups: degree, associate degree and 5-year diploma programs, after being invited to join in by their teachers. In two weeks, 50 students finished the scale and their QQ numbers were obtained, which allowed the researcher to access them again to apply test-retest.

3.4.8 Test-retest

The time interval for conducting the retest was set as 10 days, so as not to cause students' expectations to change and thus avoid memory biasing response. Ten days later, after the first point, all of the 50 students were invited to complete the online questionnaire. In one week 19 completed the retest.

3.4.9 Survey

Data collection methods and response rate management

Pilot participants were not included in the main survey as the pilot was carried out in July in the former year group at the end of their clinical study, while the main survey took place in another year group who were at the beginning of their clinical study. In the first week of online survey, only 20 students completed the questionnaires. More advertising was done in the QQ group of each class from which approximately 900 students could in theory be reached by this advertisement. The possible reasons for the low response rate were investigated; some students claimed it was difficult to access the scale on an overseas website because of their low surfing speed, and they gave up after several unsuccessful attempts; some students did not have enough interest; some did not

use the QQ group often, which implied that this information did not reach them. Then the teachers in charge of each class were invited to text the web link and research information to students via mobile phone, but one week later, the response rate was still low with just another 17 students completing the questionnaire. At last this situation was discussed within the supervision team and a hard copy survey was thought necessary to increase the response rate. This amendment of the survey method was approved by the chair of the faculty ethics committee.

Three face-to-face surveys were organized and students from different programs (degree, associate degree, 5-year diploma, and 3-year diploma) in three hospitals of Luzhou city were invited to complete the questionnaires at the end of a lecture. Twenty minutes later assistants went back to collect the questionnaires.

The online survey continued until the end of August 2014 - 50 days - as many students were studying in different cities so it was not easy to reach them physically.

Data management

For the hard copy survey, after data collection the questionnaires were checked manually to exclude unfinished or/and unengaged questionnaires (0 or very small variance)； then valid questionnaires were entered into Excel manually by three colleagues and the researcher herself; the questionnaires received and the cases entered into data analysis are presented in Figure 4.2 (see p161). They worked in pairs--one person read the numbers, and the other one typed in computer. Then 10% of the data (74 questionnaires) were checked for correctness. The data that the researcher entered had no errors, while the data entered by the other pair was problematic, so their whole data was checked by their colleagues. The online survey required participants to complete all

questions before submission but unengaged cases were also checked using Excel. Then missing data was checked and the missing pattern was analysed.

Data analysis plan

Because questions about the importance (expectation) of mentors' behaviour and the assessment of mentors' behaviour were asked separately, two data sets were set up: the importance data set and the assessment data set. The importance data set (n=669) was used to explore the structure of the new scale and to explore the hierarchical properties, so statistical strategies such as exploratory factor analysis (EFA) and Mokken scale analysis (MSA) were used. Then the structure identified from the importance data was tested in the assessment data to evaluate if this model would fit the data of assessment of mentors' real performance (n=634) using confirmatory factor analysis (CFA). Discriminant validity was tested in the groups which experienced one-to-one mentorship and non-one-to-one mentorship using the *t*-test.

3.4.10 Exploratory factor analysis

Normal distribution assumption check

First the assumption of the distribution of variables for factor analysis was checked. The 47 items had values of skewness within -/+3.0 and values of kurtosis under 7, which implies it is acceptable to do factor analysis (Ferguson and Cox, 1993; Kline, 2005). An interval level of the variables from a Likert scale with five steps was assumed (Kline, 2005).

Sample size and correlation check

KMO showed that the sample size for this analysis was very satisfactory (KMO=0.95), and the KMO values for all individual items were over 0.85, which is above the acceptable criterion of 0.5 (Field, 2009). Bartlett's test of sphericity χ^2 (1081) =13460.94, $p<0.001$ implied that the correlation among the 47 items is significantly different from 0 and thus it is suitable for factor analysis.

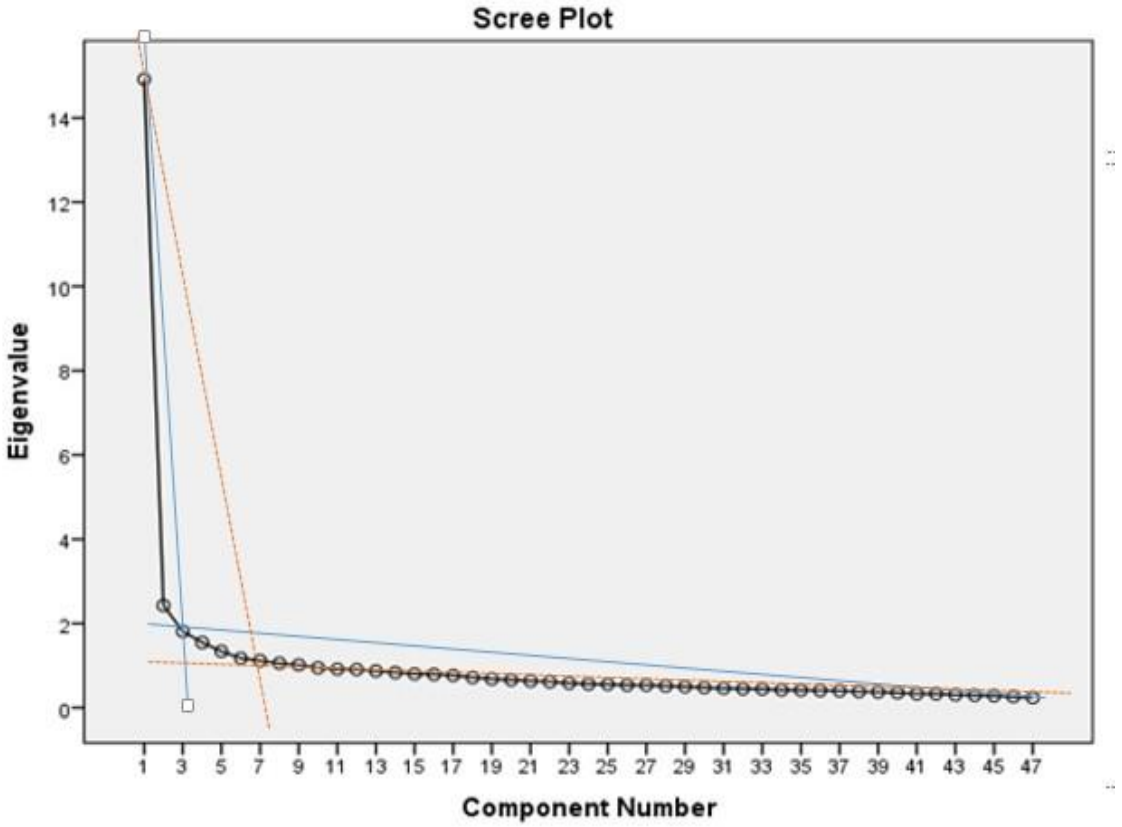
Methods of extraction

In this study both PCA and factor analysis were conducted and they provided identical solutions. Finally principal axis factoring (PAF) was selected for this research aimed to explore the theory of mentorship rather than data reduction.

Number of factors to extract

Based on eigenvalue >1, there were 9 factors, explaining 56.59% communality, which was thought to have overestimated the factors, particularly as the number of variables is greater than 30, communalities <0.4, and sample size <200, the eigenvalue is not accurate (Field, 2009). Scree plot was a little obscure, as the inflexion points showed that there might be two or six factors, shown in Figure 3.6, so it was difficult to decide the number of factors based on these two methods. Therefore another method, Monte Carlo PCA for parallel analysis software, was used. The following data were entered in to calculate random eigenvalues: the number of variables (47); sample size (669) and replication times (250). The estimated random eigenvalues were then compared with the original Eigenvalues in EFA. It suggested that there were four factors (the first four original eigenvalues are greater than corresponding estimated random eigenvalues) as shown in Table 3.6.

Figure 3.6 Scree plot with 47 items



There are two inflexions, one is at the third factor, the other is at the seventh factor, which suggests there might be 2 or 6 factors in the 47 items.

Table 3.6.Comparison of the eigenvalues from the original PAF and the Monte Carlo estimation with 47 items

Component	**Original Eigenvalue**	**Random Eigenvalue**
1	**14.824**	**1.5492**
2	**2.280**	**1.4925**
3	**1.711**	**1.4489**
4	**1.540**	**1.4140**
5	1.340	1.3827
6	1.165	1.3546
7	1.118	1.3278
8	1.049	1.3012
9	1.007	1.2770

The bold numbers show that there are four Eigenvalues in the original analysis larger than their equivalents among the Monte Carlo parallel analysis estimated values, which suggests there should be four factors in the 47 items.

Methods of rotation

In this study as the relationships between factors are not clear yet, both orthogonal rotation and oblique rotation were tried and the results suggested that oblique rotation could give a better solution and the factors do correlate with each other as shown in Table 4.10 in Chapter 4.

The criterion for loading and cross loading was set at 0.4, and based on this, items with loading below 0.4 and cross loading over 0.4 were deleted. After the deletion, the following cycles of decision of number of factor, extraction and rotation were conducted to reach a simple construct.

3.4.11 Confirmatory factor analysis

Introduction

The model generated from EFA was further tested by confirmatory factor analysis with AMOS 22.0 in the assessment data set (n=634). Requirements such as multivariate normality, missing data, multicollinearity and sample size were checked; procedures such as model specification, estimation methods selection and model fit index inspection, model modification, alternative model consideration and cross check model stability are reported in this section.

Requirements check

Distribution of variables was checked first as it can affect the model fit index and accuracy of model estimation. In addition, distribution can also guide the selection of estimation methods. Because multivariate normality inspection is difficult to carry out, univariate normality was checked as a base. According to Kline (2005), multivariate

normality is usually met when univariate normality holds. All the measured variables are normally distributed, as all the absolute values of skewness are <3 and all values of kurtosis are <7.

AMOS also requires large sample size for accurate estimation and this was satisfied in this data set (n=634): the ratio of cases to variables is >20:1 (634:29), which is over the recommended rule of thumb value (5~10:1). At the same time no case with missing data was included to assure stable and precise model estimation.

Multicollinearity was also checked using linear regression, putting each variable in the dependent variable box in turn, and other variables in the independent variable box in SPSS 22.0. No tolerance is below 0.1, nor is any Variance Inflation Factor (VIF) is over 10. No VIF is even over 5. The correlation matrix was also checked: no correlation coefficient is over 0.85, so these suggest that there is no multicollinearity among all the observed variables (Kline, 2005; Field, 2009).

Specifying the model

According to the results generated from EFA, the model was specified. Latent variables were represented with ovals, the observed variables were put in rectangles and errors in circles were named. After that bi-directional arrows were drawn among the factors to co-vary them, as EFA has shown that the factors are correlated. The preliminary modelling found that the three factors were highly correlated (r>0.8) which suggested a general second-order factor ‘mentorship’ may exist. So the final model was modified as mentors’ behaviour is a general factor at the second-order, having a direct effect on the three first-order factors.

Model fit estimation and modification

Maximum likelihood method (ML) was used to estimate the model fitness as to whether it is a robust method. This data set, with large sample size, normality and no missing data basically meets all the requirements. The preliminary model fit index showed that the model did not fit the data well, so modification based on the model modification indices was conducted by co-varying several pairs of errors, and the corresponding observed variables of correlated errors measure similar concepts; so this will not cause change to the hypothesised model.

Equivalent model and model stability

When establishing a model, an alternative model or equivalent model should be considered to find out which model is preferable. Under some conditions there are infinite equivalent models (Kline, 2005). After a model is established in one data set, model stability across data sets should be checked (Kline, 2005).This model was also checked in the importance data set (N=669) using CFA with ML in AMOS 22.0.

3.4.12 Mokken scale analysis

A Mokken scale analysis was carried out to explore whether there were hierarchical properties in mentors' behaviour and to investigate the scalability of this new scale.

Assumption check

The assumptions of Mokken scale analysis were checked, such as unidimensionality, local independence and monotonicity. The whole scale showed unidimensionality from c=0.05 to 0.3, and the three subscales show unidimentionality from c=0.05 to 0.40. In the monotonicity check, all items showed critical values under 40, implying no violation

of the never decreasing assumption. With regard to local independence, all H_is ranged from 0.32 to 0.41, which is not over 0.5; each item asks different aspects of mentorship; neither show any violation of local independence (Watson et al., 2014). This suggests that Mokken analysis is suitable for this data set.

Sample size check

Minimum sample size requirement to do MSA has been studied recently (Straat et al., 2014). It is related to the quality of items and whether items have good differentiability: with higher H_i, smaller samples are needed, and vice versa. In the current study, with Hi ranging from 0.32 to 0.41, the sample size is adequate but not good (n=669) (Straat et al, 2014).

Mokken scale analysis process

The items identified in factor analysis were checked first for scalability coefficients and any item with H_i under 0.3 or its 95% CI (confidence interval) covering low bound below 0.3 was excluded. Then scale partition was carried out to explore the dimensions of mentors' behaviour through increasing c by 0.05 increments. Scalability as ordering people and items was checked at subscale level and at whole scale level.

3.5 Summary

The methodology and specific methods in each steps of the new scale development and validation have been discussed here and all the results are reported in Chapter 4.

Chapter 4. Results

4.1. Introduction

In this chapter, results from the scale development and validation stages are presented (shown in Figure 4.1). The content mainly includes the construction of the pool of items through literature review and focus group interview, reduction of the items through discussion with the supervision team, and scale validation through expert panel review and survey. All results of planned psychometric tests, such as content validity, factorial validity, discriminant validity, test-retest reliability and internal consistency reliability from classic test theory, and scalability and hierarchy from Mokken Scale Analysis, are reported. All participants, such as those participated in focus group, expert panel and survey are presented in Table 4.1.

Figure 4.1 Results of scale development and validation

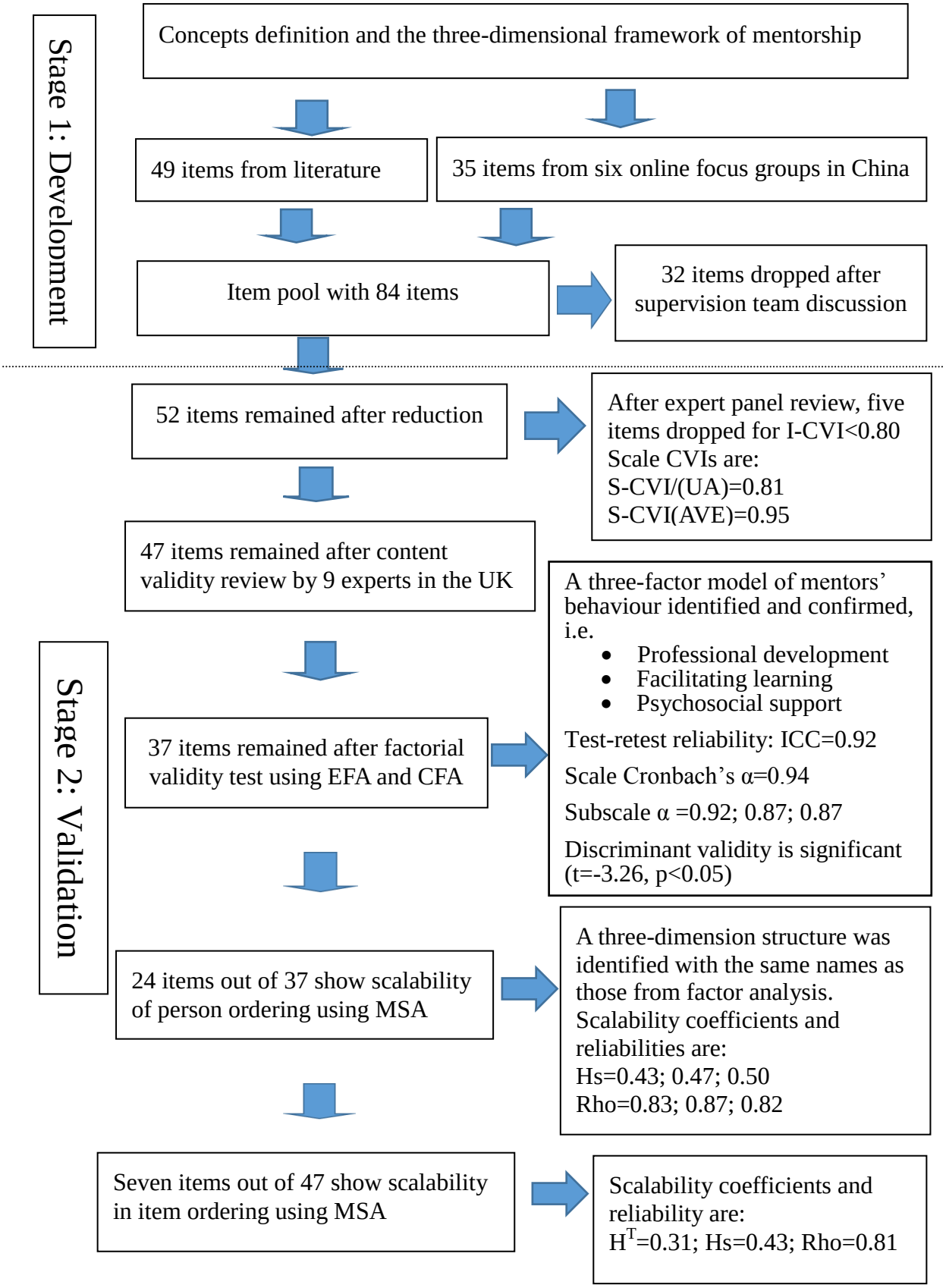


Table 4.1 Participants of this study

Area participants came from		Focus group participants		Survey respondents	Mentor experts
		Students	Mentors	Students	Mentors
China	Southwest China	17 degree and associate degree students	11	742 students from degree, associate degree and diploma programs in three hospitals	0
	South China	12 degree and associate degree students	5	0	0
	Eastern China	0	2	0	0
	Middle China	0	2	0	0
The UK	Yorkshire	0	0	0	9
Summary		29	20	742	9

4.2. Items from the literature

At first, 140 items were selected or adapted. After discussion in the supervision team and grouping similar items together, 49 items remained; the sources of all items were identified and the dimension that each item belonging to was also marked (Appendix 19). This preliminary item pool contained 16 items about professional development, 19 items about facilitating learning and 14 items related to psychosocial support, shown in Table 4.2.

Table 4.2 Items and resources from literature

Source	Eight mentorship scales in nursing
	NMC guideline
	Five studies
Items selected (n=49)	Professional development (n=16)
	Facilitating learning (n=19)
	Psychosocial support (n=14)

4.3. Items from the focus group

Both synchronous and asynchronous online focus groups were conducted to confirm items from the literature and generate new items. The composition of groups, members' qualifications and characteristics, time spent and new themes that emerged are reported in this section. The demographic data and themes from focus groups are reported in Table 4.3; items generated and reworded are shown in Table 4.4; and the whole item pool is displayed in Appendix 20.

4.3.1 Synchronous online focus group

Student groups

All the students from two universities were at their end of clinical learning (final year of study with about consecutive 10 months of clinical learning). The first focus group, including one male student and four female students with mean age 22, came from one university in the southwest of China (named University 1 in this study) and studied in three teaching hospitals, two non-teaching hospitals: all the hospitals are tertiary hospitals located in Sichuan Province. The discussion lasted 120 minutes and generated approximately 8000 words. Themes such as professionalism, deep learning, feedback-giving, evaluation and assessment, learning environment and psychosocial support and relationship were highly stressed.

Students in the second focus group came from a university in South China (named University 2 in this study) where a more prosperous economy and better educational resources exist and mentors have more training and are more diverse; there are more male mentors and mentors with overseas working or learning experience. Five degree students studying in different hospitals in that province were invited but only four of them (one male and three females) participated in the discussion. This group discussion

also lasted 120 minutes and 7000 words were generated. The students in this group used longer sentences than the first group and they seemed less emotional. They had different experiences, with more formal one-to-one mentorship and more positive experiences from placement learning. This group provided new informations about teamwork and teaching methods, in addition, they raised the question of the demographic characteristics of their ideal mentors, e.g. mentors should be young, experienced and have higher qualifications, at least to degree level.

The third group came from the same population as the first group but seven students were invited in case some students would not turn up for different reasons. The composition was a little different from the former two (two students from associate degree program and five from degree program); 120 minutes were spent and approximately 9000 words were generated. Two new themes, i.e. inter-professional learning and accepting students' difference emerged.

Mentor group

Five female mentors from one tertiary hospital in Sichuan Province attended this study. They each had over two years of students mentoring experience at least but had not attended a mentor training program. This mentor group generated approximately 4400 words after 130 minutes discussion. One new theme, student and mentor match, was added.

4.3.2 Asynchronous online focus group

The two asynchronous online focus groups (one mentor group and one student group) were open for three weeks (1 – 22 May 2014). Basically the synchronous group showed more emotional involvement, short and instant responses, while the asynchronous groups showed more reflective thinking, more logical and longer opinions.

Asynchronous online mentors focus group

The asynchronous online mentor focus group consisted of 15 mentors (two males and 13 females) from seven hospitals (one from a teaching hospital in eastern China, Jiangsu Province; one from Hunan Province in the middle of China; two in Guangdong Province, south east of China; three in Sichuan Province, southwest of China). All mentors came from tertiary hospitals, except one from a secondary hospital; three of the mentors had master's degrees and the others had undergraduate degrees; they had worked more than five years and mentored students for at least two years; three had experienced mentor training provided by their own hospitals but without updating mentoring knowledge and skills. The age ranged from 26 to 42 years and mean age was 32; basically they were young and felt comfortable with online chat. This group generated 30,000 words and gave two new themes: evidence-based nursing practice, and fostering critical thinking.

Asynchronous online student focus group

Thirteen students attended the asynchronous focus group. The students had or had not had one-to-one mentorship from the same two universities as the three synchronous student focus groups and they were distributed in different hospitals in different cities such as Guangdong, Zhejiang, Shanghai, Sichuan, and Chongqing. Finally, 15,000 words were contributed. No new themes appeared and data saturation was reached.

4.3.3 Items generated and reworded

Data from focus groups were analysed using thematic analysis. According to the themes and raw data, items were drafted. For example, one item 'transmit positive image of nursing' was drafted based on the theme 'professionalism'. The raw data are presented below. Both mentor focus groups and student groups highlighted that mentors must help

nursing students to develop their nursing professional identity through demonstrating a positive image. One student said ‘as students we need positive power and inspiration to take on the role as a nurse. But my mentor always complains about being a nurse. She said that ‘to be a nurse is meaningless, stressful, earning little money, with low social status…’ (作为学生，我们需要正能量和鼓励来承担护士角色。但是我的导师总是抱怨当护士，她说当护士没有意思， 压力大，薪水少，社会地位低.....). One mentor said ‘At this moment, the most important thing as mentors is that we need to protect and foster students’ enthusiasm of becoming a nurse.’ (一位导师说：‘现在对我们导师来说，最重要的就是要保护和培养学生当护士的热情’。)

Based on the six focus groups, all items except ‘showing leadership’ from the literature were confirmed; 35 new items were generated and five items were reworded (11 for psychosocial support, 12 for professional development and 17 for facilitating learning), shown in Table 4.4.

Table 4.3 Demographic information and themes from focus groups

	Synchronous online focus group				Asynchronous online focus group	
	Group1	Group2	Group3	Group4	Group5	Group6
Group size	5	4	7	5	15	13
Time	120 minutes	120 minutes	120 minutes	130 minutes	3 weeks	3 weeks
Participants	Students from University 1	Students from University 2	Students from University 1	Mentors from Hospital NO.1	Mentors from 7 hospitals	Students from Universities 1 & 2
Words	8000	7000	9000	4400	30,000	15,000
Age (Mean ±SD)	22 ±1.52	22 ± 1.48	22± 1.89	30± 3.54	32 ± 5.12	22± 1.67
Gender	One male	One male	One male	No male	Two males	Two males
Qualification or program	Degree	Degree	Degree and associate degree	Degree and associate degree	Master's degree and Degree	Degree and associate degree
New themes	Professionalism Deep learning Psychosocial support Learning environment Relationship Facilitating learning Feedback giving Evaluation	Teamwork Teaching methods Demographic characteristics	Inter-professional learning Individualised teaching	Mentor and student match	Evidence based practice Critical thinking	NONE

Table 4.4 Items generated and reworded after focus group

Dimension	Item
Psychosocial support (11 items)	Listens attentively reworded as *Listens to students' ideas and suggestions*.
	Confirms and affirms students
	Establishes a good relationship reworded as *Establishes a teacher-colleague-friend relationship with students*
	Takes students as learners, not a pair of hands
	Be caring
	Demonstrate empathy
	Call students their preferred title
	Get to know and understand students
	Guide personal development of students
	Be patient
	Co-working reworded as *Does work together with students*
Professional development (12 items)	Acts as a good role model for nurse
	Discusses nursing professional prospects with students
	Gives students career guidance
	Fosters professional growth of students
	Shows prioritising task and working methods
	Shows clinical competence *(deal with emergency)*
	Maintains standards and principles of nursing procedure
	Transmits negative information of nursing profession
	Instils positive professional attitude
	Promotes students to acquire nursing identity
	Fosters clinical and critical thinking
	Shows team work cooperation
	Co-working reworded as *Does work together with students*
Facilitating learning (17items)	Gives students objective and comprehensive assessment
	Never gives any negative feedback in front of others,
	Corrects mistakes without embarrassing students
	Supervises students
	Asks students questions to facilitate and assess learning
	Arranges interdisciplinary learning activities
	Provides peer students sharing opportunity on wards
	Has a clear plan for students' learning
	Follows the clinic learning plan strictly
	Discusses learning objectives
	Accepts student's individual differences
	Is accountable and responsible for students' learning
	Is active in teaching and instruction
	Helps students to linking theory to practice
	Widens and deepens students' clinical knowledge base
	Helps students develop the identified psychomotor skills *Reworded: Widens and deepens students' clinical skills*
	Demonstrates standard nursing procedures and skills

4. 4 Reduction of items after supervision team discussion

All 84 items from the literature review and focus groups were discussed in the supervision team for proper phrasing and duplication reduction (summary shown in Table 4.5). After that, 52 items remained (shown in Appendix 21).

All the items were phrased positively; one items was divided into two, i.e. Item 63 (see in Appendix 20) 'Does work together with students' was divided as 'My mentor is on the same shift with me' and 'My mentor works with me when we are on the same shift'. One item 'showing leadership' was dropped as irrelevant, which was also not confirmed in focus groups, 31 items were dropped for duplication and 52 remained.

Table 4.5 Decision after the supervision team discussion

Action	Number
Separated	One item divided into two items
Dropped	One irrelevant item and 31 duplications
Remained	52 items

Summary of scale development

Based on the theoretical framework generated from the literature review using content analysis, a scale was developed through literature review and six online focus groups. The item pool expanded from 49 to 84 and decreased to 52 after the supervision team discussion. Meanwhile, all items were phrased positively to improve face and content validity and reduce response bias. Results of the validation stage are presented in the following sections.

4.5. Expert panel review for content validity validation

Fifty two items remained after the supervision team discussion, then this scale was sent to 12 mentor experts in the UK for content validity testing. Nine of them responded, therefore, the response rate was 75%. The demographic data of experts, plus the item level and scale level CVIs are reported below.

4.5.1 Experts' demographic information

The content validity was tested by nine mentor experts with years of registration ranging from 10 to 28 and years of mentoring students ranging from 8 to 22. The experts had at least degree education and four mentors had master's degree or master certificate; they were practice teachers acting as resources for both students and mentors in clinical placement. The demographic information is shown in Table 4.6.

Table 4.6 Demographic data of mentor experts

Qualification	Value
Years of nursing registration	10-28 years
Years of mentoring student	8-22 years
Educational qualification	Four experts with master's degree
	Five with undergraduate degree

4.5.2 Item content validity index (I-CVI)

All item content validity indices were calculated according to Formula 1 (shown in Chapter 3). Forty-one items scored 100%; five items scored 88.9%; two items scored 77.8%; five items scored lower than 70%. The summary is shown in Table 4.6 and more details are displayed in Appendix 21.

4.5.3 Scale content validity index (S-CVI)

S-CVI/UA (the proportion of items on an instrument that achieved a rating of 3 or 4 by all the content experts) was calculated and the value was:

S-CVI/UA=42/52=0.81

S-CVI/Ave (average proportion of items rated as 3 or 4 across the various judges) was also calculated and the value was:

S-CVI/Ave= (47/52+3*48/52+51/52+2*50/52+2*52/52)/9=0.95

S-CVI/UA is greater than 0.4 and S-CVI/Ave is over 0.9 and both imply good scale level content validity according to Polit and Beck (2006 p. 492).

New item recommendation was also invited from the expert panel, but no items were recommended.

4.5.4 Decision after expert panel review

Finally, five items (item 16, 17, 23, 24, 52) were dropped for low content validity (CVI<0.80). While item 26 (CVI =0.78) had greatly differing opinions among experts (two rated somewhat relevant (2) and seven rated highly relevant (4) therefore it remained for further testing, shown in Table 4.7.

Table 4.7 Content validity and decision after experts review

Content validity index (CVI)			
Item content validity index (I-CVI)		Scale content validity index (S-CVI)	
>90%	41 items (remained)	S-CVI/UA	0.81
>80%	Five items (remained)	S-CVI/Ave	0.95
>70%	Two items (dropped one)		
<70%	Four items (dropped)		

4.6. Translation and back translation

4.6.1 Translation

The Chinese translations from four Chinese translators were compared and the differences were discussed. Particularly one phrase ‘a pair of hands’ in Item 46 ‘My mentor treats me as a student rather than a pair of hands’ was translated into 帮手 (helper) or 工具 (instrument) but 劳动力 (labour) would be better to reflect the real meaning and the language that students and mentors used in focus groups. At last, 劳动力 (labour) was selected.

4.6.2 Back translation

The first version of back translation was checked by Professor Roger Watson. Only the translation of Item1 ‘My mentor seems enjoying teaching’ (看起来喜欢教学) was problematic as implied by its back translation (‘My mentor looks like teaching’). The meaning that ‘seems enjoying teaching’ carries may imply that, in fact, a mentor does not like teaching at all, just pretends to enjoy teaching. So this item was reworded as 喜欢教学 (enjoys teaching) and the second back translation was done and checked again.

The final version of this new scale was shown in Appendix 22. It was ready for further testing among Chinese nursing students.

4.7 Pilot study

Twenty three degree students, nineteen associate degree and eight 5-year diploma students (50 in total) participated in the pilot study to rate each behaviour’s importance in one week. The time taken to complete the questionnaire, clarity and understanding, statistics and decision after the pilot are presented below.

Time to complete the questionnaire

The range of time taken to complete the questionnaire was wide (from 8 to 30 minutes), and the reason might be that there were two versions of questionnaires shown at the same time (one was English and the other was Chinese). Some read both before completing the questionnaire, while some just read Chinese. The rationale for using both languages is that Chinese is preferred by most of the students in China. However the Bristol online survey tool does not support Chinese during exporting the online survey results, therefore English was used to help identify results, otherwise it would be chaotic.

Clarity and understanding

No student mentioned any problems of understanding and all of them typed ‘it is easy to understand’ or ‘no problem of understanding’.

Statistic results of the pilot

The internal consistency reliability test and dimension reduction were tried using the 50 responses. The Cronbach’s alpha was very high (0.99) but it was still very likely that there were several dimensions in the scale, thus all the items remained for the main survey.

Decision for the main survey after pilot

To reduce the time of completing the questionnaire in the main survey, the English wording was deleted and numbers (1-5) for options were added to be the signal for exported results. To investigate psychometrics using CFA, another data set is needed: therefore in the main survey students were asked to rate the extent to which they did witness the behaviour of their mentors besides rating the importance of each behaviour. So a clear stem question: ‘Thinking about your most recent mentor, how much do you

agree that they show the following behaviours?', was presented in the questionnaire and students also responded on a 5-point Likert scale from 'strongly disagree' to 'strongly agree'.

4.8 Test-retest reliability and stability

The test-retest was carried out among 50 students in the pilot and 19 students responded (response rate was 38%). The total scores of the 17 students (two responses showing no engagement were excluded) were used to calculate Intra-class Correlation Coefficient (ICC=0.92, over 0.7), choosing two-way random model and absolute agreement type, which implies that this scale shows almost perfect agreement in this group of students according to Landis and Koch (1977), shown in Table 4.7.

The total scale score was stable with this sample (t=2.02, p=0.06) although some items were unstable, such as item16, 17, 45 and 47 ($p<0.05$). All the subscales' and items' test-retest reliability coefficients and means are reported in Appendix 23.

Table 4.8 Test-retest reliability and stability

	Mean	SD	t	p-value	ICC
Test	161.35	24.94	2.02	0.06	0.92
Re-test	154.59	27.19			

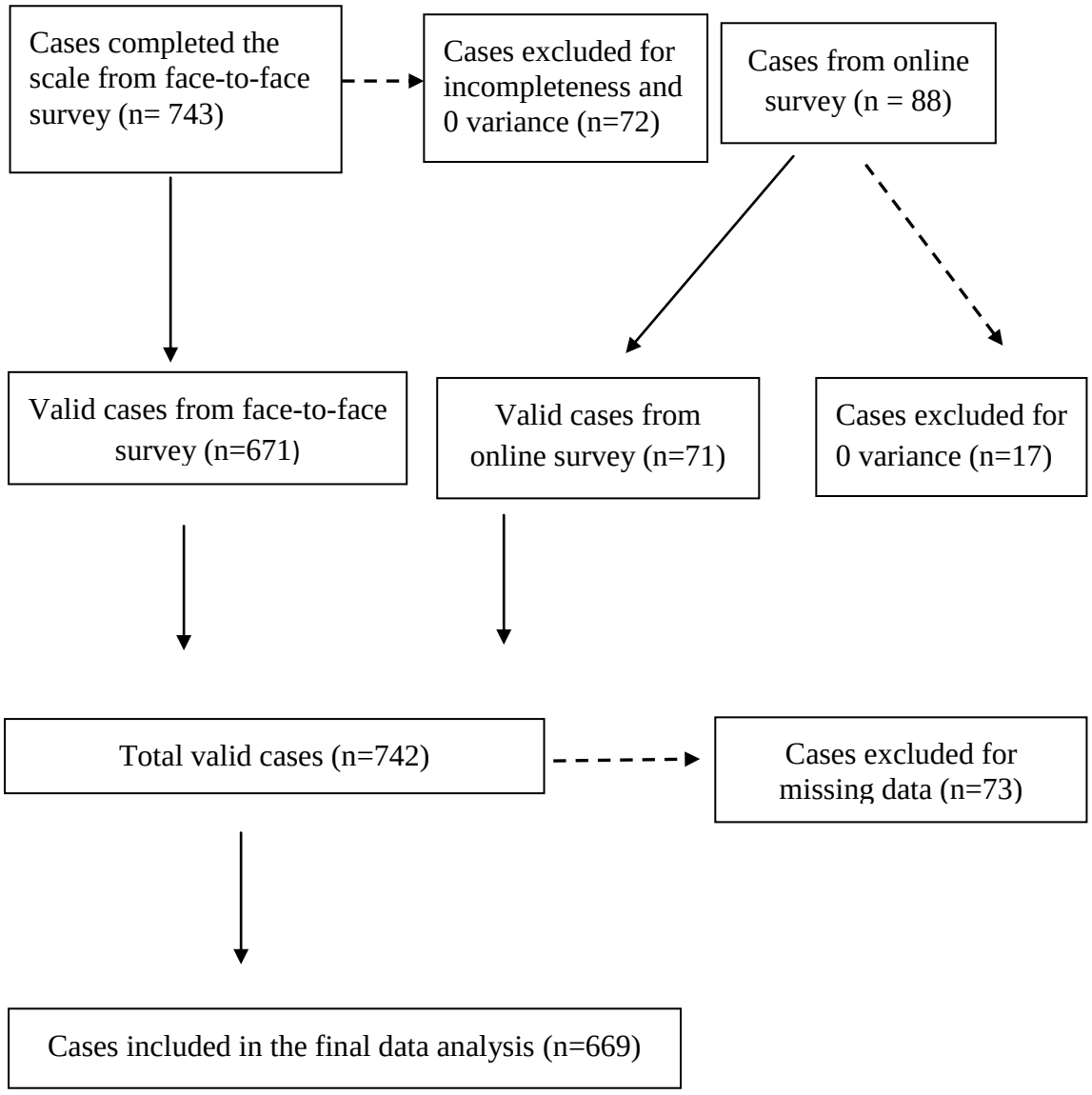
4.9 The cross-sectional survey for factorial validity, scalability and reliability

In the main survey, two data sets were collected, i.e. importance data set and assessment data set (The onine survey questionnaire format was shown in Appendix 24). In each data set, different statistics strategies were applied and different psychometrics was investigated. Results of EFA and MSA and internal consistency reliability generated from the importance data set and results of CFA, discriminant validity generated from the assessment data set are presented in this part, and the syntaxes of the procedures are shown in Appendix 25.

4.9.1 Importance data set

A face-to-face survey was carried out in three hospitals. In the first hospital (identified as Hospital 1 in this study) 362 out of 399 students completed the questionnaire (response rate was 89%); in the second hospital (identified as Hospital 2) 260 out of 300 students responded (response rate was 86.7%); in the third one (identified as Hospital 3) 121 out of 136 responded (response rate was 89%). Together 743 questionnaires were collected, but 72 of them not completed or not properly completed (showing no variance) were discarded; finally 671 cases were eligible for analysis. Eighty-eight students completed the online questionnaire; 17 cases with no variance were excluded, and 71 cases were included. Both the face-to-face and online surveys provided 742 valid questionnaires but only 669 cases were used in the final data analysis (73 cases with missing data were excluded, which is reported in detail in the next section), the data management process shown in Figure 4.2.

Figure 4.2. Flow chart of data management in importance data set



4.9.2 Missing data management

Before conducting statistical analysis, the data needed to be cleaned through checking outliers and missing data. In this data set, there were no outliers, while 73 missing data existed in the 47 items of the scale, particularly focused on item 4 (11 missing), item17 (11 missing), item 24 (7 missing), item 20 (5 missing), which may imply a problem in understanding. In other items the number of missing data was below 5 and only seven items (item1, 2, 11, 12, 38, 46, and 47) had no missing data. The cases with missing data were excluded because Mokken scale analysis does not allow missing data and the sample is big enough as well. At this point, 669 cases were included in the analysis.

4.9.3 Demographic information

The demographic data of the survey are shown in Table 4.9. Students in this sample came from four programs, degree 137 (20.5%), associate degree 238 (35.6%), 5-year diploma 97 (14.5%) and 3-year diploma 196 (29.3%); the majority of them were female 643 (96.1%). These students mainly had following-shifts mentorship 300 (44.9%) and one-to-one mentorship 257 (38.5%), while all other mentorships such as group mentoring (one-to-multiple students mentoring) comprised a very small proportion (16.6%). The students mainly came from Hospitals 1 and 2, accounting for 44.7% (n=299) and 31.1% (n=208) respectively, while the online survey only had 69 (10.3%) students. The average age was 20.29, ranging from 16 to 24 years; average days of study was 75. After the data were cleaned it was ready for factor analysis and the results of EFA are reported in the next section.

Table 4.9 Demographic information of survey

			Frequency	Percentage
Program	Degree		137	20.5
	Associate degree		238	35.6
	5-year diploma		97	14.5
	3-year diploma		196	29.3
Gender	Female		643	96.1
	Male		25	3.7
Experienced mentorship	One-to-one		257	38.5
	Group		84	12.6
	Following shift		300	44.9
	Other		22	3.3
	No clear		5	0.7
Preferred mentorship	One-to-one		585	87.6
	Group		45	6.7
	Following shift		32	4.8
	Other		4	0.6
	Do not mind		2	0.3
Location	Hospital 1		299	44.7
	Hospital 2		208	31.1
	Hospital 3		93	13.9
	Online		69	10.3
	Min	Max	Mean	SD
Age	16.00	24.00	20.29	1.37
Days of study	3.00	330.00	75.64	39.48
Number of units	1.00	12.00	2.73	1.48

4.9.4 Exploratory factor analysis

Three-factor model of mentorship

Principal axis factoring analysis was carried out in 47 items with oblique rotation (direct oblimin). The KMO measure found that the sample is adequate for the analysis (KMO=0.95 and all KMO values for each individual items > 0.85). Bartlett's test of sphericity (χ^2 (1081) =13460.94, $p<0.001$) implies that the correlation coefficients among all the items are large enough to do EFA.

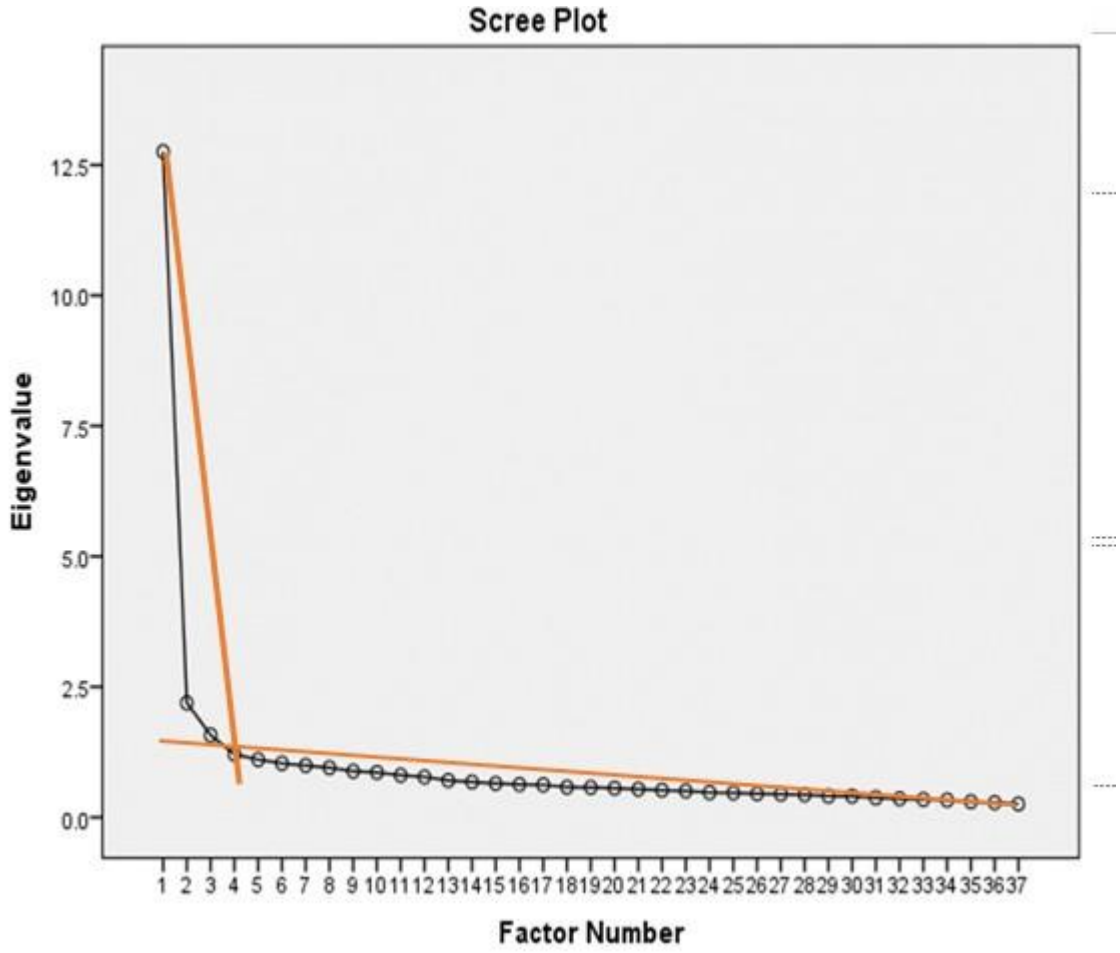
In the exploratory process, several cycles of extraction and rotation were conducted; 10 items were dropped for low loading (under 0.4) and cross-loading (over 0.4). Finally a simple three-factor construct with 37 items was reached, which was supported by both Monte Carlo Parallel Analysis (see Table 4.10) and the scree plot (see Figure 4.3), explaining 44.65% of the variance. Common variances of all items were greater than 0.3 except item 14 and all of them loaded on one factor (all loadings > 0.4) in the initial solution without rotation, shown in Table 4.11. The factor correlation matrix in Table 4.12 shows that the three factors are correlated with each other and the correlation coefficients range from 0.56 to 0.69, which supports the correctness of the selection of rotation method.

Table 4.10. Comparison of the eigenvalues from the original PAF and the Monte Carlo estimation with 37 items

Component	Original Eigenvalue	Random Eigenvalue
1	**12.755**	**1.475**
2	**2.191**	**1.414**
3	**1.575**	**1.376**
4	1.208	1.337
5	1.105	1.308
6	1.034	1.280

Six original eigenvalues from EFA were over 1 but only the former three were larger than those of Monte Carlo parallel analysis estimation, which means that three factors should be extracted.

Figure 4.3 Three factors suggested by scree plot with 37 items



The inflexion point is at the fourth factor, so three factors are suggested by the scree plot. Together with Monte Carlo parallel analysis, this implies the correctness of the decision of factor number to extract.

Naming the factors

The names of factors were given based on the content they contained (the labels of each item were shown in Table 4.15).

Factor 1 - Professional development (PD)

There were 16 items (items 16-17, 20-30, 32-34) in this factor, concerned with showing students nursing skills of patient caring, nursing clinical competency and professionalism, evaluation and assessment, deep learning and challenge.

Factor 2- Facilitating learning (FL)

There were 11 items (items 2-11, 14) in this factor, including identifying and realizing students' learning objectives, arranging learning environment, teaching methods as linking theory and practice, reflective learning, active instruction and interdisciplinary learning activity.

Factor 3 – psychosocial support (PS)

There were 10 items (items 37-42, 44-47) in this factor, concerned with respecting students, treating them as learners not pairs of hands, listening attentively and being friendly, supportive and encouraging.

Table 4.11 Principal axis factoring with oblimin rotation

Item	Commonality	Unrotated loading	Professional development	Facilitates learning	Psychosocial support
v27	0.52	0.67	**0.73**	-0.03	0.02
v22	0.44	0.61	**0.66**	0.09	-0.08
v30	0.36	0.54	**0.65**	-0.05	-0.03
v23	0.42	0.60	**0.65**	0.02	-0.02
v26	0.40	0.58	**0.64**	-0.05	0.04
v29	0.40	0.60	**0.60**	-0.05	0.09
v25	0.38	0.59	**0.54**	0.04	0.07
v21	0.38	0.59	**0.54**	0.15	-0.04
v34	0.45	0.65	**0.51**	-0.04	0.24
v20	0.33	0.56	**0.47**	0.12	0.02
v33	0.45	0.66	**0.46**	0.08	0.20
v28	0.31	0.53	**0.44**	0.00	0.14
v17	0.31	0.54	**0.44**	0.10	0.06
v16	0.39	0.60	**0.43**	0.28	-0.04
v24	0.32	0.55	**0.42**	0.13	0.06
v32	0.37	0.60	**0.42**	0.05	0.19
v7	0.46	0.56	-0.14	**0.71**	0.10
v6	0.46	0.55	-0.02	**0.71**	-0.04
v5	0.46	0.57	-0.05	**0.70**	0.03
v8	0.44	0.55	-0.03	**0.69**	-0.01
v9	0.44	0.58	0.05	**0.61**	0.02
v10	0.40	0.57	0.08	**0.56**	0.03
v3	0.33	0.50	0.06	**0.53**	-0.01
v11	0.40	0.58	0.10	**0.53**	0.04
v2	0.31	0.50	0.09	**0.50**	-0.01
v4	0.31	0.51	0.05	**0.48**	0.06
v14	0.27	0.49	0.13	**0.41**	0.02
v39	0.39	0.52	-0.04	0.01	**0.65**
v38	0.33	0.45	-0.09	-0.01	**0.64**
v46	0.42	0.56	-0.03	0.07	**0.63**
v40	0.49	0.63	0.11	0.00	**0.62**
v42	0.48	0.63	0.05	0.07	**0.62**
v44	0.39	0.55	0.08	-0.03	**0.58**
v45	0.39	0.55	0.04	0.03	**0.58**
v47	0.46	0.62	0.12	0.03	**0.58**
v37	0.41	0.57	0.08	0.01	**0.57**
v41	0.48	0.63	0.14	0.06	**0.55**
Cronbach's α		0.94	0.91	0.87	0.87

*For clarity loadings on putative factors are shown in bold

Table 4.12 Factor Correlation Matrix

Factor	Professional development	Facilitates learning	Psychosocial support
Professional development	1.00		
Facilitating learning	0.69	1.00	
Psychosocial support	0.69	0.56	1.00

Internal consistency reliability of the construct

Internal consistency reliability was conducted to test the homogeneity of the new scale, and it also served the function of testing the reliability of the construct (if the three-factor structure was reliable). Thus the results were reported after factorial validity exploration. The whole scale reliability was 0.94 and the reliabilities of the three subscales ranged from 0.87 to 0.91, shown in Table 4.9, which suggests that this is a reliable construct in this sample and all items are correlated, measuring the same construct—mentorship – and the three factors are reliable as well.

Summary of EFA

EFA suggested that the conceptualization of mentorship is a three-correlated-factor model, i. e. professional development, facilitating learning and psychosocial support. All the assumptions of EFA such as sample size and correlation of variables were tested and met; the key steps, such as extraction method, number of factors to extract and rotation method were checked carefully and chosen sensibly and the results imply correct decisions. This model is identical with the theoretical framework generated

through the literature review. This structure was tested further using Mokken scale analysis and CFA and their results are reported in the following sections.

4.9.5 General level of students' expectation about mentorship

Mean scores of the three factors of mentorship were calculated. The means show that professional development (mean=4.42) and psychosocial support have similar scores (mean=4.41), while facilitating learning has the lowest score (mean=4.36), shown in Table 4.13.

Table 4.13. Mean score of each factor of students' expectation from mentorship

	N	Min	Max	Mean	S D
Mean of FL	669	2.27	5.00	4.36	0.45
Mean of PD	669	2.31	5.00	4.42	0.43
Mean of PS	669	2.50	5.00	4.41	0.48
Mean of total score	669	2.49	5.00	4.40	0.40

4.9.6 Mokken scale analysis

In this part, MSA, a non-parametrical item response theory, was used to explore the dimensions and hierarchy among 37 items identified by EFA. The assumptions of MSA were checked first, then based on the condition that scalability coefficients (H_is and their 95% CI) should be over 0.3, 11 items were dropped and 24 items remained and three subscales were selected among the 24 items. All of them were moderate Mokken scales, but no scale showed invariant item ordering (IIO). Then exploration was carried out in the whole 47 items and a small scale showing IIO was identified. Invariant item ordering (IIO) means that the order of items in terms of difficulty or importance in a certain scale is the same for all respondents. This invariant ordering is useful in prediction, assessment (Sijtsma and Junker, 1996; Ligtvoet et al., 2010), such as in intelligence testing for children; if the items are invariantly ordered, a younger age group can start from easier questions and an older age group can start from more difficult ones, for example. The results were compared with that of EFA.

Mokken scales

To explore the dimensions of the 24 items, c lower bound started from 0.05 and increased to 0.45 with 0.05 increments. From 0.05 to 0.35, all of the items formed a single scale after which three reliable scales were selected at c = 0.40. At c = 0.45 five scales were selected; one of them included just two items; two scales had three items respectively and no more meaningful information was discovered. So the final solution to the Mokken scaling was set at c = 0.40, shown in Table 4.14.

Basically both EFA and MSA generated identical results: both structures have three dimensions e.g. professional development, facilitating learning and psychosocial support. The differences were that EFA included more items (n=37), while MSA had 24

items. Three items about deep learning and challenge (item 32-34) in the professional development factor in EFA were apportioned to psychosocial support factor in MSA. The subscale, psychosocial support, included most items (n=10), the other two Mokken scales shared similar numbers of items (n=6 and 8 respectively). All subscales were moderate Mokken scales (0.4<Hs<0.5) and were reliable (Rho>0.8) but no scale demonstrated IIO, shown in Table 4.15.

Hierarchical properties of the subscales

Mokken ordering according to means showed in Table 4.15.

Scale 1 – psychosocial support – has 10 items mainly concerning support, encouragement and respect. This scale describes the hierarchy of importance of each psychosocial support behaviour. The most highly endorsed concept is 'respect' ('Treats me as a learner, not a pair of hands' and 'Shows respect to me'). This is the most basic need as a person and a student, which may not be met adequately now. After this is encouragement and support ('Instils confidence in me', 'Encourages deep-learning', 'Guides personal development' and 'Makes me feel part of the team'). The least endorsed concept is challenge ('Encourage evidence-based practice' and 'Gives best possible care'). The hierarchy of endorsement is from respect to support and encouragement, ending with challenge. This is a moderate Mokken scale (Hs=0.47) but does not show IIO either, which means that it is reliable and precise to order students according to their expectations, but they may not all rate the items in the same way.

Scale 2 – facilitating learning – has six items and it describes the hierarchy of the importance of behaviour which can facilitate nursing students to learn effectively in clinical placement. The most highly endorsed concept is 'linking theory with practice' ('Helps me to link theory to practice' and 'Actively instructs me'), which is the main purpose and way of clinical learning and teaching. The less endorsed concept is

reflective learning (‘Encourages me to reflect on my learning’), which is a key step in experiential learning. The least likely endorsed concept is learning objectives and plan (‘Has a clear plan for my learning’, ‘Discusses learning objectives with me in placement’, ‘Helps me achieve learning objectives and goals’), which may be more work for mentors other than direct behaviour influencing students’ learning as perceived by students. So the hierarchy is from linking theory with practice to reflective learning, ending at planning and discussing learning objectives. This is a moderate Mokken scale (Hs=0.50) but does not show IIO either.

Scale 3 – professional development – has eight items and it describes the hierarchy of importance of the behaviours that can promote nursing students’ professional development. The most highly endorsed concept is professionalism (‘Demonstrates professional integrity’ and ‘Shows me how to make decisions about patient care’), which is the most important concept: being a nurse, everybody must show professional integrity to save life, be caring and compassionate and understand professional boundaries. This is followed by the concept of professional competency (‘Displays clinical competence’, ‘Shows me how to prioritise tasks’ and ‘Facilitates good communication skills with staff and patients’) and the least endorsed concept is assessment and giving feedback (‘Gives me continuous assessment’ and ‘Gives me constructive feedback’). Therefore, the hierarchy of importance is from showing professionalism to fostering professional competency and assessment. This scale is a moderate Mokken scale (Hs=0.43) but does not show IIO.

Table 4.14 Partitioning of items across Mokken scales with increasing lowerbound values of C (n = 669)

item	c= 0.05	c = 0.10	c=0.15	c=0.20	c=0.25	c=0.30	c=0.35	c=0.40
5	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale 2
6	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale 2
7	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale 2
9	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale 2
10	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale 2
11	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale 2
16	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
20	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
21	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
22	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
23	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
25	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
26	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
27	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale3
29	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
32	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
33	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
34	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
37	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
40	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
41	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
42	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
46	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
47	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1	Scale1
Reliability	Scale1 (0.93)	Scale1 (0.93)	Scale1 (0.93)	Scale1 (0.93)	Scale1 (0.93)	Scale1 (0.93)	Scale1 (0.93)	Scale1 (0.87)
								Scale2 (0.82)
								Scale3 (0.83)
When c increased to 0.40, three reliable Mokken scales were selected. Scale 1 included 11 items; scale 2 included 6 items; scale 3 included 7 items.								

Table 4.15 Mokken scaling with items ordered according to their mean score (n=669)

Item	Label	Mean	EFA	MSA	H_i
2	Takes responsibility for my learning	4.54	FL	DNS	<0.3
3	Orientates me to the clinical environment.	4.43	FL	DNS	<0.3
4	Provides a supportive practice environment	4.38	FL	DNS	<0.3
5	Has a clear plan for my learning	4.27	FL	FL	0.39
6	Discusses learning objectives with me	4.20	FL	FL	0.40
7	Helps me achieve learning objectives and goals	4.32	FL	FL	0.37
8	Asks me questions to facilitate and assess learning	4.39	FL	DNS	<0.3
9	Actively instructs me	4.53	FL	FL	0.38
10	Encourages me to reflect on my learning	4.39	FL	FL	0.37
11	Helps me to link theory to practice	4.53	FL	FL	0.38
14	Arranges interdisciplinary learning activities	3.94	FL	DNS	<0.3
16	Assesses my achievements continuously	4.30	PD	PD	0.39
17	Gives me objective and comprehensive assessment	4.44	PD	DNS	<0.3
20	Gives me constructive feedback	4.38	PD	PD	0.37
21	Facilitates good communication skills with staff and patients	4.47	PD	PD	0.37
22	Shows me how to make decisions about patient care	4.50	PD	PD	0.40
23	Shows me how to prioritise tasks	4.47	PD	PD	0.38
24	Guides me to become a registered nurse.	4.36	PD	DNS	<0.3
25	Displays clinical competence	4.47	PD	PD	0.37
26	Demonstrates professional integrity	4.54	PD	PD	0.37
27	Transmits a positive image of the nursing profession	4.45	PD	PD	0.41
28	Fosters critical thinking in me	4.22	PD	DNS	<0.3
29	Makes me feel part of the team	4.50	PD	PS	0.37
30	Makes me aware of the legal implications of nursing care	4.48	PD	DNS	<0.3
32	Encourages the use of evidence-based practice	4.31	PD	PS	0.39
33	Motivates me to give the best possible care	4.36	PD	PS	0.42
34	Encourages in-depth learning about clinical practice	4.49	PD	PS	0.40
37	Always makes time to teach me	4.36	PS	PS	0.38
38	Works the same shifts as me	4.13	PS	DNS	<0.3
39	Works with me while on the same shift	4.19	PS	DNS	<0.3
40	Supports and encourages me	4.47	PS	PS	0.40
41	Instils confidence in me	4.51	PS	PS	0.40
42	Shows respect to me	4.54	PS	PS	0.39
44	Has a warm and friendly attitude	4.43	PS	DNS	<0.3
45	Listens to my ideas and suggestions	4.41	PS	DNS	<0.3
46	Treats me as a learner, not a pair of hands	4.60	PS	PS	0.38
47	Guides my personal development	4.49	PS	PS	0.38
	PS= psychosocial support FL=facilitating learning PD=professional development DNS = did not scale. EFA = factors identified by exploratory factor analysis MSA=scales selected by Mokken scale analysis Hi<0.3: the scalability of a item is lower than 0.3 or its 95% CI <0.3 For mean scores, scores are on Likert scale, 1 = not important at all, 3 = not clear, 5 = quite important; a high score indicates more important attitude towards mentors' behaviour. Mokken Scale 1: PS: Hs = 0.47; Rho = 0.87; H^T= 0.08; Mokken Scale 2: FL: Hs = 0.50; Rho = 0.82; H^T= 0.11 Mokken Scale 3: PD: Hs = 0.44; Rho = 0.83; H^T= 0.02.				

Exploring items showing IIO

All the subscales show monotonicity, which means that as the expectations of students increase, the accumulated scores of items will increase; but people with different levels of expectations may not respond to the items in the scale in the same order, as no IIO shows in any of the scales. Then the IIO property was explored in the whole scale.

All 47 items were tested using Mokken scale analysis. First five items scaling 0 (item 1, 13, 18, 19, 31) and one item scaling 2 (item 43) were removed, then IIO was checked in the remained 41 items. The items that violated IIO were dropped one by one. Finally 7 items remained and formed a reliable moderate Mokken scale showing weak IIO property (H^T=0.31, Hs=0.43, Rho=0.81), shown in Table 4.16.

This scale with 7 items describes the hierarchy of importance of behaviours that mentor should have towards nursing students in clinical placement. The most highly endorsed concept is ‘respect and support’ (Treats me as a learner, not a pair of hands; Shows respect to me; Instils confidence in me; Listens to my ideas and suggestions), followed by professionalism (Adheres to recognized standards of practice). The least likely endorsed item is ‘Arranges interdisciplinary learning activities’. So the hierarchy is from respecting and supporting students to showing professionalism, ending at ranging interdisciplinary learning activities.

Table 4.16 Items showing IIO

Item	Labels	Mean	Hi
14	Arranges interdisciplinary learning activities	3.93	0.31
36	Adheres to recognized standards of practice	4.15	0.40
40	Supports and encourages me	4.47	0.48
41	Instils confidence in me	4.51	0.50
42	Shows respect to me	4.54	0.49
45	Listens to my ideas and suggestions	4.41	0.44
46	Treats me as a learner, not a pair of hands	4.60	0.45
	Hs=0.43, implying a moderately precise Mokken scale in ordering people. H^T=0.31, implying a weakly precise Mokken scale in ordering items. Rho=0.81, implying a reliable Mokken scale.		

Summary of MSA

Mokken Scale Analysis as a non-parametric item response theory was conducted to explore the structure of mentors' behaviour in different way, the hierarchy in mentors' behaviour and the scalability of the new scale. Three reliable Mokken scales were identified, showing the hierarchy of student's expectations, which is identical to the three-factor solution from EFA. This implies that the conceptualization of mentorship is stable as classical test theory and item response theory shows the same solution. But the more useful and unique aspect of MSA is that it found out the hierarchical properties of mentors' behaviour from respect and support to professionalism and to inter-professional learning. It also shows the precision of the new scale in ordering students according to their expectations about mentors' behaviour. These findings make people understand more about mentors' behaviour and students' needs.

4.9.7 Summary of the analysis in the importance data set

In this data set, first, EFA was carried out and a mentoring model with three correlated factors (professional development, facilitating learning and psychosocial support) was identified, in which all the factors showed homogeneity (internal consistency reliability coefficients >0.8). More advanced test theory such as MSA was used, which partitioned three subscales and these were identical with those of EFA. All the subscales are reliable showing the hierarchy of respondents ordering but none shows IIO. A small scale with seven items showing IIO was identified from the whole scale with 47 items. Then this model was tested in the assessment data set using CFA to find out if this model can stand when it is applied to assess mentors' actual performance. Other validities such as discriminant validity were also tested in this data set and the results are presented in the next section.

4.9.8 Assessment data set

The assessment data set collected information about how students rated their mentors' real performance using the same questions as in the importance data set. The analysis in this data set aimed to find out if the model identified through EFA can be confirmed and to explore discriminant validity as well. This data set was checked first for outliers and missing data, like the importance data set. Finally 634 cases without missing data were included in the analysis.

4.9.9 Confirmatory factor analysis (CFA)

Model fit estimation

The maximum likelihood method (ML) was used. The preliminary model fit index showed that the model did not fit the data well, so modification based on the model modification indices was conducted by co-varing 14 pairs of errors measuring similar concepts. The final indices showed that the model fits the data; both the original fit index and final one are shown in Table 4.17 and error pair's correlation coefficients are shown in Table 4.18.

The indices for the final model show model fit and the model improvement. For example, the original chi-square value (χ^2 (612) =1381.44 (p=0.001)) did not suggest a model fit. Due to sensitivity to sample size, it is not surprising to get a large chi-square with this big sample size (N=634) but it changed from 2269.10 (626) to 1381.44 (612) after correlating the errors, which suggests substantial model improvement. CMIN/DF (χ^2/df) is around 2 (2.26) which is under 3, suggesting model fit too.

The comparative fit index (CFI) is 0.95, suggesting model fit; parsimony-adjusted index (root mean squared error approximation, RMSEA) having a correction for model complexity is 0.04 (90% confidence from 0.04 to 0.05) which is under 0.05, indicating a

close approximate fit (Kline, 2005 p.139). Goodness of fit index (GFI=0.89) is around 0.9 which suggests model fit. Standardised root mean squared residual (SRMR) is 0.03, which is under 0. 08 showing model fit too.

The observed variables in the model have strong loadings (regression weights) on first-order factors; loadings of variables on F1 (professional development) range from 0.62 to 0.74; loadings of variables on F2 (facilitating learning) range from 0.58 to 0.78; loadings of variables on F3 (psychosocial support) range from 0.58 to 0.82, shown in Table 4.19. The three first-order factors have loadings from 0.90 to 0.97 on the second-order general factor, shown in Figure 4.4, and the squared multiple correlation coefficients of indicators are shown in Table 4.18, which denotes the variance each observed variable can explain.

Table 4.17 Model fit index

FIT index	Preliminary value	Value after Modification
CMIN/DF	3.34	2.26
Chi-square	2269.10 (df=626)	1381.44 (df=612)
GFI	0.82	0.89
CFI	0.90	0.95
RMSEA	0.06 (90% CI: 0.06~0.07)	0.04 (90% CI: 0.04~0.05)
SRMR	0.04	0.03

Table 4.18 Correlation between error variances

Error pairs	Correlations
1-2	0.18
5-6	0.32
8-9	0.39
8-10	0.25
9-10	0.24
14-15	0.27
18-19	0.23
19-20	0.24
23-24	0.30
25-26	0.26
28-29	0.26
31-32	0.24
33-34	0.36
35-36	0.55

Table 4.19 Standardised regression weights of items on first-order factors and squared multiple correlations of predictors.

Item	Facilitating learning	Professional development	Psychosocial support	Squared multiple correlations
2	0.76			0.57
3	0.58			0.34
4	0.67			0.45
5	0.78			0.61
6	0.79			0.63
7	0.79			0.62
8	0.68			0.46
9	0.78			0.61
10	0.71			0.50
11	0.70			0.49
14	0.68			0.46
16		0.72		0.52
17		0.62		0.38
20		0.69		0.47
21		0.69		0.47
22		0.69		0.48
23		0.66		0.44
24		0.72		0.53
25		0.64		0.40
26		0.65		0.42
27		0.70		0.48
28		0.71		0.50
29		0.74		0.55
30		0.68		0.46
32		0.74		0.55
33		0.74		0.55
34		0.74		0.55
37			0.75	0.56
38			0.58	0.33
39			0.64	0.41
40			0.80	0.64
41			0.81	0.66
42			0.74	0.55
44			0.75	0.56
45			0.78	0.61
46			0.81	0.65
47			0.82	0.68

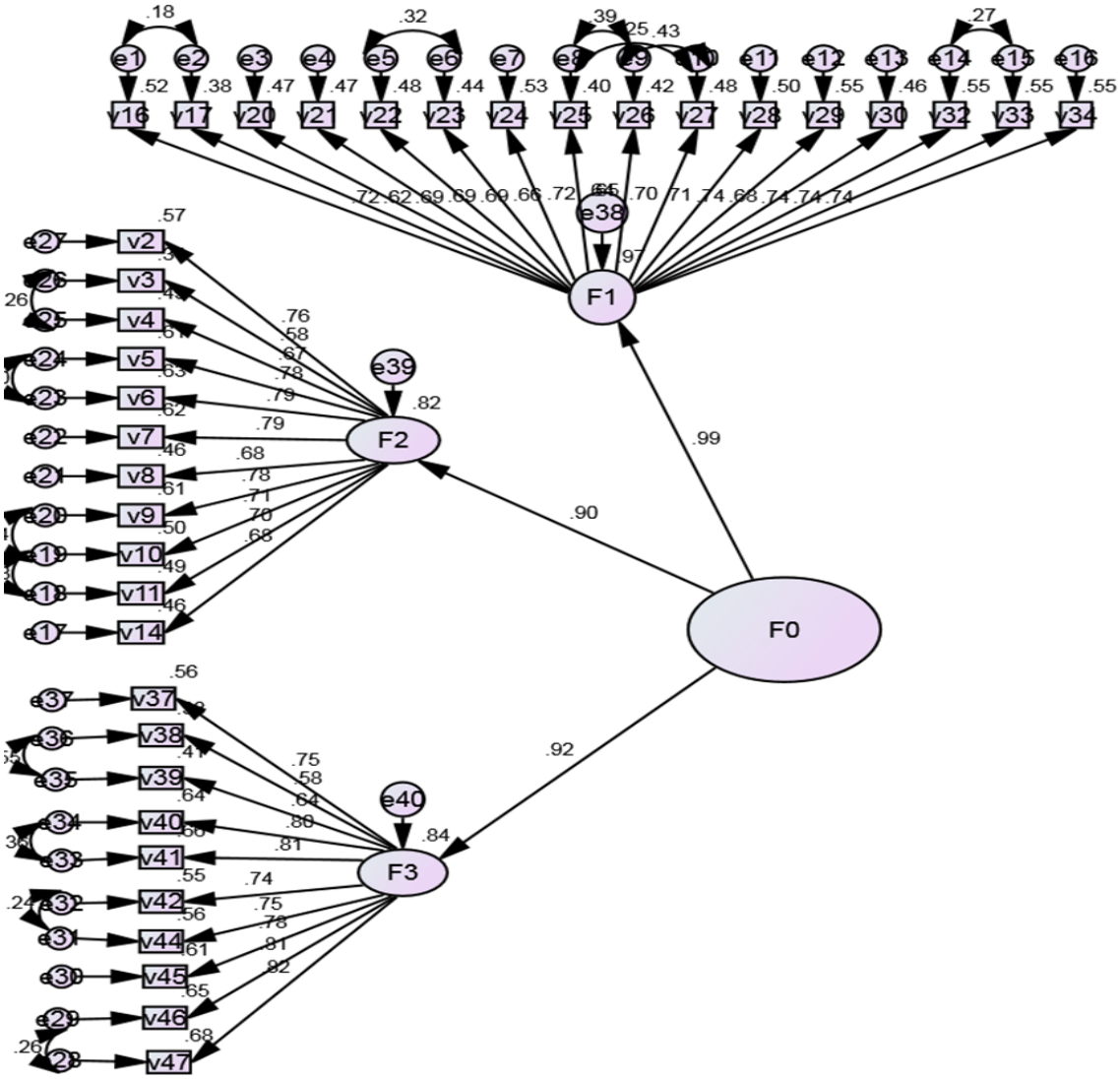


Figure 4.4 Structure of mentorship

This figure represents the structure of mentorship which reflects the relationship between variables in mentors' behaviour scale. Squares denote observed behaviour; ovals represent three first–order latent variables and a second-order factor. Bold arrows denote loadings (standard regression weights) of first-order factors on the second-order factor and of observed variables on first-order factors. The numbers beside three first-order factors and observed variables denote squared multiple correlation of first-order factors and the squared multiple correlation of observed variables, which is shown in Table 4.18 as well. The curved arrows represent correlation of errors and the correlation coefficients are shown in Table 4.17.

The final model implies that when students perceive mentorship, the differences students perceive derive from the 37 observed behaviours, which can be agglomerated into three factors which are correlated; then the difference in the three factors can be agglomerated further into one general mentorship concept. One noteworthy thing is that, although the three factors are correlated, there is no significant cross loading and the loadings of indicators on putative factors are significant.

Equivalent model and model stability

In this study, mainly one alternative model is discussed. The model with three correlated first-order factors and 37 indicators (Appendix 26) is the equivalent model to the hierarchical, second-order model (Fig 4.4), which has the equivalent model fit index, but the second-order model conveys a broader understanding and richer information about the mentorship construct. For model stability check, this model was also checked in the importance data set (N=669) using CFA with ML in AMOS 22.0 and the model fits that data set well (Appendix 27). Therefore the hierarchical, second-order mentorship model is confirmed, showing model stability (Kline, 2005).

Summary of confirmatory factor analysis

In this part the model identified in the importance data set with EFA was confirmed in the assessment data set. All the recommended procedures for conducting CFA (model specification and re-specification, model estimation, equivalent model consideration and model stability cross check) were carried out step by step; all the requirements of doing CFA with AMOS were met; the model was modified based on the modification indices by relaxing 14 pairs of errors correlation. Finally a hierarchical model, a second-order factor (one general mentorship factor) with three first-order factors (professional development, facilitating learning and psychosocial support) and 37 measured variables, was confirmed. It is equivalent to the model of three-correlated-factors from EFA in the

importance data set, which implies that this model can be applied to assess mentors' real behaviour and students' expectations.

4.9.10 Discriminant validity

The discriminant validity was tested between the two groups (one-to-one mentorship and non-one-to-one mentorship groups) with the *t*-test. The assumption of equivalent variance between the two groups was checked and it is not significant (F=2.57, p=0.11), which implies that the data is suitable to do an independent *t*-test. The means of total scores were compared and the difference was significant (t=-3.96, p=0.001), shown in Table 4.20, which implies that this scale is capable of differentiating the mentors who showed expected mentoring behaviour from those who did not show it.

Table 4.20 Comparison of the means of extreme groups

	N	Mean	SD	t	P
Non-one-to-one mentorship	389	101.17	20.38	-3.96	0.01
One-to-one mentorship	245	107.63	19.31		

4.10 Summary of results

In this part, the results of scale development and validation have been reported. The item pool was expanded from 49 items to 84 after focus group interviews among Chinese nursing students and mentors and reduced to 52 after the supervision team discussion. At the validation stage, 47 items showing high content validity remained after review by nine mentor experts in the UK. Through EFA, a correlated three-factor mentorship model (professional development, facilitating learning and psychosocial support) was identified in the importance data set. This model was further tested using CFA in the assessment data set, from which a hierarchical structure, a second-order general factor (mentorship) with three first-order factors, was confirmed, which is equivalent to the model from EFA. The model from EFA was tested using MSA as well and a model with three Mokken scales was generated, which is basically identical to the construct from EFA. In addition, a small Mokken scale with seven items showing hierarchy of students' expectations about mentors' behaviour was investigated.

Other reliabilities were also tested, such as test-retest reliability, which showed that the scale was stable from item level to scale level, and the internal consistency reliabilities were over 0.8. Also the *t*-test in extreme groups showed the minimal discriminant validity. The conceptualization of mentorship, its hierarchical property and theoretical implications are reported in Chapter 5.

Chapter 5. Discussion

5.1 Introduction

This study was carried out to develop and validate a new instrument related to mentors' behaviour, in order to enhance the understanding of students' expectations from mentors and to measure mentors' real performance, to inform their training needs. The study addressed the following two research questions:

1. What is the conceptualization of mentorship in the field of clinical nursing education?

2. Is the instrument based on the conceptualization psychometrically sound?

In this study, systematic and rigorous procedures were conducted to develop and validate a mentors' behaviour instrument including aspects such as: concept definition; theoretical framework construction; item pool generation; content validity review through mentor experts; and a large sample survey for other psychometric evidence, as discussed in Chapter 3. As a result, a novel and unique instrument related to mentors' behaviour has been developed and validated. This instrument showed three inter-related factors of mentors' behaviour and also a wide range of psychometric properties such as stability over time, homogeneity in content, differentiability between extreme groups, hierarchical properties of students' expectations and scalability in ordering them. Therefore, the utility of this new tool to assess students' expectations from mentors and to measure mentors' real behaviour has been demonstrated.

This chapter critically discusses the structure, content and hierarchy of mentors' behaviour and its psychometrics. The limitations, implications and recommendations of this study are also covered.

5.2 Structure and content of mentorship

5.2.1 Overview of mentorship

Both classical test theory (CTT): exploratory and confirmatory factor analysis and item response theory (IRT): Mokken Scale Analysis (MSA) were used in this study and showed that mentorship in the field of clinical nursing education can be conceptualized as a model with three correlated factors: professional development; facilitating learning; and psychosocial support. In this section, this three-factor structure is compared to the three-dimension theoretical framework generated in the literature review, the eight domains of the *Standards to Support Learning and Assessment in Practice* in the UK (Nursing and Midwifery Council, 2008) and other mentoring instruments in the nursing field.

This three-factor structure has confirmed the three dimensional theoretical framework (professional development, facilitating learning and psychosocial support) generated in the literature review (Chapter 2), which included 46 international qualitative and quantitative studies. This theoretical framework guided the construction of the item pool from the literature (including scales related to mentoring in the nursing field and studies of nursing students mentoring) and from the contextualised six focus group interviews in China. This structure then went through expert review in the UK and rigorous statistical testing using CTT and IRT in two large sample data sets (n=669 & 634, respectively) of Chinese nursing students, which implies that the three-factor structure is guided by theory and empirically precise and stable, mirroring both international perspectives (mainly the UK experts' perspectives) and the local context in China. One observation is that giving feedback is shifted from the facilitating learning

dimension in the theoretical framework to the professional development dimension in the final model, which implied the the necessity to tes theory with large samples.

One advantage of the final model against the preliminary framework is that the relationships of the three dimensions were explored and their interrelated nature was discovered, which gives a clearer theoretical understanding, interpretation of and practical guidance to mentors' behaviour. This relationship suggests that a mentor who does not show good professional support behaviour may be perceived as not facilitating learning nor demonstrating sufficient psychosocial support. This is new, compared to the former studies (Suen and Chow, 2001, Hou et al., 2011, Knox and Mogan, 1985), which mainly simplified a set of observed variables to several factors, without investigating their relationships. This is not sufficient for construction and understanding of a model or a theory without exploring relationships of concepts or internal structure (Achinstein, 1965).

The three-factor structure identified in this study is student-centred, contextualized and parsimonious compared with the eight-domain theoretical framework of *Standards to Support Learning and Assessment in Practice in the UK* (Nursing and Midwifery Council, 2008). The NMC mentorship framework was generated mainly from a statutory perspective and it is orientated to mentors based in the UK nursing education and management system (NMC, 2008); for example, the 'context of practice' domain focusing on clinical practice enhancement on wards to provide a better learning environment, not directly aiming at students learning (Nursing and Midwifery Council, 2008 p.25). This NMC framework may not be compatible with Chinese nursing system. In China, there is no national guidance or job description of mentors' role and responsibility. Furthermore the nursing education and registration system are different: for instance, mentors do not have reasonable responsibility and accountability for

assessment and evaluation of students' learning, as the certificates of graduates are issued mainly based on subject exams in nursing school; registration is based on provincial level exams; and mentors' assessments are not concerned to a large extent. Therefore to guide mentors' behaviour in China, at this point, the students' expectations and needs are more likely to provide practical guidance such as this three-dimension construct.

Some domains in the NMC standard have overlapped contents, such as leadership is redundant with six other domains, covering concepts like learning, planning and assessment (Nursing and Midwifery Council, 2008 p.25); two domains: assessment and accountability and evaluation of learning, also overlap. However, it is not surprising to have this complex structure in the NMC guideline, as this theoretical framework has not been tested or refined empirically. It also suggests that the definitions of these abstract concepts such as leadership, learning environment planning, assessment and accountability are ambiguous, not mutually exclusively and operationally defined. In addition, the relationships of the eight domains are not stated.

The three-factor mentorship model is clearer conceptually and more reality-centred when compared with other mentoring instruments in the field of student nurse mentoring. Some existing instruments fail to recognize and reflect adequately the mentors' function at a very explicit and parsimonious structural level (Berk et al., 2005; Lee et al., 2009; Löfmark et al., 2012) without being guided by a theoretical framework or structure statistical exploration. One instrument (Chow and Suen, 2001) conceptualized mentorship as befriending, guiding, assisting, counselling and advising but it was too obscure to guide mentors' behaviour as it failed to distinguish the five factors theoretically and operationally and also failed to report any supportive psychometrics as discussed in Chapter 2. Some instruments (Knox and Mogan, 1985;

Hou et al., 2011) may be more concerned with mentors' competency or traits like teaching, leadership and problem solving ability, evaluation, interpersonal relations, personality, educational intelligence and professional competence. This may not necessarily be manifested directly in mentors' behaviour towards students and therefore may not be beneficial towards them.

5.2.2 Factor 1-Professional Development

This factor accounts for the largest amount of variance in the data and the largest number of items (n=16), including concepts such as evaluation and assessment, giving feedback, critical thinking, nursing skills and competency nurturing, role modelling of professional integrity, positive image and challenging students (Table 4.13), possessing the highest mean scores (Table 4.17). Professional development is of substantial importance for students, as they can get a proper understanding of nursing culture and acquire professional identity and competency through professional socialization and practice in the real world of nursing care.

This professional development factor is different from the career development function in the business field (Kram, 1983; Dreher and Ash, 1990; Ragins and McFarlin, 1990; Scandura, 1992), which includes concepts like sponsorship, visibility, challenging assignments, and coaching. These concepts are more related to helping staff career development such as achieving a higher level in the organization and/or a salary increment. Staff in organizations in the business field and nursing students are at different stages of their career development (strictly speaking, nursing students have not yet started their careers). Therefore, career advancements to higher levels are too early and not relevant for them. These should be the task of continuing professional development after registration. While at the learning stage of students on wards, complying with professional codes, fostering nursing competence and professional

identity are pivotal to be a registered nurse (Cohen, 1981). In the field of general college and university education, since mentoring takes place in campus, academic learning and subject success is emphasised (Nora and Crisp, 2007), corresponding professional development is missing or is not as conspicuous as in the field of clinical nursing education(Nora and Crisp, 2007).

The most important behaviour of a mentor in promoting nursing students' professional development is not to show clinical competence or psychomotor skills but to demonstrate professional integrity, according to the highest mean score (4.54) of item 26 'shows professional integrity' in this factor (Table 4.13). Professional integrity in nursing means being caring, compassionate, adhering to a professional code of practice and ethics and being professionally consistent (Tyreman, 2011; Edgar and Pattison, 2011), which complies with the 6Cs (care, courage, competence, compassion, communication and commitment) of nursing (Department of Health, 2012). Having and demonstrating professional integrity is emphasized by researchers from the field of higher education (Rose, 2003) and nursing faculty mentoring (Berk et al., 2005). It was highly valued by Chinese student nurses in the focus groups and in the large sample but was ignored by former nurse researchers (Chow and Suen, 2001; Hou et al., 2011). Existing mentoring instruments contain the concepts of role modelling but emphasis the role modelling of teamwork, communication and clinical competency more (Chow and Suen, 2001; Hou et al., 2011; Knox and Mogan, 1985; Lee et al., 2009; Lofmark et al., 2012). A possible reason might be that professional integrity is assumed in mentors and nurses as they are educated and required to be so. However, recent studies suggest that nursing students often observed behaviours showing no integrity (Rees et al., 2015); these behaviours of lack of caring, compassion, morality are also exposed in the *Francis Report* (Francis, 2013). In China, researchers also recognized integrity erosion in the

health care profession (Zhang and Lei, 2012; Ran and Zou, 2015; Sun et al., 2014). Evidence from China and the UK demonstrates that professional integrity needs revisiting and stricter and constant inspection is necessary and essential. The item about professional integrity in the model in this Thesis reflects the current picture of world nursing and it should be upheld firmly and constantly by mentors and all health care professionals.

One unique item, namely 'transmit a positive image', generated from the focus group, also remained in the final instrument, with mean score 4.47. The reason why it emerged might be that the nursing profession has a relatively negative image as a young and underprivileged profession, theoretically and practically less developed, with low career progression in China (Eddins et al., 2011): this negative image was transmitted to students from mentors, which may have a negative influence on nursing identity formation among students. This is also true internationally: an item 'get a better image of nursing profession' in Jakubik's instrument (2012) indirectly implied that a negative image of nursing is perceived in USA. Adding this item in the new instrument may encourage nurse mentors to face reality, understand the importance and meaning of transmitting positive image and strive to change it through valuing nursing, showing the benefits it has brought to the public and discussing the professional prospect internationally, instead of constantly complaining about the current social status, discouraging students' enthusiasm for being nurses and shaking their decision or motivation to choose nursing. Otherwise it may prevent students from forming a positive attitude towards the nursing profession (Chiarella, 2002; Smedley, 2008), and in turn, attrition and turnover rate will increase further (Lee et al, 2009; Fox, 2010).

The item with the lowest mean score (4.22) in this factor is 'to foster critical thinking' which was highlighted in mentors' focus groups but thought to be very demanding

(Darling, 1984; Chow and Suen, 2001). As nursing is experiencing the challenge of patient care in a more complicated environment, critical thinking is unanimously recognized as a necessary thinking skill that a nurse should have in data appraisal, analysis and patient care decision-making (Sullivan, 2012). But it is very demanding due to its high level reasoning skills; therefore proper guidance and training is imperative (Sullivan, 2012). This should be a part of the mentors' preparation program; after mentors grasp critical thinking skills and training methods they can help students to improve it.

The final concept to mention here is 'to encourage evidence-based-practice (EBP)', which implies that mentors should help students to form evidence-based nursing competency. EBP refers to make decisions about patients care, based on three key factors: best possible evidence, clinical expertise and patients' preference and values (Wallen et al, 2010). A similar concept, 'Solving problem based on information and evidence', emerged in Hou et al.'s instrument (2011), however it stressed more mentors' own EBP behaviour and competency rather than nurturing students' attitude and competency in evidence-based-practice. In this research, EBP showed a low mean (4.31), which may suggest that nursing students have other priorities in learning and they do not fully recognize its significance as they are at a very early stage of clinical practice (days of study=75; the total days should be over 300, shown in Table 4.7). But it is the trend and requirement in health care (NMC, 2008) and it is studied and implemented internationally as an imperative and effective way to achieve professional development and good patient outcome (Wallen et al, 2010). In the UK, it has been added to the NMC's standard for students mentoring (Nursing and Midwifery Council, 2008). It is also conducted in China and supported by the Evidence-Based Medicine Centre in Sichuan University and the Evidence-Based Nursing Centre in Shanghai

Fudan University. As EBP becomes more popular, some research and training programs have been conducted to enhance mentors' and preceptors' attitudes and skills about it (Wallen et al., 2010); furthermore, mentors need to demonstrate EBP in daily nursing to nurture a positive attitude, corresponding skills and competence of EBP in students.

5.2.3 Factor 2-Facilitating learning

The second factor 'facilitating learning' means guiding and supporting students' learning. It includes concepts such as being responsible for students' learning, helping to link theory and practice, questioning, reflection on learning and organizing inter-disciplinary learning activity. It emphasises theoretical knowledge and practical psychomotor skills learning through real clinical experience and this is different from teaching in business (Alleman and Clarke, 2002), which is focused on teaching the job and teaching the policies to make a mentee know the organizational culture, rules and routines (Kram, 1984; Drefer and Ash, 1990; Ragin & McFarlin, 1990).

In the nursing field, a mentor needs to demonstrate pedagogical knowledge and competency for they shoulder the responsibility of cultivating and teaching the next generation of nurses effectively, as stated by Nursing and Midwifery Council (2008), which is widely supported and extensively studied by other researchers and instrument developers (Chow and Suen, 2001; Knox and Mogan, 1985; Hou et al., 2011; Lofmark et al., 2012); they went into very subtle details about teaching and learning strategies such as 'elaborate clearly', 'stimulates student interest', 'quickly grasps what students are asking or telling' (Knox and Mogan， 1985). However, clinical learning is highly related to experiential learning theory and social learning theory (Kolb, 1984; Bandura and McClelland, 1977), which proclaims learning through experience and reflection on experience of observing and doing (Kolb, 1984; Yardley et al., 2012). Therefore,

facilitating learning behaviour is mainly about establishing a supportive environment, planning and organizing activities and guiding learning and reflection (NMC, 2008).

In this factor, the highest mean score (4.54) was achieved by the item 'takes responsibility for my learning'. It implies that mentors should bear the duty and obligation to mentor students. In the UK, supporting learning is stated as a requirement of all nurses ('you must support students and colleagues to help them to develop their competence and confidence.') in *The Code: professional standards of practice and behaviour for nurses and midwives* (Nursing and Midwifery Council, 2015 p.9), which may transcend mentors' personal interests in and motivation for, mentoring and teaching. Mainly, mentors should be responsible and accountable for organising learning activities, supervising, providing constructive feedback, setting learning objectives, assessing total performance and providing evidence of their achievement (Nursing and Midwifery Council, 2008). Responsibility is conceived as an ethical issue in student mentoring (Hilli et al., 2014) as well.

In the Chinese focus groups, students and mentors raised the phenomena that mentors did not care and ignored students, showing no interest in teaching them; some students were idling on wards without any mentors to supervise them; to the other extreme, some mentors just ordered students to do endless chores; these are irresponsible behaviours. The reasons might include: there are no national or local guidelines or standards to guide mentors' behaviour; mentoring students is not included in any nursing job descriptions and there is no pay or reward from that either; and that mentors have high workloads and multiple roles to play on the wards without sufficient support from ward managers, nursing teachers and other staff (Eddins et al., 2011). It shows the necessity of training mentors to be aware of and take this responsibility of mentoring nursing students, which should also be listed in job descriptions and guidelines. In addition,

mentors' support and reward should be considered (Dibert and Goldenberg, 1995; Hallin and Danielson, 2010).

Item 14, 'arranges interdisciplinary learning activities', was retained in the factor analysis (both EFA and CFA) although showing the lowest mean score (3.96) in the whole scale (Table 4.13). Interdisciplinary learning was advocated by the World Health Organisation (World Health Organization, 1988; 2010) with the objective of 'learn together, work together', because health care is undergoing profound changes, e.g. ageing population and complicated health problems. Hence teamwork and cooperation of multi-professionals are critical (World Health Organization, 1988; 2010). Research has shown the effectiveness of inter-professional education in enhancing communication and cooperation skills (Netherwood and Derham, 2014; Wilhelmsson et al., 2013). This is popular in the UK, Canada, Australia, USA and the Scandinavian countries, and it has also recently been carried out in some universities and hospitals (Zhang and Chen, 2013) in China and recognized as a promising strategy in medical education (Wang et al, 2014). In the field of nursing, *Standards for pre-registration nursing education* (Nursing and Midwifery Council, 2010 p.9) in the UK, stated that nursing programs should provide inter-professional learning activities to students in school and in placements. It was revealed to be a very rare, special and beneficial experience for students in accordance with the nursing student and mentor focus groups in China, which implies it is relatively new in the field of clinical nursing education. Therefore mentors should be trained about what kind of inter-professional teaching and learning activities are feasible and beneficial and how to arrange and conduct these activities efficiently (NMC, 2010).

5.2.4 Factor 3-Psychosocial support

This factor includes concepts about treating students as learners, with respect, guiding personal development, providing support and encouragement, which is similar to those in business (Kram, 1984; Noe, 1988) and education (Nora and Crisp, 2009; Rose, 2003); however, it stresses particularly the concept of respect, e.g. treating a student as a learner, instead of a pair of hands. This does not mean nursing students ask for more or have higher expectations than people from other fields. On the contrary, it may reflect the actual situation that nursing students are at the bottom of the hierarchical health care setting (Seibel, 2014) and respect is their most common need (100% agreement) (Mao et al., 2014), but they have not been respected sufficiently (Liu, 2014). Previous studies on nursing education placed substantial emphasis on mentors' teaching behaviour and clinical competency (Löfmark et al., 2012; Hou et al., 2011) or simply being friendly (Chow and Suen, 2001), whilst the real experiences and expectations of nursing students may be overlooked or ignored to some extent.

Traditionally, respect to teachers from students was profoundly emphasised and respect towards patients from nurses was also stressed in *The Code Professional standards of practice and behaviour for nurses and midwives* (Nursing and Midwifery Council, 2015), while respect to students and nursing students from a mentor or a teacher was seldom discussed (Giesinger, 2012). Respect is defined as 'a basic moral principle and human right that is accountable to the values of human dignity, worthiness, uniqueness of persons and self-determination' (Browne, 1993 p.213). Mentors should be trained morally to recognize the basic human rights of nursing students in China, and this is the foundation to establish an effective relationship (Browne, 1993), which was reflected in the mentors' focus group: 'the relationship of mentor-mentee can be concluded in one word that is respect.' Currently, how to transmit and show respect operationally varies

across cultures and it is difficult to measure (Browne, 1993). However common behaviours such as addressing students, eye contact and listening attentively should be addressed.

The behaviour, 'treats me as a student rather than a pair of hands' in item 46, was advocated by all students and mentors repeatedly. It also appeared in the literature (Hakojärvi et al., 2014), which implies that it is an international phenomenon, reflecting students' real experience. They were ordered around the wards to do the things that their mentors did not want to do as human power (Eddins et al., 2011). Furthermore the worst aspect was that after the whole day of hard work on the ward, students had learned nothing, as reflected in the students' focus groups. The only motivation that made some mentors teach students was that after the students had learned the skills, they could do more for the mentors, rather than thinking that students need to learn to become qualified nurses. This concept, 'treat students as learners' should be recognized as a standard, common sense and mandatory requirement by clinical educators, mentors, ward managers and doctors in China. It should be applied to a broader practical situation and higher level such as local and national nursing and nursing education regulatory bodies to raise awareness and to improve the situation that students are experiencing.

The other concept—personal development—is universally mentioned in mentorship studies (Allen et al., 2006; Ragins et al., 2000) but not presented as an item in any instrument in any field. The reason might be that personal development guidance is assumed in the intimacy of the mentor and mentee relationship, or it is part of professional development, or it is not highly and urgently recognised and appealed for. The concept 'guides personal development' is deeply and widely reflected in both student and mentor focus groups. It challenges mentors to understand that, usually,

nursing students in China are at the early adult transition stage (17-22 years old) and, according to Levinson (1986), facing a profound change, moving away from adolescence and entering into the adult world; their life structures are undergoing rapid transformation, as old relationships with family will be modified and new relationships in the working and learning environment need to be formed (Levinson, 1986): they are experiencing pressure and anxiety due to ill-preparation for the new adult world (Levinson, 1986). This item reflects students' expectations of seeking useful guidance related to personal psychological and social development from mentors who are working with them side-by-side and are experienced adults. This concept also complies with Chinese education philosophy which perceives teachers or mentors as 'human spirit engineers' (人类灵魂的工程师) who should guide students in the correct direction through solving puzzles in life (Liu et al 2012).

5.2.5 Consideration of some discarded items in validation

It is important to discuss some items discarded in the validation stage and carefully consider the decision further. To drop too many items may damage content validity and precision in measuring the latent trait and the cause of failure in testing should be analysed carefully (Polit and Beck, 2006); proper changes should be made to them and further tests need to be carried out; while, for example, some redundant and poor items should be excluded to maintain the simplicity of the final scale.

One concept 'enjoys teaching' is widely cited in mentoring instruments, quantitative studies (Knox and Mogan, 1985; Hou et al., 2011) and qualitative study (Liu et al., 2011), but it may considers mentorship from the mentors' perspectives. In this study, the item 'enjoy teaching' was discarded because it was too weak to load on any factor in a test of a big sample (loading criterion was set to be 0.4). It may be more suitable in

choosing a mentor or choosing to be a mentor, because 'enjoys mentoring' as a vocational interest or disposition can influence an individual's performance and development (Holland, 1997), while if a nurse has already acted as a mentor, responsibility should be the baseline (NMC, 2008; 2015). Therefore the decision is that if the scale is used to measure mentors' performance it could be excluded, while if it is used to choose a mentor this item can be included.

Another item, 'Organizes learning activities from observing to independent practice', was excluded by the test of a big sample for low loading but it is a very important part of mentors' behaviour as illustrated in the literature (Löfmark et al., 2012), the NMC standard (Nursing and Midwifery Council, 2008) and in the students' focus groups. This item may be too broad and general; different students may perceive it differently and it can be changed into as 'gives me more hands on experience', or 'gives me more practising opportunity' which may be more specific and also reflect one of the most important tasks of psychomotor skills training and development.

Summary

This section has answered the first research question about the conceptualization of mentors' behaviour. The three-interrelated-factor structure of mentorship, its uniqueness and meaning were interpreted through comparing with and contrasting to the originally generated theoretical framework, the NMC model and models from the business, education and nursing fields. Some items reflecting an authentic picture of clinical nursing education have been found and the rationalities of their existence and implication have been explained. In the next section, another property, i.e. hierarchy of mentors' behaviour is discussed.

5.3 Hierarchical property of mentors' behaviour

In this study, item response theory, specifically, Mokken scale analysis was used to explore any hierarchical property and structure of mentors' behaviour. Among all the mentoring scales in business, education and the nursing field, only one instrument – the Manchester clinical supervision scale (MCSS) (Winstanley and White, 2011)– used an advanced statistical method, i.e. Rasch model. This is a parametric item response theory to explore differential item functioning which was used as the sample was inhomogeneous and to reduce redundancy but not to investigate its hierarchical property (Winstanley and White, 2011). This study used Mokken scale analysis to provide a new angle of understanding of the hierarchical property of mentors' behaviour.

In this newly developed mentor's behaviour instrument, three Mokken scales showing monotonicity demonstrated moderate to strong accuracy (Hs>0.4-0.5) in differentiating students' expectations (Table 4.12). The monotonicity model is used to order people when selection of people with a certain trait is needed (Sijtsma and Junker, 1996). In this study, the three Mokken scales, manifesting moderate to strong precision in ordering students' expectation, give a basis to match students with mentors, e.g. match students with high expectations with mentors with high quality of mentorship.

But none of the three scales showed invariant item ordering (IIO); only a small Mokken scale with seven items selected from the total 47 items showed weak IIO (Table 4.14). This small Mokken scale suggests that there is hierarchy in importance of mentors' behaviours and that students from any program, no matter whether diploma, degree or associate degree, will rate the importance of these behaviours in the same way (Sijtsma and Junker, 1996). The item 'treat me as a learner, not a pair of hands' was ubiquitously agreed as being more important or more popular than others; the item 'arrange interdisciplinary learning opportunity' had the lowest mean score which means that the

lowest rated need of students is inter-professional learning. A mentor should understand this ordering and meet the most common and important need first when mentoring any student from any program, at any learning stage and then think about other expectations.

The small scale showing IIO can be applied to help mentors to communicate and understand students better using less time, as the items are ordered (Sijtsma and Junker, 1996). Students from a degree program or at the late clinical learning stage may have high expectations, mentors can start with topics with low mean score (means a less popular need). If the least endorsed item is required by the student, other questions with higher mean sores need not to be asked as every student will be more likely to endorse more popular items. If students have low expectations (students from a diploma program or at early stage of clinical learning), mentors can start with the question with highest mean score and stop at questions which students do not endorse. They do not need to go further to ask other questions with lower mean score behaviours.

Scales showing IIO can also be used to identify aberrant response patterns using person-fit statistics (Sijtsma and Junker, 1996; Meijer et al., 2015). In this study, as the seven items are ordered in terms of importance, if this ordering does not show in any case, an invalid response pattern can be detected. For example, if a student rates the least important behaviour higher than the most important behaviour, which can be detected using the R-package PerFit (Tendeiro and Tendeiro, 2014), the potential reasons can be explored, such as conscientiousness, idiosyncratic response behaviour, understanding problems and response motivation (Meijer et al., 2015).

The reason why three subscales from the Mokken analysis (professional development, facilitating learning and psychosocial support) did not show IIO might be that the items are too close conceptually which may cause intersection (Watson et al., 2014). Under this condition further item selection should be carried out to include a fuller spectrum of

behaviour, from the least to the most popular ones, from least difficult to most difficult or discard some close items. The other reason might be the differential item functioning (Sijtsma and Junker, 1996) in some items, which means that people with the same level of trait from different groups rate the items in different ways. The cause is that some items are multidimensional (tapping both the trait under investigation and other factors like gender, race and age). With respect to this study, the sample included students from four programs (degree, associate degree, 5-year diploma and 3-year diploma) and the requirements for them from nursing schools in higher education institutes are different in China as stated before in Chapter 1; and also the students were at different clinical learning stages (5-year diploma students studied about 300 days and other students studied about 30 days); both probably influenced the rating of their expectations from mentorship. Under this condition, in the future, items with differential functioning should be detected using different methods such as Rasch and multiple-group measurement in CFA, or more homogenous samples are required.

The seven-item Mokken scale showing IIO had five items (treat students as learners, respect, instil confidence, support and encourage, listen) from the psychosocial support dimension according to the three-dimensional theoretical framework and factors identified from EFA. The possible explanation would be that these five behaviours are common expectations of all students despite clinical learning time, learning programs and learning hospitals; while professional development and facilitating learning expectation may potentially be influenced by factors such as the learning program and clinical learning time in the heterogeneous sample, which made the two subscales unable to show IIO. Surprisingly, the psychosocial support factor including 10 items (shown in Table 4.13) did not show IIO. These items had very similar mean score so it was not possible to form a hierarchical Mokken scale displaying IIO.

As a complement to factor analysis, the three subscales from the Mokken scale analysis to a large extent confirmed the three-factor structure of mentorship identified by EFA. The differences between the solutions of MSA and factor analysis are that MSA selected fewer items (n=24) while EFA retained more (n=37), and items 32, 33 and 34 concerning challenge concepts (encourage deep learning, best possible care and evidence-based nursing) were apportioned to the psychosocial support subscale in MSA, while in EFA they belonged to professional development. The reason for this difference might be that MSA has stricter assumptions, such as monotonicity, which is likely to exclude more items violating these assumptions (Sijtsma and Junker, 1996); the partitioning method in MSA is different from the factor extraction method in EFA, so the concept 'challenge' conveyed by items 32-34 is more related to psychosocial support in MSA rather than professional development in EFA; it may also be caused by their conceptual multidimensionality (Sijtsma and Junker, 1996).

5.4 A reliable, valid and scalable mentorship scale

This new scale is verified to be valid, reliable and scalable by the systematic procedures applied in the scale development and validation stage, e.g. a large sample, the application of advanced statistical strategies in a rigorous way, and significant and cohesive psychometric evidence. Exploratory factor analysis generated a three-factor mentorship model in the importance data set, which reflected the three dimensional theoretical framework generated from the literature review; confirmatory factor analysis confirmed this model in another data set (assessment dataset) shown in Fig 4.4; this structure is also confirmed by MSA. Taken together, these reinforce the psychometric properties of the construct of mentorship using different methods.

The expert panel review showed acceptable content validity index (CVI) at item level and scale level (Table 4.6); test-retest also showed substantial reliability coefficient at item,

subscale and scale level (Table 4.7; Appendix 23); internal consistency reliability coefficients at subscale level and scale level manifested homogeneity (Table 4.10 and 4.13). It also showed discriminant validity, which demonstrated the ability to differentiate the quality of mentors' real behaviour. Mokken scale analysis found three moderate to strong Mokken scales which suggests that the scale is scalable in ordering students' expectations and reliable in using raw score to infer students expectations and needs (shown in Table 4.14). One small Mokken scale showing IIO orders the items according to their importance, shown in Table 4.15, which is reliable in predicting students' expectations about mentors' behaviour.

5.5 Conclusion

In this part, the limitations of this research, further research, theoretical and educational implication and recommendation, are discussed.

5.5.1 Limitations

This cross-sectional survey design provides a one-point observation of mentors' performance, and this only allows description and inference of correlation and comparison across groups but not causality (Gordis, 2009). Response rate is also a main concern when using a survey; in the test-retest using an online survey, the low response rate (38%) may cause instability of some items, while in the main survey, the response rate was managed by adding a hard copy survey to compensate the online survey (Online response rate was about 9.8%: 88 out of 900 potential respondents responded in this study.). Here the final response rate was over 80%, which was very good (Babbie, 1990). The low response rate and low engagement of the online survey drew the attention of supervision team; the low response may have been due to low access to the website abroad and low access to computers; while the reason for the low engagement

(questionnaires showing high percentage of no and low variance in item scores) was not clear. To avoid this limitation, further psychometrical tests need more advanced design such as longitudinal design and experiment design, which will be discussed in detail in the next section.

Other limitations need to be addressed as well, such as the single method of data collection, i.e. self-report data collection. There might be response bias, known as the halo effect (Streiner and Norman, 2008). If a mentor was perceived well in one aspect, e.g. humorous, she/he might be rated highly in every item, and vice versa. In addition, students completed the questionnaires from their perspective and they might not understand some requirements from professional body and higher educational institutes, such as leadership, planning and organising inter-professional learning opportunity: in this situation, participants' triangulation is needed (Halcomb and Andrew 2005). Also using this triangulation can find out if the scale is suitable for mentors' self-assessment. Triangulation of the data collection method is also needed (Halcomb and Andrew 2005); for example, observation should be used to assess mentors' real performance, which will be more objective.

Sample limitation should not be ignored either. This study used convenient sampling; one local big sample from southwest China may not be representative of all Chinese nursing students, for there is no national clinical mentoring guideline or routine monitoring system of clinical nursing education: the mentoring style and quality varies across the huge country. This limitation may exert influence on the external validity and generalizability of this study (Babbie, 1990). In the future, more representative data from randomized multiple centres should be collected. Sample size may be a problem for the stability test. In the test-retest study, 19 students completed the questionnaire again 10 days after the first survey. 17 cases were included in the analysis and the

results show some unstable items (Appendix 23), which may have been influenced by this relatively small sample, as study shows it should be around 50 respondents (Streiner and Norman, 2008).

5.5.2 Future study

Responsiveness test

One crucial quality of an instrument is responsiveness or sensitivity to change and it is the capability of an instrument to detect the change in respondents and the change difference across group (Streiner and Norman, 2008). This mentors' behaviour instrument can be tested for responsiveness to students' expectations and mentors' real behaviour. For the former, a longitudinal study can be carried out, which is a better design in inferring causal-effect relationship than a cross-sectional survey (Gordis, 2009), and can also be used to test the measurement invariance over time (this will be further explained in the next section). In longitudinal cohort studies students can be recruited to complete questionnaires related to their expectations about mentorship at several points to test the responsiveness of this new instrument to students' expectations (the hypothesis is that students' expectations vary across the spectrum of clinical learning) (Gordis, 2009)

Experimental studies, such as training mentors, should be carried out as well to test if this instrument can spot the mentors' behaviour change after training; so before and after training, mentors and/or students will be invited to complete the questionnaires. In addition, control groups and randomizing methods need to be considered to minimize biases (Gordis, 2009).

Measurement Invariance test

Measurement invariance is the basis for carrying out multiple group comparison as people from different cultures or sub-groups may conceive a latent structure differently, and it is becoming more and more popular (Vandenberg and Lance, 2000, Cheung and Rensvold, 2002). Measurement invariance testing has been done in business mentoring studies (Crisp and Cruz, 2010; Hu et al., 2011a) but not as yet in the nursing field. Measurement invariance tests usually have four levels: configural invariance (same number of factors and each factor has the same items across groups), metric invariance (each factor has the same loading across groups), scalar invariance (item intercept invariance across groups) and measurement error invariance. Statistics such as multiple group invariance test in CFA and Rasch can be used to compute the index.

In this study, the concept 'students' was defined as students studying in any pre-registration program (from degree to diploma), therefore, students from all programs were recruited and it was assumed that all students had similar expectations as all the items were considered to be generic. But in reality the invariance across programs needs to be tested before their expectations can be compared. We can even go further to test if there is measurement invariance among student pharmacists, therapists, dentist and doctors. If measurement invariance does not hold it can show how people from different groups perceive mentorship differently and it can serve as a basis for treating them differently, to some extent, on some aspects.

Measurement invariance across cultures can be carried out as well, as cross-culture testing can pave the way for international use of this new scale and provide a possibility of comparing differences across countries in nursing education and approaching a unanimous understanding. In addition, over time measurement invariance testing can be done. Longitudinal study can be carried out to detect whether at different stages

students' perception of mentorship varies or not, which will give a foundation for understanding that students' needs vary over time (Streiner and Norman, 2008).

Further hierarchical property tests

The hierarchy of each Mokken scale could be tested with more homogenous samples, such as degree or diploma group in the future. In this study, none of the three Mokken scales showed invariant item ordering in the students' expectation data set, which was collected from an inhomogeneous sample. Hierarchy in difficulty of mentors' behaviour can be explored in mentors' real performance data set. If there is hierarchy in difficulty of the behaviours, difficult ones should be targeted and trained for with more effort.

Mentee's behaviour

This study focused on mentors' behaviour, but to make mentorship efficient, mentees' behaviour is equally important (Ensher and Murphy, 2011). In Ensher and Murphy's study (2011), requiring commitment and resilience, measuring up to mentors' standards, and career goal and risk orientation were identified to be mentees' responsibility in business mentorship. In nursing, Hallin and Danielson (2010) identified that nursing students should be prepared well for clinical practice, such as being responsible for their own clinical study, comprehending the outcomes of learning, but this study did not focus on the relationship between mentor and mentee. In the future specific emphasis could be put on mentees' behaviour in mentorship. Questions, for instance, what proper behaviours a mentee should have and what mentors expect from mentees could be addressed.

Other studies

Other construct validities, such as convergent and divergent validity of this mentors' behaviour scale can be tested (Streiner and Norman, 2008). For convergent validity,

some concepts related to mentorship should be found and a hypothesis (Streiner and Norman, 2008);, such as ‘mentoring quality is correlated with students’ professional identity building’ can be raised. Both students’ professional identity (using a valid measurement tool) and mentorship quality (using this new scale as measurement tool) should be measured to calculate the correlation. If they have good correlation, then convergent validity is suggested (Streiner and Norman, 2008). For divergent validity, some concept irrelevant to mentorship should be identified and a hypothesis, such as ‘mentoring quality is not associated with students’ personality’ can be set. In addition, the correlation coefficients of the three factors were very high (0.63-0.81) in the importance data set and 0.73-0.87 in the assessment data set in our sample, which might suggest items in one factor may be correlated with another factor other than the putative one. The convergent and divergent validity of the factors in this instrument needs to be tested with bigger and more representative samples as well. Further research to explore whether mentorship can influence students’ learning qualities, such as nursing competency and quality, professional identity, pass rate, can be conducted.

5.5.3 Implications and recommendations

Theoretical implication

This study provides a new understanding of pre-registration nursing students’ mentorship and a new mentorship model with three interrelated factors, i.e. professional development, facilitating learning, and psychosocial support, which is student-centred and parsimonious. It shares the common concept of psychosocial support with other fields such as business (Kram, 1983) and education (Crisp, 2009; Crisp and Cruz, 2010) but focuses particularly on ‘respect, treats me as a learner’. Its professional development factor is similar to the concept of career development in business (Kram, 1983) but stresses professional integrity; and it shares a common facilitating learning factor with

the education field (Crisp, 2009) but focuses more on experiential learning and being responsible.

For the first time, the hierarchical property of mentors' behaviour was investigated in the nursing field and it is also true among all other disciplines. This hierarchical theory is practical and useful, which informs nursing educators that some expectations perceived universally to be more important need to be met first and training for some behaviours more difficult than others needs to be carried out with more rigour. This study has deepened our understanding of mentorship to more precise features of behaviour ordering, from several equal factors.

Nursing education recommendation

One-to-one mentorship should be conducted as a rule in clinical nursing education as it showed higher score of mentoring against none- one-to-one mentorship (Table 4.20). At least 50% of clinical learning students' time should be allocated with a named mentor and an alternative mentor should also be available (Nursing and Midwifery Council, 2008).

Mentor preparation programs should include at least the content of professional development, psychosocial support and facilitating learning. To help students' professional development, mentors should emphasise professionalism and professional integrity as reflected in this study. Mentors as role models just having nursing competency is not sufficient as ethics will not be guaranteed without monitoring and education (Francis, 2013; Zhang and Lei, 2012; Ran and Zou, 2015; Sun et al., 2014). Pedagogic knowledge and competency needs to be fostered as well, particularly adult learning, and experiential learning theory should be integrated into training programs (Gopee, 2011). Personal respect and treating students as learners is required to establish workable relationships and realize learning and teaching objectives (Browne, 1993).

Students' supernumerary status must be acknowledged at regulatory body level, hospital level and individual level, including ward managers and mentors in China, as 'treat me as a learner, not a pair of hands' showed the highest score among all the behaviours, and students' supernumerary status in clinical learning has been stated in Standards for pre-registration nursing education in the UK (Nursing and Midwifery Council, 2010).

This research outcome should be disseminated to enhance management and educational staff awareness of students' expectations and the role requirement and responsibility of a mentor. This instrument should be widely and constantly used to boost a mentoring culture, and in turn, this will serve as a way to improve nursing image and the students' quality of learning and to decrease attrition substantially.

5.5.4 Reflection

Theoretical frameworks are necessary in instrument development (Devis, 2003): they have an impact on item drafting and selection from literature and on the questions to ask in focus groups. It is based on thorough study and understanding of former research but is prone to be changed in empirical studies (Devis, 2003).

Instruments without psychometric tests cannot give confidence in measuring any behaviour or attitude (Streiner and Norman, 2008), as sampled items selected from literature or a small number of experts may not represent a population's opinion in specific contexts. This selection is influenced by the researcher's and/or participants' personal perspectives: large sample testing can correct this bias (Streiner and Norman, 2008). This study reflected that items such as 'giving feedback' and 'assessment' were conceptualized to be part of the 'facilitating learning' factor by the researcher, but the large sample test showed the different opinion that it belongs to professional development; some behaviours, such as arranging inter-discipline learning activity,

were advocated in the focus group but rated lowest in the big sample test. People should be cautious when they select an instrument and the psychometric evidence should be carefully examined.

Statistics can be a powerful tool, revealing complicated patterns (Field, 2009) among such simple numbers as 1-5. It is amazing to observe a pattern which corresponds to one's framework or shows that a prediction is right, but sometimes it provides quite bizarre results or leads one to a wrong conclusion. At this stage, the researcher must take the responsibility to make a wise selection of strategies and have sensible judgements according to the data and the theories (Field, 2009).

Focus group is a really imperative way to understand participants to enhance content validity (Krueger and Casey, 2009; Morgan, 1997; Morgan et al., 1997),while theory and literature can just give some idea which already exists in other contexts. However as a researcher one could not get the real understanding of the research project before conducting focus groups or other qualitative studies to listen to and understand how participants think, feel and experience (Creswell and Planoclark, 2011). Experiencing the process of participants' interaction, the process of knowledge-building and problem-solving was joyful and inspiring and it could provoke one's thinking and vision (Krueger and Casey, 2009; Morgan, 1997; Morgan et al., 1997). Also the process of moderating proved important, fun and fruitful, although this was the first time that the researcher had used this data collection method. It became clear that qualitative and quantitative methods should be integrated together undoubtedly and necessarily (Krueger and Casey, 2009; Creswell and Planoclark, 2011).

Rigour of methodology is critical to ensure the quality of a study. Quality assurance and maintaining is imbedded in every step of research and some limitations will be caused by the design (Creswell and Planoclark, 2011). In this study the design of the

cross-sectional survey influenced the result, which just allowed the factor construct, while responsiveness to change cannot be tested; causality cannot be inferred. But the content validity and construct validity were trustworthy which was supported by the research steps conducted systematically and scientifically.

5.6 Summary

The conceptualization of mentors' behaviour in the field of clinical nursing education has been established after systematic and rigorous study. This model has an inter-related three-factor structure, the three factors being professional development, facilitating learning and psychosocial support, including 37 concrete mentors' behaviours towards students. Mentors' behaviours have hierarchical property and 'respect, treat students as learners' is the most important concept, which needs to be met universally.

This new mentors' behaviour scale based on the conceptualization is reliable and valid in measuring students' expectations and assessing mentors' real performance. It is also scalable, which could be used to order and differentiate students with different levels of expectation and will be useful to match students with proper mentors.

Rerefences

AJZEN, I. (2002) Constructing a TPB questionnaire: Conceptual and methodological considerations. [Online]. Available: http://chuang.epage.au.edu.tw/ezfiles/168/1168/attach/20/pta_41176_7688352_57138.pdf [Accessed 30/06 2015].

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. (2003) Relationship effectiveness for mentors: Factors associated with learning and quality. Journal of Management. Vol. 29 (4) pp. 469-486.

ALLEN, T. D., EBY, L. T. & LENTZ, E. (2006) Mentorship Behaviors and Mentorship Quality Associated With Formal Mentoring Programs: Closing the Gap Between Research and Practice. Journal of Applied Psychology. Vol. 91 (3) pp. 567-578.

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.

ALLEN, T. D., EBY, L. T., POTEET, M. L., LENTZ, E. & LIMA, L. (2004) Career benefits associated with mentoring for protégé: a meta-analysis. The Journal Of Applied Psychology. Vol. 89 (1) pp. 127-136.

ALLISON-JONES, L. L. & HIRT, J. B. (2004) Comparing the teaching effectiveness of part-time & full-time clinical nurse faculty. Nursing Education Perspectives. Vol. 25 (5) pp. 238-243.

ALTMANN, T. K. (2006) Preceptor selection, orientation, and evaluation in baccalaureate nursing education. International Journal Of Nursing Education Scholarship. Vol. 3 pp. 1-16.

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

ANDERSON, E. M. & SHANNON, A. L. (1988) Toward a Conceptualization of Mentoring. Journal of Teacher Education. Vol. 39 (1) pp. 38-42.

ANDREWS, G. J., BRODIE, D. A., ANDREWS, J. P., HILLAN, E., THOMAS, B. G., WONG, J. & 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. & CHILTON, F. (2000) Student and mentor perceptions of mentoring effectiveness. Nurse Education Today. Vol. 20 (7) pp. 555-562.

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

ANDRUSYSZYN, M., IWASIW, C. L., GOLDENBERG, D., MAWDSLEY, C., SINCLAIR, B., BEYNON, C., PARSONS, C., LETHBRIDGE, K. & BOOTH, R. (2007) Chapter 4: registered nurses mentoring baccalaureate nursing students: instrument development and preliminary testing. Annual Review of Nursing Education. Vol. 5 pp. 77-98.

APONTE, M. 2007. Mentoring: career advancement of Hispanic Army nurses. Unpublished thesis (PhD.), George Mason University.

ATKINS, S. & WILLIAMS, A. (1995) Registered nurses' experiences of mentoring undergraduate nursing students. Journal of Advanced Nursing. Vol. 21 (5) pp. 1006-1015.

BABBIE, E. R. (1990) Survey research methods. Belmont, CA: Wadsworth Publishing Company.

BANDURA, A. & MCCLELLAND, D. C. (1977) Social learning theory. Englewood Cliff, N. J.: Prentice-Hall.

BANNIGAN, K. & WATSON, R. (2009) Reliability and validity in a nutshell. Journal of clinical nursing. Vol. 18 (23) pp. 3237-3243.

BENOR, D. E. & LEVIYOF, I. (1997) The development of students' perceptions of effective teaching: the ideal, best and poorest clinical teacher in nursing. The Journal of nursing education. Vol. 36 (5) pp. 206-211.

BERGJAN, M. & HERTEL, F. (2013) Evaluating students' perception of their clinical placements—Testing the clinical learning environment and supervision and nurse teacher scale (CLES+ T scale) in Germany. Nurse education today. Vol. 33 (11) pp. 1393-1398.

BERGLUND, M., SJÖGREN, R. & EKEBERGH, M. (2012) Reflect and learn together - when two supervisors interact in the learning support process of nurse education. Journal of Nursing Management. Vol. 20 (2) pp. 152-158.

BERK, R. A., BERG, J., MORTIMER, R., WALTON-MOSS, B. & YEO, T. P. (2005) Measuring the Effectiveness of Faculty Mentoring Relationships. Academic Medicine. Vol. 80 (1) pp. 66-71.

BISHOP, V. (1998) Clinical supervision: what is going on? Results of a questionnaire... including commentary by Butterworth T. NT Research. Vol. 3 (2) pp. 141-151.

BOND, M. & HOLLAND, S. (2010) Skills of clinical supervision for nurses : a practical guide for supervisees, clinical supervisors and managers. 2nd ed. Maidenhead: McGraw Hill/Open University Press.

BOS, E., ALINAGHIZADEH, H., SAARIKOSKI, M. & KAILA, P. (2012) Validating the 'clinical learning environment, supervision and nurse teacher' CLES+T instrument in primary healthcare settings using confirmatory factor analysis. Journal of Clinical Nursing. Vol. 21 (11-12) pp. 1785-1788.

BRAUN, V. & CLARKE, V. (2006) Using thematic analysis in psychology. Qualitative research in psychology. Vol. 3 (2) pp. 77-101.

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

BROWN, S. T. (1981) Faculty and student perceptions of effective clinical teachers. Journal of Nursing Education. Vol. 20 pp. 4-15.

BROWNE, A. (1993) A conceptual clarification of respect. Journal of Advanced Nursing. Vol. 18 (2) pp. 211-217.

BROWNE, M. W., CUDECK, R., BOLLEN, K. A. & LONG, J. S. (1993) Alternative ways of assessing model fit. Sage Focus Editions. Vol. 154 pp. 136-136.

BUSCH, J. W. (1985) Mentoring in graduate schools of education: Mentors' perceptions. American Educational Research Journal. Vol. 22 (2) pp. 257-265.

BUTTERWORTH, T. & FAUGIER, J. (1992) Clinical supervision and mentorship in nursing. London: Chapman and Hall.

BUTTERWORTH, T., FAUGIER, J. & BURNARD, P. (1998) Clinical supervision and mentorship in nursing. 2ed ed.Gloucester: Nelson Thornes

BUUS, N. & GONGE, H. (2013) Translation of the Manchester Clinical Supervision Scale (MCSS) into Danish and a preliminary psychometric validation. International Journal of Mental Health Nursing. Vol. 22 (2) pp. 145-153.

BYRNE, M. W. & KEEFE, M. R. (2002) Building research competence in nursing through mentoring. Journal of Nursing Scholarship. Vol. 34 (4) pp. 391-396.

CAHILL, H. A. (1996) A qualitative analysis of student nurses' experiences of mentorship. Journal of Advanced Nursing. Vol. 24 (4) pp. 791-799.

CANADIAN NURSING ASSOCIATION. (2004) Achieving Excellence in Professional Practice – A Guide to Preceptorship andMentoring. Ottawa: CAN

CELE, S. C., GUMEDE, H. A. & KUBHEKA, B. A. (2002) An investigation of the roles and functions of nurse preceptors in the clinical areas. Curationis. Vol. 25 (1) pp. 41-51.

CHARLESTON, R. & HAPPELL, B. (2005) Attempting to accomplish connectedness within the preceptorship experience: the perceptions of mental health nurses. International Journal of Mental Health Nursing. Vol. 14 (1) pp. 54-61.

CHARLESTON, R. & HAPPELL, B. (2006) Recognising and reconciling differences: mental health nurses and nursing students' perceptions of the preceptorship relationship. Australian Journal of Advanced Nursing. Vol. 24 (2) pp. 38-43.

CHEUNG, G. W. & RENSVOLD, R. B. (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. Structural equation modeling. Vol. 9 (2) pp. 233-255.

CHOW, F. L. W. & SUEN, L. K. P. (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. & 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, H. A. (1981) The nurse's quest for a professional identity. Menlo Park, CA: Addison-Wesley.

COHEN, N. H. (1995) Mentoring Adult Learners: A Guide for Educators and Trainers. Malabar, Fl: Kreiger Publishing Co.

CRESWELL, J. & PLANOCLARK, V. (2011) Designing and conducting mixed methods research. 2011 London: SAGE Publications

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. (2009) Mentoring College Students: A Critical Review of the Literature Between 1990 and 2007. Research in Higher Education. Vol. 50 (6) pp. 525-545.

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.

CROPANZANO, R. & MITCHELL, M. S. (2005) Social Exchange Theory: An Interdisciplinary Review. Journal of Management. Vol. 31 (6) pp. 874-900.

CUTCLIFFE, J. R. & LOWE, L. (2005) A comparison of North American and European conceptualizations of clinical supervision. Issues in Mental Health Nursing. Vol. 26 (5) pp. 475-488.

DARLING, L. A. (1986) What to do about toxic mentors. Nurse Educator. Vol. 11 (2) pp. 29-30.

DARLING, L. A. W. (1984) What do nurses want in a mentor? Journal of Nursing Administration. Vol. 14 (10) pp. 42-44.

Department of Health. (1993) A Vision for the Future. Report of the Chief Nursing Officer. London: Department of Health.

Department of Health. (2012) Compassion in Practice. Nursing, Midwifery and Care Staff: Our Vision, Our Strategy. London: Department of Health.

Department of Health. (2010) Preceptorship Framework for Newly Registered Nurses, Midwives and Allied Health Professionals. London: Department of Health.

DE WITTE, N., LABEAU, S. & DE KEYZER, W. (2011) The clinical learning environment and supervision instrument (CLES): Validity and reliability of the Dutch version (CLES+NL). International Journal of Nursing Studies. Vol. 48 (5) pp. 568-572.

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.

DIMITRIADIS, K., VON DER BORCH, P., STÖRMANN, S., MEINEL, F. G., MODER, S., REINCKE, M. & FISCHER, M. R. (2012) Characteristics of mentoring relationships formed by medical students and faculty. Medical Education Online. Vol. 17 pp. 1-9.

DOBBS, K. K. (1988) The senior preceptorship as a method for anticipatory socialization of baccalaureate nursing students. Journal of Nursing Education. Vol. 27 (4) pp. 167-171.

DREHER, G. F. & ASH, R. A. (1990) A Comparative Study of Mentoring Among Men and Women in Managerial, Professional, and Technical Positions. Journal of Applied Psychology. Vol. 75 (5) pp. 539-546.

EARP, J. R. (1937) The Supervising Nurse. American Journal of Public Health and the Nations Health. Vol. 27 (8) pp. 817-818.

EBY, L. T., ALLEN, T. D., EVANS, S. C., NG, T. & DUBOIS, D. L. (2008) Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. Journal of vocational behavior. Vol. 72 (2) pp. 254-267.

EDDINS, E. E., JIE, H. & HUAPING, L. (2011) Baccalaureate Nursing Education in China : Issues and Challenges. Nursing Education Perspectives. Vol. 32 (1) pp. 30-33.

EDGAR, A. & PATTISON, S. (2011) Integrity and the moral complexity of professional practice. Nursing philosophy. Vol. 12 (2) pp. 94-106.

EDMOND, C. B. (2001) A new paradigm for practice education. Nurse Education Today. Vol. 21 (4) pp. 251-259.

ELCIGIL, A. & SARI, H. Y. (2008) Students' opinions about and expectations of effective nursing clinical mentors. Journal of Nursing Education. Vol. 47 (3) pp. 118-123.

ENB (1988) Circular 39/APS. English National Board for Nursing, Midwifery and Health Visiting, London. in SUEN, L. K. P. & CHOW, F. 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.

ENSHER, E. A. & MURPHY, S. E. (2011) The Mentoring Relationship Challenges Scale: The Impact of Mentoring Stage, Type, and Gender. Journal of Vocational Behavior. Vol. 79 (1) pp. 253-266.

FERGUSON, E. & COX, T. (1993) Exploratory factor analysis: A users' guide. International Journal of Selection and Assessment. Vol. 1 (2) pp. 84-94.

FERGUSON, M. & HAUF, B. (1973) The preceptor role: implementing student experience in community nursing. Journal of continuing education in nursing. Vol. 4 (1) pp. 12-16.

FIELD, A. (2009) Discovering statistics using SPSS. 3rd ed. London: Sage publications

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.

FLYNN, J. P. (1997) The Role of the Preceptor: A Guide for Nurse Educators and Clinicians. The Nurse Practitioner. Vol. 22 (5) pp. 200.

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.

FOX, K. C. (2010) Mentor program boosts new nurses' satisfaction and lowers turnover rate. Journal of Continuing Education in Nursing. Vol. 41 (7) pp. 311-316.

FRANCIS, R. (2013) Report of the Mid Staffordshire NHS Foundation Trust public inquiry: executive summary. London: The Stationery Office

FRANDSEN, G. M. 2003. Mentoring nursing faculty in higher education. Unpublished thesis (PhD.), Saint Louis University.

FULTON, J., BØHLER, A., HANSEN, G. S., KAUFFELDT, A., WELANDER, E., SANTOS, M. R., THORARINSDOTTIR, K. & ZIARKO, E. (2007) Mentorship: an international perspective. Nurse Education in Practice. Vol. 7 (6) pp. 399-406.

GAO, L.-L., CHAN, S. W.-C. & CHENG, B.-S. (2012) The past, present and future of nursing education in the People's Republic of China: a discussion paper. Journal of Advanced Nursing. Vol. 68 (6) pp. 1429-1438.

GASKIN, C. J. & HAPPELL, B. (2014) On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. International journal of nursing studies. Vol. 51 (3) pp. 511-521.

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.

GIESINGER, J. (2012) Respect in education. Journal of Philosophy of Education. Vol. 46 (1) pp. 100-112.

GOPEE, N. (2011) Mentoring and supervision in healthcare. London: Sage Publications

Gordis, L. (2009) Epidemiology, 4th edition. Philadelphia, Pa.: Saunders Elsevier.

GRANEHEIM, U. H. & LUNDMAN, B. (2004) Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Education Today. Vol. 24 (2) pp. 105-112.

GRAY, M. A. & SMITH, L. N. (2000) The qualities of an effective mentor from the student nurse's perspective: findings from a longitudinal qualitative study. Journal of Advanced Nursing. Vol. 32 (6) pp. 1542-1549.

GUO, Y. H. (2011) Review of nursing management in 2010 and nursing focus in 2011 in China. Chinese Nursing Management (Chinese journal). Vol. 11 (1) pp. 7-8. (Chinese language journal)

GWYN, P. G. (2011) The quality of mentoring relationships' impact on the occupational commitment of nursing faculty. Journal of Professional Nursing. Vol. 27 (5) pp. 292-298.

HAITANA, J. & BLAND, M. (2011) Building relationships: the key to preceptoring nursing students. Nursing Praxis in New Zealand. Vol. 27 (1) pp. 4-12.

HAKOJÄRVI, H.-R., SALMINEN, L. & SUHONEN, R. (2014) Health care students' personal experiences and coping with bullying in clinical training. Nurse education today. Vol. 34 (1) pp. 138-144.

Halcomb EJ; Andrew S. (2005) Triangulation as a method for contemporary nursing research. Nurse Researcher, Vol. 13(2) pp. 71-82.

HALL, L. (1997) Professional issue. Mentoring relationships of New Zealand nurses: am empirical study (part 1). Collegian. Vol. 4 (4) pp. 28.

HALL, L. (1998) Mentoring relationships of New Zealand nurses: results of an empirical study (part 2). Collegian. Vol. 5 (1) pp. 32-35.

HALLIN, K. & DANIELSON, E. (2009) Being a personal preceptor for nursing students: Registered Nurses' experiences before and after introduction of a preceptor model. Journal of Advanced Nursing. Vol. 65 (1) pp. 161-174.

HALLIN, K. & DANIELSON, E. (2010) Preceptoring nursing students: registered nurses' perceptions of nursing students' preparation and study approaches in clinical education. Nurse Education Today. Vol. 30 (4) pp. 296-302.

HAMILTON, E. M., MURRAY, M. K., LINDHOLM, L. H. & MYERS, R. E. (1989) Effects of mentoring on job satisfaction, leadership behaviors, and job retention

of new graduate nurses. Journal of Nursing Staff Development. Vol. 5 (4) pp. 159-165.

HARDYMAN, R. & HICKEY, G. (2001) What do newly-qualified nurses expect from preceptorship? Exploring the perspective of the preceptee. Nurse Education Today. Vol. 21 (1) pp. 58-64.

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.

HEFFERNAN, C., HEFFERNAN, E., BROSNAN, M. & BROWN, G. (2009) Evaluating a preceptorship programme in South West Ireland: perceptions of preceptors and undergraduate students. Journal of Nursing Management. Vol. 17 (5) pp. 539-549.

HEISLER, A. (1940) Supervision of Public Health Nurses: A Continuous Educational Program—From the Point of View of the Rural Agency. American Journal of Public Health and the Nations Health. Vol. 30 (7) pp. 749-754.

HENRIKSEN, N., NORMANN, H. K. & SKAALVIK, M. W. (2012) Development and Testing of the Norwegian Version of the Clinical Learning Environment, Supervision and Nurse Teacher (CLES+T) Evaluation Scale. International Journal of Nursing Education Scholarship. Vol. 9 (1) pp. 1-15.

HILLI, Y., SALMU, M. & JONSÉN, E. (2014) Perspectives on good preceptorship A matter of ethics. Nursing ethics. Vol. 21 (5) pp. 565-575.

HOGAN, T. P. & CANNON, B. (2003) Psychological testing: A practical introduction. New York*:* Wiley

HOHMAN, J. (1979) Nurse mentor system cuts costs, boosts quality of patient care. Hospitals. Vol. 53 (1) pp. 93-94.

HOU, X., ZHU, D. & ZHENG, M. (2011) Clinical Nursing Faculty Competence Inventory–development and psychometric testing. Journal of advanced nursing. Vol. 67 (5) pp. 1109-1117.

HOU, X. J., ZHU, D. & JIANG, X. L. (2005) Constructing an Evaluation Model for Core Competence of Clinical Nursing Teachers. Chinese Journal of Evidence-Based Medicine. Vol. 15 (11) pp. 822-828. (Chinese language journal)

HSIEH, H. & SHANNON, S. E. (2005) Three approaches to qualitative content analysis. Qualitative Health Research. Vol. 15 (9) pp. 1277-1288.

HU, C., PELLEGRINI, E. K. & SCANDURA, T. A. (2011a) Measurement invariance in mentoring research: A cross-cultural examination across Taiwan and the U.S. Journal of Vocational Behavior. Vol. 78 (2) pp. 274-282.

HU, L. T. & BENTLER, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal. Vol. 6(1) pp. 1-55.

HU, L., CHU, Y. & LI, X. (2011b) Design of the Environment Evaluation Scale for Clinical Nursing Internship: testing of reliability and validity. Journal of Nursing Science. Vol. 9 pp. 002-004. (Chinese language journal)

HUANG, B. Z. & WU, X. L. (1999) Effectiveness evaluation of mentoring New nurses. Journal of Chinese nursing. Vol. 34 (2) pp. 95-97. (Chinese language journal)

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.

HURLEY, C. & SNOWDEN, S. (2008) Mentoring in times of change. Nursing in Critical Care. Vol. 13 (5) pp. 269-275.

HYRKÄS, K., APPELQVIST-SCHMIDLECHNER, K. & PAUNONEN-ILMONEN, M. (2003) Translating and validating the Finnish version of the Manchester Clinical Supervision Scale. Scandinavian Journal of Caring Sciences. Vol. 17 (4) pp. 358-364.

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

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

JOHANSSON, U., KAILA, P., AHLNER-ELMQVIST, M., LEKSELL, J., ISOAHO, H. & SAARIKOSKI, M. (2010) Clinical learning environment, supervision and nurse teacher evaluation scale: psychometric evaluation of the Swedish version. Journal of Advanced Nursing. Vol. 66 (9) pp. 2085-2093.

JOKELAINEN, M., JAMOOKEEAH, D., TOSSAVAINEN, K. & TURUNEN, H. (2013) Finnish and British mentors' conceptions of facilitating nursing students' placement learning and professional development. Nurse Education in Practice. Vol. 13 (1) pp. 61-67.

JOKELAINEN, M., TURUNEN, H., TOSSAVAINEN, K., JAMOOKEEAH, D. & COCO, K. (2011) A systematic review of mentoring nursing students in clinical placements. Journal of Clinical Nursing. Vol. 20 (19/20) pp. 2854-2867.

JONES, A. M. I. 2000. Students' perceptions of the preceptorship clinical experience. Unpublished thesis (PhD.), Rush University.

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

JONES, R. & BROWN, D. (2011) The mentoring relationship as a complex adaptive system: Finding a model for our experience. Mentoring & Tutoring: Partnership in Learning. Vol. 19 (4) pp. 401-418.

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

KANTER, R. M. (1977) Men and Women of the Corporation. New York: Basic books

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.

KENNY, A. J. (2005) Interaction in cyberspace: an online focus group. Journal of Advanced Nursing. Vol. 49 (4) pp. 414-422.

KIM, K. H. (2007) Clinical competence among senior nursing students after their preceptorship experiences. Journal of Professional Nursing. Vol. 23 (6) pp. 369-375.

KLINE, R. (2005) Principles and practice of structural equation modeling. 2nd ed. New York: Guilford Press.

KNOX, J. E. & MOGAN, J. (1985) Important clinical teacher behaviours as perceived by university nursing faculty, students and graduates. Journal of Advanced Nursing. Vol. 10 (1) pp. 25-30.

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

KOLB, D. A. (1984) Experiential learning: Experience as the source of learning and development. New Jersey: Prentice-Hall

KOMARATAT, S. & OUMTANEE, A. (2009) Using a mentorship model to prepare newly graduated nurses for competency. Journal of continuing education in nursing. Vol. 40 (10) pp. 475-480.

KOTZABASSAKI, S., PANOU, M., DIMOU, F., KARABAGLI, A., KOUTSOPOULOU, B. & IKONOMOU, U. (1997) Nursing students' and faculty's perceptions of the characteristics of 'best'and 'worst'clinical teachers: A replication study. Journal of Advanced Nursing. Vol. 26 (4) pp. 817-824.

KRAM, K. E. (1983) Phases of the mentor relationship. Academy of Management journal. Vol. 26 (4) pp. 608-625.

KRAM, K. E. & ISABELLA, L. A. (1985) Mentoring alternatives: the role of peer relationships in career development. Academy of Management Journal. Vol. 28 (1) pp. 110-132.

KRISTOFFERZON, M.-L., MÅRTENSSON, G., MAMHIDIR, A.-G. & LÖFMARK, A. (2013) Nursing students' perceptions of clinical supervision: the contributions of preceptors, head preceptors and clinical lecturers. Nurse education today. Vol. 33 (10) pp. 1252-1257.

KRUEGER, R. A. & CASEY, M. A. (2009) Focus groups: A practical guide for applied research. 4th ed. London:Sage Publications

LAMBERT, V. & GLACKEN, M. (2005) Clinical education facilitators: a literature review. Journal of Clinical Nursing. Vol. 14 (6) pp. 664-673.

LANDIS, J. R. & KOCH, G. G. (1977) The measurement of observer agreement for categorical data. Biometrics. Vol. 33 (1) pp. 159-174.

LANDMARK, B. T., HANSEN, G. S., BJONES, I. & BØHLER, A. (2003) Clinical supervision -- factors defined by nurses as influential upon the development of competence and skills in supervision. Journal of Clinical Nursing. Vol. 12 (6) pp. 834-841.

LEE, C. & CARMEN MONTIEL, E. (2011) The Correlation of Mentoring and Job Satisfaction: A Pilot Study of Mental Health Professionals. Community Mental Health Journal. Vol. 47 (4) pp. 482-487.

LEE, T., TZENG, W., LIN, C. & YEH, M. (2009) Effects of a preceptorship programme on turnover rate, cost, quality and professional development. Journal of Clinical Nursing. Vol. 18 (8) pp. 1217-1225.

LEVINSON, D. J. (1986) A conception of adult development. American psychologist. Vol. 41 pp. 3-13.

LEVINSON, D. J., DARROW, C. M., KLEIN, E. G., LEVINSON, M. H. & MCKEE, B. (1978) The seasons of a man's life. New York: Ballentine.

LI, F. & HAN, J. (2010) Thinking on training of master degree postgraduate in clinical nursing profession. Chinese Nursing Research. Vol. 24 pp. 266-267. (Chinese language journal)

LIGHTFOOT, C. A. (2001). The impact of academic, social and mentoring experiences on the persistence of minority and non-minority college students. Unpublished thesis (PhD.), University of Houston.

LIGTVOET, R., VAN DER ARK, L. A., TE MARVELDE, J. M. & SIJTSMA, K. (2010) Investigating an invariant item ordering for polytomously scored items. Educational and Psychological Measurement. Vol. 70(4) pp.578-595

LINDQUIST, I., JOHANSSON, I. & SEVERINSSON, E. (2012) Evaluation of process-oriented supervision of student nurses: A Swedish case study. Nursing & Health Sciences. Vol. 14 (1) pp. 2-7.

LIU, C. (2014) Status analysis and strategy of clinical nursing teaching in primary hospital. Chinese Community Doctors. Vol. 30 (13) pp. 163-164. (Chinese language journal)

LIU, L. X., DAI, H. S. & GAO, L. L. (2012) An application of mentoring program in nursing teaching. China higher education. Vol. 11 pp. 50-51. (Chinese language journal)

LIU, M., CHEN, W.-J., LAN, Y.-T., WANG, Y. & LI, Y.-E. (2011) Effective clinical teaching behavioral characteristics as perceived by clinical preceptors:a qualitative study. Chinese journal of nursing education. Vol. 8 (9) pp. 387-389. (Chinese language journal)

LÖFMARK, A., THORKILDSEN, K., RÅHOLM, M.-B. & KARIN NATVIG, G. (2012) Nursing students' satisfaction with supervision from preceptors and teachers during clinical practice. Nurse Education in Practice. Vol. 12 (3) pp. 164-169.

LUHANGA, F. L., BILLAY, D., GRUNDY, Q., MYRICK, F. & YONGEYY, O. (2010) The one-to-one relationship: is it really key to an effective preceptorship experience? A review of the literature. International Journal of Nursing Education Scholarship. Vol. 7 (1) pp.1-14.

LYNN, M. R. (1986) Determination and quantification of content validity. Nursing research. Vol. 35 (6) pp. 382-386.

LYTH, G. M. (2000) Clinical supervision: a concept analysis. Journal of Advanced Nursing. Vol. 31 (3) pp. 722-729.

MA, Y. P., KANG, J. F. & TAN, C. H. (2007) Research progress on tutor system of college students in medical and nursing fields. Chinese nursing research. Vol. 21 (12B) pp. 3203-3204. (Chinese language journal)

MA, Y. P. & TAN, C. H. (2010) The construction of appraisal system for core capability of nursing undergraduates tutors. Journal of Nursing Science. Vol. 25 (7) pp. 67-69. (Chinese language journal)

MÃ¥RTENSSON, G., ENGSTRÃ¶M, M., MAMHIDIR, A.-G. & KRISTOFFERZON, M.-L. (2013) What are the structural conditions of importance to preceptors' performance? Nurse Education Today. Vol. 33 (5) pp. 444-449.

MACDONALD, J. (2002) Clinical supervision: a review of underlying concepts and developments. Australian & New Zealand Journal of Psychiatry. Vol. 36 (1) pp. 92-98.

MAO, X., LIU, W. & CHEN, W. (2014) Survey of nursing students' need studying in surgical department in Traditional Chinese Medicine hospital. Journal of Traditional Chinese Medicine Management. Vol. 8 pp. 1225-1226. (Chinese language journal)

MATTHEWS, B. & ROSS, L. (2010) Research methods : a practical guide for the social sciences. Harlow : Pearson Longman

MCCLOUGHEN, A., O'BRIEN, L. & JACKSON, D. (2006) Positioning mentorship within Australian nursing contexts - a literature review. Contemporary Nurse: A Journal for the Australian Nursing Profession. Vol. 23 (1) pp. 120-134.

MCDOWELL, I. (2006) Measuring health: a guide to rating scales and questionnaires. Oxford: Oxford University Press.

MCGRAW, K. O. & WONG, S. P. (1996) Forming inferences about some intraclass correlation coefficients. Psychological methods. Vol. 1 (1) pp. 30-46.

MCKENNA, L., MCCALL, L. & WRAY, N. (2010) Clinical placements and nursing students' career planning: A qualitative exploration. International journal of nursing practice. Vol. 16 (2) pp. 176-182.

MEIJER, R. R., NIESSEN, A. S. M. & TENDEIRO, J. N. (2015) A Practical Guide to Check the Consistency of Item Response Patterns in Clinical Research Through Person-Fit Statistics: Examples and a Computer Program. Assessment. Vol. pp.1-26

MENO, K. M., KEAVENY, B. M. & O'DONNELL, J. M. (2003) Education news. Mentoring in the operating room: a student perspective. AANA Journal. Vol. 71 (5) pp. 337-341.

MILLS, J. E., FRANCIS, K. L. & BONNER, A. (2005) Mentoring, clinical supervision and preceptoring: clarifying the conceptual definitions for Australian rural nurses. A review of the literature. Rural & Remote Health. Vol. 5 (3) pp. 1-10.

MORGAN, D. L. (1997) Focus groups as qualitative research. Newbury Park, CA: Sage

MORGAN, D. L., KRUEGER, R. A. & KING, J. 1997. Focus group kit: Volumes 1-6. Thousand Oaks, CA: Sage Publications.

MORTON -COO PER, A. & PALMER, A. (2000) Mentoring, Preceptorship and Clinical Supervision: A Guide to Professional Roles in Clinical Practice, 2nd ed. Oxford: Blackwell Science

MOSELEY, L. G. & DAVIES, M. (2008) What do mentors find difficult? Journal of Clinical Nursing. Vol. 17 (12) pp. 1627-1634.

MURRAY, P. J. (1997) Using virtual focus groups in qualitative research. Qualitative health research. Vol. 7 (4) pp. 542-549.

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 (14) pp. 1834-1842.

MYRICK, F. (1988) Preceptorship -- is it the answer to the problems in clinical teaching? Journal of Nursing Education. Vol. 27 (3) pp. 136-138.

MYRICK, F. & BARRETT, C. (1994) Selecting clinical preceptors for basic baccalaureate nursing students: A critical issue in clinical teaching. Journal of Advanced Nursing. Vol. 19 (1) pp. 194-198.

MYRICK, F. & YONGE, O. (2002) Preceptor questioning and student critical thinking. Journal of Professional Nursing. Vol. 18 (3) pp. 176-181.

MYRICK, F. & YONGE, O. (2004) Enhancing critical thinking in the preceptorship experience in nursing education. Journal of Advanced Nursing. Vol. 45 (4) pp. 371-380.

NEARY, M. (2000) Supporting students' learning and professional development through the process of continuous assessment and mentorship. Nurse Education Today. Vol. 20 (6) pp. 463-474.

NETHERWOOD, M. & DERHAM, R. (2014) Interprofessional education: merging nursing, midwifery and CAM. British Journal of Nursing. Vol. 23 (13) pp. 740-743.

NOE, R. A. (1988) An investigation of the determinants of successful assigned mentoring relationships. Personnel Psychology. Vol. 41 (3) pp. 457-479.

NORA, A. & CRISP, G. (2007) Mentoring students: Conceptualizing and validating the multi-dimensions of a support system. Journal of College Student Retention: Research, Theory and Practice. Vol. 9 (3) pp. 337-356.

Nursing and Midwifery Council. (2010) Standards for pre-registration nursing education. London: NMC

Nursing and Midwifery Council.(2008) Standards to Support Learning and Assessment in Practice. London: NMC

Nursing and Midwifery Council. (2015) The Code Professional standards of practice and behaviour for nurses and midwives. London: NMC.

OHRLING, K. & HALLBERG, I. R. (2000a) Nurses' lived experience of being a preceptor. Journal Of Professional Nursing: Official Journal Of The American Association Of Colleges Of Nursing. Vol. 16 (4) pp. 228-239.

OHRLING, K. & HALLBERG, I. R. (2000b) Student nurses' lived experience of preceptorship. Part 1--in relation to learning. International Journal Of Nursing Studies. Vol. 37 (1) pp. 13-23.

OHRLING, K. & HALLBERG, I. R. (2000c) Student nurses' lived experience of preceptorship. Part 2--the preceptor-preceptee relationship. International Journal Of Nursing Studies. Vol. 37 (1) pp. 25-36.

OMANSKY, G. L. (2010) Staff nurses' experiences as preceptors and mentors: an integrative review. Journal of Nursing Management. Vol. 18 (6) pp. 697-703.

Oxford dictionaries. (2015) Definition of behaviour in English. [Online]. Available: http://www.oxforddictionaries.com/definition/english/behaviour [Accessed 30/06 2015]

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.

PEIRCE, A. G. (1991) Preceptorial students' view of their clinical experience. Journal of Nursing Education. Vol. 30 (6) pp. 244-250.

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.

POLIT, D. F. & BECK, C. T. (2006) The content validity index: are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health. Vol. 29 (5) pp. 489-497.

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.

PRICE, P. J. (2006) Final year baccalaureate nursing students' perceptions of a perceptor model of clinical teaching. Unpublished thesis (PhD.), University of Toronto.

PROCTOR, B. (1986) Supervision: a co-operative exercise in accountability. Enabling and Ensuring. Leicester: Leicester National Youth Bureau and Council for Education and Training in Youth and Community Work. Vol. pp. 21-3.

RAGINS, B. R., COTTON, J. L. & MILLER, J. S. (2000) Marginal mentoring: The effects of type of mentor, quality of relationship, and program design on work and career attitudes. Academy of Management Journal. Vol. 43 (6) pp. 1177-1194.

RAGINS, B. R. & MCFARLIN, D. B. (1990) Perceptions of mentor roles in cross-gender mentoring relationships. Journal of Vocational Behavior. Vol. 37 (3) pp. 321-339.

RAN, Q.H. & ZOU, S. K. (2015) The causes of degrading of professional ethics among health professionals and the strategies to deal with it. Health medicine research and practice. Vol. 12 (2) pp. 91-93

REES, C. E., MONROUXE, L. V. & MCDONALD, L. A. (2015) 'My mentor kicked a dying woman's bed…'Analysing UK nursing students''most memorable'professionalism dilemmas. Journal of advanced nursing. Vol. 71 (1) pp. 169-180.

RICHARD, L. W. (1995) The relationship between organizational culture and the frequency, quality and array of mentoring activities among nurse executive mentors. Unpublished thesis (PhD.), University of Colorado.

RIDLEY, M. J., LASCHINGER, H. K. S. & GOLDENBERG, D. (1995) The effect of a senior preceptorship on the adaptive competencies of community college nursing students. Journal of Advanced Nursing. Vol. 22 (1) pp. 58-65.

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 (4) pp. 473-494.

ROSE, G. L. (2005) Group Differences in Graduate Students' Cconcepts of The Ideal Mentor. Research in Higher Education. Vol. 46 (1) pp. 53-80.

SAARIKOSKI, M., ISOAHO, H., LEINO-KILPI, H. & WARNE, T. (2005) Validation of the clinical learning environment and supervision scale. International Journal of Nursing Education Scholarship. Vol. 2 (1) pp. 1-16.

SAARIKOSKI, M., ISOAHO, H., WARNE, T. & LEINO-KILPI, H. (2008) The nurse teacher in clinical practice: Developing the new sub-dimension to the Clinical Learning Environment and Supervision (CLES) scale. International Journal of Nursing Studies. Vol. 45 (8) pp. 1233-1237.

SAARIKOSKI, M. & LEINO-KILPI, H. (2002) The clinical learning environment and supervision by staff nurses: developing the instrument. International Journal of Nursing Studies. Vol. 39 (3) pp. 259-267.

SAARIKOSKI, M., MARROW, C., ABREU, W., RIKLIKIENE, O. & OZBICAKÇI, S. (2007) Student nurses' experience of supervision and Mentorship in clinical practice: a cross cultural perspective. Nurse Education in Practice. Vol. 7 (6) pp. 407-415.

SALAMI, S. O. (2008) Psychosocial factors as predictors of mentoring among nurses in southwestern Nigeria. Journal of Workplace Learning. Vol. 20 (5) pp. 348-363.

SALMONS, J. (2011) Cases in online interview research. Thousand Oaks, CA:Sage Publications.

SANDS, R. G. & PARSON, L. A. (1991) Faculty mentoring faculty in a public university. Journal of Higher Education. Vol. 62 (2) pp. 174-193.

SANDSTRÖM, B., BORGLIN, G., NILSSON, R. & WILLMAN, A. (2011) Promoting the Implementation of Evidence-Based Practice: A Literature Review Focusing on the Role of Nursing Leadership. Worldviews on Evidence-Based Nursing. Vol. 8 (4) pp. 212-223.

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.

SCALES, F. S., ALVERSON, E. & HARDER, D. L. (1993) The effect of a preceptorship on nursing performance. Nursing Connections. Vol. 6 (2) pp. 45-54.

SCANDURA, T. A. (1992) Mentorship and career mobility: An empirical investigation. Journal of Organizational Behavior. Vol. 13 (2) pp. 169-174.

SCANDURA, T. A. & RAGINS, B. R. (1993) The effects of sex and gender role orientation on mentorship in male-dominated occupations. Journal of Vocational Behavior. 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.

SEIBEL, M. (2014) For us or against us? Perceptions of faculty bullying of students during undergraduate nursing education clinical experiences. Nurse education in practice. Vol. 14 (3) pp. 271-274.

SEVERINSSON, E. & SAND (2010) Evaluation of the clinical supervision and professional development of student nurses. Journal of Nursing Management. Vol. 18 (6) pp. 669-677.

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.

SHROUT, P. E. & FLEISS, J. L. (1979) Intraclass correlations: uses in assessing rater reliability. Psychological bulletin. Vol. 86 (2) pp. 420.

SIJTSMA, K., EMONS, W. H., BOUWMEESTER, S., NYKL ĺCEK, I. & ROORDA, L. D. (2008) Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). Quality of Life Research. Vol. 17 (2) pp. 275-290.

SIJTSMA, K. & JUNKER, B. W. (1996) A survey of theory and methods of invariant item ordering. British Journal of Mathematical and Statistical Psychology. Vol. 49 (1) pp. 79-105.

SLOAN, G. & GRANT, A. (2012) A rationale for a clinical supervision database for mental health nursing in the UK. Journal of Psychiatric & Mental Health Nursing. Vol. 19 (5) pp. 466-473.

SMEDLEY, A., MOREY, P. & RACE, P. (2010) Enhancing the Knowledge, Attitudes, and Skills of Preceptors: An Australian Perspective. Journal of Continuing Education in Nursing. Vol. 41 (10) pp. 451-461.

SMITH-STRØM, H., OTERHALS, K., RUSTAD, E. C. & LARSEN, T. (2012) Culture crash regarding nursing students' experience of implementation of EBP in clinical practice. Nordic Journal of Nursing Research & Clinical Studies / V ård i Norden. Vol. 32 (4) pp. 55-59.

SONG, C. Y. & SHEN, N. (2004) A scale for appraisal of clinical nursing instructors. Chinese Journal of Nursing. Vol. 39 (12) pp. 888-890.

SPENCER, R. & WINKELS, R. (1978) Treatment of diminutive mucosal lesions of the terminal colon by nurse endoscopists. Mayo Clinic Proceedings. 391-392.

SPOUSE, J. (2001) Bridging theory and practice in the supervisory relationship: a sociocultural perspective. Journal Of Advanced Nursing. Vol. 33 (4) pp. 512-522.

STEIN, R. F. (1964) THE NURSING INSTRUCTOR AS MODEL AND PRECEPTOR. Mental hospitals. Vol. 15 pp. 28-31.

STOCHL, J., JONES, P. B. & CROUDACE, T. J. (2012) Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. BMC Medical Research Methodology. Vol. 12 (1) pp. 74-89.

STRAAT, J. H., VAN DER ARK, L. A. & SIJTSMA, K. (2014) Minimum Sample Size Requirements for Mokken Scale Analysis. Educational and Psychological Measurement. Vol. 74 (5) pp. 809-822.

STREINER, D. L. (2010) Measure for measure: new developments in measurement and item response theory. Canadian Journal Of Psychiatry. Revue Canadienne De Psychiatrie. Vol. 55 (3) pp. 180-186.

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

SUEN, L. K. P. & CHOW, F. 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.

SULLIVAN, E. A. (2012) Critical thinking in clinical nurse education: Application of Paul's model of critical thinking. Nurse education in practice. Vol. 12 (6) pp. 322-327.

SUN, Y. L., LYU, Z. F., WANG, X. Y., HUANG, H., LI, L., WANG, Y. & CHEN, S. J. (2014) The Moral Status of Rural Health Personnel at the Township and Village Level. Chinese Medical Ethics. Vol. 27 (4) pp. 555-557.(Chinese language journal)

TENDEIRO, J. N. & TENDEIRO, M. J. N. (2014) Package 'PerFit'. [Online]. Available: http://cran.csiro.au/web/packages/PerFit/PerFit.pdf. [Accessed 30/06 2015]

THOMPSON, B. (2004) Exploratory and confirmatory factor analysis: Understanding concepts and applications. Washington, DC, US: American Psychological Association

TYREMAN, S. (2011) Integrity: is it still relevant to modern healthcare? Nursing Philosophy. Vol. 12 (2) pp. 107-118.

UDLIS, K. A. (2008) Preceptorship in undergraduate nursing education: an integrative review. Journal of Nursing Education. Vol. 47 (1) pp. 20-29.

UKCC (1996) Position statement on clinical supervision for nursing and health visiting. London: UKCC. The NMC replaced the UKCC and four National Boards in April 2002

USHER, K., NOLAN, C., RESER, P., OWENS, J. & TOLLEFSON, J. (1999) An exploration of the preceptor role: preceptors' perceptions of benefits, rewards, supports and commitment to the preceptor role. Journal of Advanced Nursing. Vol. 29 (2) pp. 506-514.

VAN EPS, M. A., COOKE, M., CREEDY, D. K. & WALKER, R. (2006) Student evaluations of a year-long mentorship program: a quality improvement initiative. Nurse Education Today. Vol. 26 (6) pp. 519-524.

VANDENBERG, R. J. & LANCE, C. E. (2000) A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational research methods. Vol. 3 (1) pp. 4-70.

VOGT, D. S., KING, D. W. & KING, L. A. (2004) Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychological assessment. Vol. 16 (3) pp. 231-243.

WALLEN, G. R., MITCHELL, S. A., MELNYK, B., FINEOUT-OVERHOLT, E., MILLER-DAVIS, C., YATES, J. & HASTINGS, C. (2010) Implementing evidence-based practice: effectiveness of a structured multifaceted mentorship programme. Journal of Advanced Nursing. Vol. 66 (12) pp. 2761-2771.

WALSH, P. & MCNEIL, E. (1951) Field Training Program in Public Health Nursing Supervision. American Journal of Public Health and the Nations Health. Vol. 41 (7) pp. 846-851.

WANG, J. P. (2008) Comparative study of the effective characteristics of clinical nursing teachers as accepted by teachers and students. Shanghai Nursing Vol. 8 (1) pp. 5-9. (Chinese language journal).

WANG, R.M.,SHI, N.K., ZHAO, Y.. (2015) Application of medical-nursing students collaboration combined with scenario simulation teaching in the operating room nursing. Chinese Journal of Nursing. 50(3) pp.336-339 (Chinese language journal)

WANG, S. X., GUO, X. L., ZHANG, H. & ZHANG, L. (2011) Cultivation of managerial expertise of newly-appointed head nurses by using preceptorship system. Journal of Nursing Science . Vol. 126 (111) pp. 47-49. (Chinese language journal).

WARNE, T., JOHANSSON, U., PAPASTAVROU, E., TICHELAAR, E., TOMIETTO, M., DEN BOSSCHE, K. V., MORENO, M. F. V. & SAARIKOSKI, M. (2010) An exploration of the clinical learning experience of nursing students in nine European countries. Nurse Education Today. Vol. 30 (8) pp. 809-815.

WATSON, N. A. (1999) Mentoring today -- the students' views. An investigative case study of pre-registration nursing students' experiences and perceptions of mentoring in one theory/practice module of the Common Foundation Programme on a Project 2000 course. Journal of Advanced Nursing. Vol. 29 (1) pp. 254-262.

WATSON, R., DEARY, I., THOMPSON, D. & LI, G. (2008) A study of stress and burnout in nursing students in Hong Kong: a questionnaire survey. International journal of nursing studies. Vol. 45 (10) pp. 1534-1542.

WATSON, R. & THOMPSON, D. R. (2006) Use of factor analysis in Journal of Advanced Nursing: literature review. Journal of advanced nursing. Vol. 55 (3) pp. 330-341.

WATSON, R., VAN DER ARK, L. A., LIN, L. C., FIEO, R., DEARY, I. J. & MEIJER, R. R. (2012) Item response theory: how Mokken scaling can be used in clinical practice. Journal of clinical nursing. Vol. 21 (19) pp. 2736-2746.

WATSON, R., WANG, W. & THOMPSON, D. R. (2014) Violations of local stochastic independence exaggerate scalability in Mokken scaling analysis of the Chinese Mandarin SF-36. Health and quality of life outcomes. Vol. 12 (149) pp. 1-10.

WATSON, R., YANHUA, C., IP, M. Y., SMITH, G. D., WONG, T. K. & DEARY, I. J. (2013) The structure of stress: Confirmatory factor analysis of a Chinese version of the stressors in Nursing Students Scale (SINS). Nurse education today. Vol. 33 (2) pp. 160-165.

WEBB, C. & SHAKESPEARE, P. (2008) Judgements about mentoring relationships in nurse education. Nurse Education Today. Vol. 28 (5) pp. 563-571.

WENG, R.-H., HUANG, C.-Y., TSAI, W.-C., CHANG, L.-Y., LIN, S.-E. & 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 (1) pp. 1-9.

WIDAMAN, K. F. (1993) Common factor analysis versus principal component analysis: Differential bias in representing model parameters? Multivariate Behavioral Research. Vol. 28 (3) pp. 263-311.

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

WILHELMSSON, M., SVENSSON, A., TIMPKA, T. & FARESJÖ, T. (2013) Nurses' views of interprofessional education and collaboration: A comparative study of recent graduates from three universities. Journal of interprofessional care. Vol. 27 (2) pp. 155-160.

WILSON, A. H., SANNER, S. & MCALLISTER, L. E. (2010) An evaluation study of a mentoring program to increase the diversity of the nursing workforce. Journal of Cultural Diversity. Vol. 17 (4) pp. 144-150.

WINSTANLEY, J. & WHITE, E. (2003) Clinical supervision: models, measures and best practice. Nurse Researcher. Vol. 10 (4) pp. 7-38.

WINSTANLEY, J. & WHITE, E. (2011) The MCSS-26: Revision of the Manchester Clinical Supervision Scale; Using the Rasch Measurement Model. Journal of Nursing Measurement. Vol. 19 (3) pp. 160-178.

WONG, F. K. Y. & ZHAO, Y. (2012) Nursing education in China: past, present and future. Journal of Nursing Management. Vol. 20 (1) pp. 38-44.

World Health Organization. 1988. Learning together to work together for health: Technical Report series 769. Geneva: WHO

World Health Organization. 2010. Framework for action on Interprofessional Education and Collaborative Practice. Geneva: WHO

XU, L. L. & KANG, Y. C. (2010) A survey of clinical nursing teachers' mentoring capability. Journal of Qilu nursing. Vol. 16 (21) pp. 45-46.(Chinese language journal)

YARDLEY, S., TEUNISSEN, P. W. & DORNAN, T. (2012) Experiential learning: transforming theory into practice. Medical teacher. Vol. 34 (2) pp. 161-164.

YE, H. F., WANG, R. & FU, R. (2013) Study on evaluation of practical nursing students on competency of clinical nursing teachers. Chinese nursing research Vol. 27 (8A) pp. 2328-2329. (Chinese language journal)

YEGDICH, T. & CUSHING, A. (1998) An historical perspective on clinical supervision in nursing. Australian & New Zealand Journal of Mental Health Nursing. Vol. 7 (1) pp. 3-24.

YOU, L.-M., AIKEN, L. H., SLOANE, D. M., LIU, K., HE, G.-P., HU, Y., JIANG, X.-L., LI, X.-H., LI, X.-M., LIU, H.-P., SHANG, S.-M., KUTNEY-LEE, A. & SERMEUS, W. (2013) Hospital nursing, care quality, and patient satisfaction: Cross-sectional surveys of nurses and patients in hospitals in China and Europe. International Journal of Nursing Studies. Vol. 50 (2) pp. 154-161.

YOU, L.-M., KE, Y.-Y., ZHENG, J. & WAN, L.-H. (2015) The development and issues of nursing education in China: A national data analysis. Nurse Education Today. Vol. 35 (2) pp. 310-314.

ZEY, M. G. (1991) The mentor connection: Strategic alliances in corporate life Transaction Publishers

ZHANG, Z. & LEI, Z. Y. (2012) Reshape doctors' professional ethics, save doctor-patient confidence crisis. Medicine and law. Vol. pp. 89-91.(Chinese laguage journal)

ZHANG, T., CHEN, H. (2014) Practice of Interprofessional Education Program in Medical Undergraduate Education. West China Medical Journal. 29（2）pp. 353-355. (Chinese language journal)

Appendices

Appendix 1. Psychosocial support dimension, subcategories, codes and references

Dimension 1: Psychosocial support	
Relationship building	**Support and encouragement**
• Welcome and accept: warm, friendly, genuine (6-9, 29, 38) • Smile (14) • Know students (24, 42) • Treat equally (7, 14, 21, 28, 31-33, 39-40) • Trust (8, 9, 31-34) • Allow autonomy (24, 29) • Gradually withdraw supervision (14) • Respect, call by name (7-9, 25, 31-33) • Share informal activities, ideas and experiences (9, 26, 36, 39-40) • Befriend and allow sensitive questions (7-9, 36, 39-40) • Develop effective relationship: constant contact, mutual approval (1, 7, 10, 14, 20, 22, 26, 31-33, 35) • Relationship four stages (26) • Understanding (13) • Attitude toward students (15) • Be open and honest (41) • Preceptor consistency (18) • Orientate to the environment (8-9, 37)	• Encourage, inspire, spur (4, 13, 22, 24, 26-27, 37) • Clarify student's role (7, 29, 34) • Listen attentively (21-21, 24, 26, 28) • Show interests and positive attitude, commitment to students learning and students (7-11, 17, 20, 22, 24-25, 38) • Approachable (4, 7, 22, 27, 39-40) • Accessible (4, 22) • Quality time together (14) • Amount of time together (7, 17, 19) • Sort out student's problems and worries (1, 10, 14, 39-40) • Reduce anxiety (39-40) • Support, help, assist, advocate (1, 3-5, 10-11, 20, 22, 24, 26) • Work on same shift (7) • Protect students (16) • Provide a safe place for better study (16) • Meet students' need (23) • Being with students (45) • Take responsibility of students' learning (45) • Affirm and confirm (4, 31-33, 45) • Work in partnership (6) • Be supportive yet challenging relationship (41)

Appendix 2. Facilitating learning dimension, subcategories, codes and references

Dimension 2: Facilitating learning		
Planning and organizing	**Teaching and guiding**	**Feedback and assessment**
• Plan educational activities (6, 8-11, 23, 26, 42-43) • Arrange activities from observation to hands on (14, 24, 29, 38) • Environment maintaining (21, 28 , 31-33, 29) • Discuss learning objectives and give clear expectation (21, 24, 25, 26 ,28, 31-33) • Set goals and objectives (34) • Provide supportive network (4, 8, 9, 21, 28) • Be prepared for student, know program (19) • Safe space for learning (9, 18, 21, 28, 45) • Connect students with patients (12, 16) • Assign tasks based on student's ability (34) • Negotiate learning aims (46) • Creating a relaxed environment (12)	• Promote theory and practice integration (13, 17, 21-22, 25, 28) • Be challenger (1, 4, 8-10, 14, 24, 29) • Promote reflective and critical thinking (8-9, 17, 21, 24, 28) • Share knowledge and skills (11, 13, 22, 24); • Recommend source of reference (8-9, 27, 39-40); • Answer questions (4, 13, 25) • Give advice (8, 9, 14, 25, 39-40) • Show and teach clinical skills (1-3, 8-10, 13, 21, 45-46) • Be a supervisor, coach, teacher, guider and facilitator (1-6, 10, 22) • Possess good teaching ability (11, 15, 19, 28, 31-33, 46) • Teaching motivation (19) • Involve students in all activities (29) • Appreciation of education (5) • Contribute to achieve the goal (41) • Direct students in learning (44)	• Feedback should be concrete, specific, skill and performance focused, continuous, consistency, timely, honest (1-3, 7, 10, 14, 24-25, 28, 30, 38) • Feedback for progress and positively (14, 38) • Negotiate learning experience (24) • Constructive criticism or feedback (4, 7, 13, 34) • Be a feedback giver, idea bouncer (1-3, 10) • Give unbiased opinion (2, 39-40) • Have assessment ability and skills (19, 21, 28) • Act as evaluator (13) • Assess student's motivation (34) • Evaluate student's strengths and weaknesses (6, 34)

Appendix 3. Professional development dimension, subcategories, codes and references

Dimension 3: Professional development	
Professional socialization and competency building	**Role modelling**
• Introduce student to professional network (4, 8-9, 21, 28) • Involve student as a member in nursing team (8-9, 21, 25, 28-29) • Career counsellor (1-3, 10) • Talk about professional interests and issues (8-9) • Promote professional identity acquirement and professional socialization (7-9, 21, 28, 30) • Improve quality of care and clinical competence (24, 39-40); • Give student a degree of responsibility (8-9, 24-25, 36) • Involve in various nursing activities (34) • Be aware of legal concerns (36) • Teach patient care and decision making (34) • Teach prioritising working (34) • Open discussion with student (23) • Competency based education (23) • Let student attend unit meetings (34) • Attend continuing education programs (34) • Enhance student's competence (44) • Enhance student's responsibility (44) • Converse about patients and nursing (46)	• Demonstrate good professional skills, knowledge and competency (4, 14, 21-22, 26-28, 38) • Show good inter-personal communication skills with different people (5, 8-9, 13, 21-22, 28, 38) • Show positive attitude toward patients and nursing (34) • Provide clinical standard practice (1, 8, 10) • Evidence based practice, research application (2, 17, 29) • Professional integrity (4) • Role model (10-11, 13-14, 17, 20-21, 26, 28, 35, 43) • Be experienced (13) • Possess a positive image of nursing (5) • Being confident (5) • Show professional attitude (16) • Take responsibility of patients' safety (45)

Reference

1. Andrews and Chilton, 2000
2. Hall, 1997
3. Hall, 1998
4. Berk et al, 2005
5. Heffernan et al., 2009
6. Spouse, 2001
7.Cahill, 1996
8. Suen & Chow, 2001
9. Chow & Suen, 2001
10. Darling, 1984
11. Dibert and Goldenberg, 1995
12.Charleston and Happell, 2006
13.Elcigil and Sari, 2008
14. Gray and Smith, 2000
15. Smedley et al., 2010
16.Charleston&Happell,2005
17. Hallin and Danielson, 2009
18. Luhanga et al, 2010
19. Cele et al., 2002
20. Jakubik, 2012
21. Jokelainen et al, 2013,
22. Knox and Mogan，1985
23. Charleston and Goodwin, 2004
24. Löfmark et al, 2012
25. Lee et al, 2009
26. Andrews & Wallis 1999
27. Meno et al, 2003
28. Jokelainen et al, 2011
29. Myall et al, 2008,
30. Neary, 2000
31-32.Saarikoski et al, 2002; 2005
33.Bos et al., 2012
34.Kim, 2007
35.Warne et al., 2010
36. Weng et al., 2010
37. Lo,, 2002
38. Webb and Shakespeare, 2008
39-40. Winstanley and White, 2003; 2011
41.Severinsson and Sand, 2010
42. Price, 2006
43. Mårtensson et al,2012
44. Öhrling and Hallberg, 2000a
45.Öhrling and Hallberg, 2000b
46. Öhrling and Hallberg, 2000c

Appendix 4 Ethic approval from the University of Hull

UNIVERSITY OF Hull

Ms Yanhua Chen
Faculty of Health and Social Care
University of Hull
HU6 7RX

FACULTY OF HEALTH AND SOCIAL CARE
T: 01482 464680
E: j.dyson@hull.ac.uk

OUR REF: 137
29April 2014

Dear Yanhua

Re: Assessment of mentors' behaviour – a scale development and validation in nursing

Thank you for advising me about your proposed minor amendment to the above.

Given the information you have provided, I am able to give Chair's approval for your amendment. However, may I remind you that ethics approval is subject to the following condition as stated in my original letter;

- The Committee has sight of the survey before distribution

I look forward to receiving the survey prior to you commencing data collection.

Yours sincerely

Dr Judith Dyson
Chair, Research Ethics Committee

cc: file/supervisors

Appendix 5. Agreement of data collection from Luzhou Medical University

燕华：你好！
来信收悉。同意研究。
祝一切顺利！
何涛

在 2014-04-11 13:36:53，YAHUA <chen_yanhua25@163.com> 写道：
-显示引用文字-

Translation

Dear Yanhua

I have accepted your application and your research is approved.

Best wishes

He tao

Appendix 6 Agreement of data collection from West China Hospital

燕华

燕华好！你是华西的学子，我们更应该帮你。希望你学业有成。祝顺利开心！胡秀英

--

Xiu-Ying Hu RN, PhD, Professor
Dean of Nursing School / Director of Nursing Dept.
West China Nursing School / West China Hospital，
Sichuan University，China.
Tel & Fax: 86-28-85422042

Translation

Dear Yanhua

You are one of our graduates, so we should help you more. I wish you very successful in your study.

Xiu-Ying Hu RN, PhD, Professor
Dean of Nursing School / Director of Nursing Dept.
West China Nursing School / West China Hospital，
Sichuan University，China.
Tel & Fax: 86-28-85422042

Appendix 7. Agreement of data collection from one Guangzhou Medical university, nursing school

燕华

你好！

我非常愿意帮助你收集资料，我会把你的研究相关信息告诉实习同学的。

祝你学习顺利！

李桃

Translation

Dear Yanhua

I am very happy to help you to collect the data. I will dissimilate information about your research among nursing students learning on wards.

Best wishes

Li Tao

Appendix 8. Agreement of data collection from one of tertiary hospital in Sichuan province

燕华

你好！

我很高兴能帮助你。 我会把你得研究信息发到医院护理 QQ 群里面，让我们的护士了解并参与你得研究。

祝一切顺利

黎雪梅

Translation

Yanhua

I am pleased that I can help you. I will advertise your research information in the QQ group of our nursing staff and let our nurses (mentors) know and participate your study.

Wish you everything is smooth.

Li Xuemei

Appendix 9. Agreement of data collection from one of tertiary hospital in Shenzheng

燕华

你好！

希望能帮到你，你的研究对改变临床护理教学很有意义，我会在护理 QQ 群里面告诉大家你得研究相关信息，我想一定有护理导师积极参与的。

祝早日完成学业

阳世伟

Translation

Dear Yanhua

I hope I can help you. Your study is meaningful to change clinical nursing education. I will advertise your research material in our staff QQ group and I believe that mentors will participate your research actively.

Yang Shiwei

Appendix 10. Research participant information sheet (Chinese version)

UNIVERSITY OF Hull

研究项目信息单

研究题目：护理临床导师行为评价——量表（问卷）发展与验证

大家好！

我是陈燕华，现于英国赫尔大学攻读护理博士学位，我的导师是Roger Watson 教授和Andrea Hilton博士。我正在做关于临床导师行为评价量表的发展与验证研究，旨在通过该研究提高临床导师教学水平和护理学生临床学习质量和学习感受。希望您能参与研究，帮助我完成课题，在你作出是否参加该研究的决定前，请花几分钟阅读本研究的相关信息。

1. 该研究要做什么？

定义：临床导师 指 一对一临床带教老师，一名老师负责一名实习同学，与同学共同工作 50%以上班次。有的医院或科室没有进行这样的教学模式，就请老师们、同学们设想假如有这样的带教模式，你愿意怎样带同学或者你期望老师怎样教你。

本研究包括两个阶段：

阶段 1 量表（问卷）发展，主要通过文献研究和 6 个焦点小组访谈建立条目池

阶段 2 量表验证，在验证阶段需要一系列步骤评价新问卷的质量，包括问卷的稳定性、一致性和真实性。12 名专家组参与内容效度评定，两次问卷调查，每次问卷调查需要 500 名护理实习学生建立问卷结构效度。

这项研究将从 2014 年 4 月开始，9 月结束，共需要三批不同研究对象参与，包括 30 名护理实习同学和 20 名护理临床导师参与焦点小组访谈，英国和中国各 12 名临床护理教育专家评定问卷内容效度，1000 名护理实习同学参加问卷调查。

2. 如果参与该研究，我需要做什么？

小组访谈

小组访谈的主要目的是建立问卷需要的问题（条目）该研究需要 6 个实习同学小组，3 个临床导师小组，每个小组需要 4-5 人参与。护理本科、大专、中专同学均可以参加。凡是具备两年以上临床带教经验的护理老师均可以参加临床导师组讨论。

访谈的形式是通过建立专门 QQ 群，约定专门时间集体讨论关于临床实习经历，你所期望的临床导师行为特征等。临床导师小组访谈将讨论带教体会。 大家通过键盘输入你的想法，并与小组其他成员讨论、互动。每个小组访谈大概需要 1-2 个小时。

如果你决定参加，我绝不会泄漏任何参与者个人的信息，该访谈不会涉及到个人隐私信息，同时也请每一位参与讨论者保守小组其他成员的信息。如果你对任何问题有疑虑或顾忌，可以不回答，任何时候，你都可以退出讨论。小组讨论将在 2014 年 4 月到 5 月进行。

专家组判断量表质量（评价内容效度）

本研究将邀请英国和中国各12名临床导师制专家对新问卷中各个问题的相关性和重要性、清晰性进行评价。专家需要评价每一个问题（条目），每个条目有三个选项：不相关，一点相关，高度相关 。你只需要在相应的方格里打钩。您也可以根据您的知识和经验增加新的条目， 整个过程大概需要半个小时。这部分研究将在2014年6月进行。

横断面调查确定问卷结构

专家评定问卷质量后，将进行两次问卷调查以明确新量表的构成（分量表，维度），每次需要调查500名护理学生，包括护理本科、大专、高职、中专。该调查应用网络调查软件设计而成，点击链接就可以找到该问卷（该链接将在7月建立，并发送给愿意参与研究的同学）。该问卷包括三个部分：第一部分是关于你的一般情况如性别，实习时间，学历等。第二部分与第三部分问题一样，但选项不同，第二部分评价临床导师行为的重要性，第三部分评价对临床导师行为的满意度。如果你有过一对一的带教经历两部分都可以填， 没有一对一带教经历者请填写第二部分。该问卷需要5-10分钟完成。你只需要在符合你观点的方格内打钩。这部分研究将在2014年7-9月进行。

3. 我必须参加这个研究吗？

参与科研完全是自愿选择，如果你决定不参加研究，这不会对你带来认为负面影响。我真诚的希望你能参与研究，相信你的参与会对本研究作出有意义的贡献。

4. 参与本研究可能带来的益处有哪些？

虽然参与本研究可能没有直接的个人益处，但你提供的信息对提高护理导师的教学能力，提高未来护理学生的学习质量和改善学习经历具有重要作用。

5. 参与本研究可能带来的危险有哪些？

参与本研究没有任何危险，但可能会给你带来不便，比如小组访谈需要大家安排相同的时间。

6. 如何为研究对象保密？

整个研究将严格匿名和保密。任何个人信息都不会被泄露。在发表论文时可能会引用小组讨论和专家的话，但任何个人的名字或者其他能辨认的信息绝对不会出现在任何地方。调查研究收集到的数据在研究期间只有导师和研究者能接触到，将来可能会有其他研究人员用到，以利于该研究的继续、深入，实现数据分享，但该数据一定是匿名的。

7. 联系方式

如果你对研究的任何一部分感兴趣（每人只能参加一部分研究），或者有任何问题请联系：

研究者：陈燕华（QQ ID707655508, email: y.chen@2012.hull.ac.uk)

或者导师 Professor Roger Watson (email: r.watson@hull.ac.uk)

Dr. Andrea Hilton (email: A.Hilton@hull.ac.uk)

感谢您的阅读。对于您的参与和帮助，我将不甚感激。

Appendix 11. Research participant information sheet

Research Participants Information Sheet

Study title: Assessment of mentor's behaviour—A scale development and validation

My name is Chen Yanhua, a PhD student at The University of Hull under the supervision of Professor Roger Watson and Dr. Andrea Hilton. I am conducting a study involving the development and validation of a rating scale of mentors' behaviour, aiming to improve mentors' teaching quality and nursing students' clinical learning quality and experience in China. I would like you to take a few minutes to read this information sheet before making your decision about whether you would like to help me with my research.

How will the research be conducted?

This research contains two phases:

Phase 1 - Scale development

Phase 2 - Validation

In phase 1 an item pool will be generated through a literature review and focus group.

During phase 2, an expert panel will be used to establish content validity; two cross-sectional surveys of up to 1000 Chinese nursing students will be conducted to investigate further the validity of the scale.

This research needs four different groups of participants and will be conducted from April 2014 to January 2015.

What will I do if I take part in the research?

Focus group (students and mentors)

To generate new items or themes for the new scale, online focus groups will be conducted among Chinese nursing students and mentors; this can also give the new scale higher content validity and more contextual information. The group

size is 6-8 participants; two student groups and two mentor groups are required. Students in degree, associate degree, and diploma programs are eligible to participate in the students' focus group, while mentors in any department with over two years of mentoring experience are welcome to participate in the mentor focus group. Purpose sampling will be used to make the group composition diverse. If any student or mentor is invited as a legitimate participant and also interested in the study, the link teacher in the Luzhou Medical College will give you the researcher's contact detail as QQ ID or email, and after you contact with the researcher, the research information sheet, cover letter, informed consent form will be sent to you and you will be invited to print out the informed consent form, sign it and email back to the researcher if you agree to participate in this research.

The students will be invited to discuss via QQ chat room about things such as clinical learning experience with a mentor and expectations from mentors.

Mentors will be asked about their mentoring experience and what actions they think are important for a mentor

Both students and mentors focus groups are interactive and you will be expected to respond (via typing) to the discussion, this will last about 1 hour.

Once the focus groups have been completed, I will copy the discussion into a Word file (which will be saved securely) and the conversion will be deleted from QQ.

If you agree to participate you should not mention the names of any individuals or organizations. Please remember to maintain the confidentiality of each member of the group. You may use a pseudonym instead of your real name; you will not be asked about any personal information. If you are not comfortable with any question being asked at anytime by any participant you are not obliged to answer. You can withdraw at any time during the session. This session will be conducted from April to May 2014.

Expert panel for content validity

People who are highly experienced and specialized in nursing students' mentoring and mentor supervision, working around Humberside and Easter Yorkshire will be invited to participate in a panel of 12 experts to rate the items.

You will be contacted through email by one colleague in the University of Hull and the research information sheet, cover letter, informed consent statement will be sent to you as well. Your opinion and judgment is important to ensure the quality of the scale. You need to rate each item for relevance to the mentors' behavior using the labels of 'not relevant', little relevance, highly relevant'. You simply tick the boxes according to your perspective. Also you can add some new items according to your experience and knowledge. It will take you no longer than 30 minutes. This part will be conducted in June 2014.

Cross-sectional survey for Factorial validity

Following the focus groups and expert panel discussion, two surveys will be carried out to test the quality of the scale. For each survey, up to 500 nursing students in degree, associate degree, and diploma programs, in any hospital will be invited to complete a survey. This is an online survey and the web link, informed consent statement, research information sheet and cover letter will be sent to you via QQID by the researcher. The questionnaire includes three parts:

Part 1 is about your general information such as how long have you been in placement; your study program etc but no personal identifying information

Parts 2 and 3 are concerned with importance of the mentors' behavior, and the assessment of your satisfaction with the actual mentoring experience. Depending on your actual situation you can choose to complete Part 2 or Part 3 or both. You simply need to find the webpage and tick the boxes which apply to you. It should not take you more than 10 minutes to complete the questionnaire. This part will be done from July to September 2014.

Do I have to take part in this research?

Your participation is absolutely voluntary. We would like you to participate in this research because we believe that you can make a meaningful contribution to this study.

What are the possible benefits of taking part?

Although there may be no personal benefits to your participation in this study, the information you provide could improve mentors' performance and the future students' learning experience in clinical placement.

What are the possible disadvantages and risk of taking part?

Although you may be invited to talk about mentors behaviours or your expectations from mentors, there is no risk involved. Depending on which stage you are taking part, the maximum time involved is 60 minutes. .

How confidentiality will be maintained?

The anonymity and confidentiality of all information from you will be maintained at all times. The data from the focus group and expert panel will be used for publication and policy recommendation. Only my supervisors and I will have access to the raw data you provide to us. Your words may be quoted in the research paper, but no mention of your name or any other identifiable information

Thank you for your consideration. Your participation is greatly appreciated.

Chen yanhua 7 April 2014

Appendix 12. Cover letter to student focus group (Chinese version)

UNIVERSITY OF Hull

护理学生小组访谈说明信

研究题目：护理临床导师行为评价——量表（问卷）发展与验证

定义：

临床导师 指 一对一临床带教老师，一名老师负责一名实习同学，与同学共同工作 50%以上班次。就请同学们设想有这样的带教模式，你期望老师怎样教你。

亲爱的同学们：

大家好！

我是陈燕华，现于英国赫尔大学攻读护理博士学位，我的导师是 Roger Watson 教授和 Andrea Hilton 博士。我正在进行小组访谈研究以了解同学们对临床带教老师行为的期望，旨在通过该研究提高临床带教老师教学水平和护理学生临床学习质量和学习感受。这是课题（护理带教老师行为评价——量表（问卷）发展与验证）中非常关键的一部分，希望您能参与研究，帮助我完成课题。

小组访谈将在 2014 年 4 月到 5 月进行，其主要目的是建立问卷需要的问题（条目）该研究需要 3 组实习同学，每组需要 5 人参与。访谈的形式是通过建立专门 QQ 群，约定专门时间集体讨论关于临床实习经历，你所期望的带教老师是怎样的等。大家通过键盘输入你的想法，可以与小组其他的人的讨论、互动、提问、答复。每个小组访谈大概需要 1-2 个小时。

如果你决定参加，我绝不会泄漏任何参与者个人的信息，该访谈不会涉及到个人隐私信息，同时也请每一位参与讨论的同学保守小组其他成员的信息。如果你对任何问题有疑虑或顾忌，可以不回答，任何时候，你都可以退出讨论。整个研究将严格匿名和保密。任何个人信息都不会被泄露。在发表论文时可能会引用你的话，但任何个人的名字或者其他能辨认的信息绝对不会出现在任何地方。研究收集到的原始数据只有导师和研究者能接触到。

如果你对该研究有兴趣请联系研究者陈燕华其邮箱为：

y.chen@2012.hull.ac.uk QQ: 707655508

我真诚的希望你能参与研究，相信你的参与会对本研究作出有意义的贡献。对于你的参与和帮助，我将不甚感激。

祝学习愉快！

陈燕华

Appendix 13. Cover letter to student focus group

UNIVERSITY OF Hull

Cover letter to student focus group

Study title: Assessment of mentor's behavior ---- A scale development and validation

Dear students

My name is Chen Yanhua, a PhD student at The University of Hull under the supervision of Professor Roger Watson and Dr. Andrea Hilton. I am conducting a study to understand how mentors should behave to benefit nursing students' clinical learning and to understand what nursing students want from their mentors on wards. Your participation will provide invaluable data to generate items for a rating scale of mentors' behavior.

I am inviting you join in a group of 6-8 students to talk through QQ chat room about things such as: 'how do you feel about your clinical learning with your mentor?'; 'Who is your ideal mentor？'. The session will last approximately 1 hour and 12-16 students are in need. If you are interested in this study you can contact the researcher via QQ ID (707655508), and then an informed consent form will be sent to you. After the QQ chat room for the group is established, a suitable time for all participants will be arranged, and then you can type your perspectives simultaneously. During the talk you can interact with other participants or ask them questions, or respond to their questions.

If you agree to participate I should not mention the names of any individuals and maintain the confidentiality of each member of the group. You may use a pseudonym instead of your real name; you will not be asked about any personal information. If you are not comfortable with any question being asked at any time by any participants you are not obliged to answer, and you can withdraw at any time during the talk.

The data will be used solely for research purposes and policy recommendations. Confidentiality and anonymity will be maintained; the data will only be seen by me and supervisors. Your words may be quoted in the research paper, but no mention of your name or any other identifiable information.

Yours sincerely,

Chen Yanhua

Appendix 14. Informed consent form (Chinese version)

UNIVERSITY OF Hull

知情同意书

研究题目：护理临床导师行为评价——量表（问卷）发展与验证

大家好！

我真诚邀请您参加该研究，决定加入之前请您仔细阅读下面研究相关信息。如果您决定参加该研究请在下面的横线上签字。

研究目的: 发展并验证护理带教老师行为评价量表（问卷）

如果你参加，你会被邀请: 参加由 5 名护理学生组成的小组讨论。访谈的形式是通过建立专门 QQ 群（请不要用真名登陆），约定专门时间集体讨论关于临床实习经历，你所期望的带教老师是怎样的等。大家通过键盘输入你的想法，并与小组其他成员讨论、互动。

需要多长时间: 每个小组访谈大概需要 60 到 90 分钟。

潜在危险: 参与本研究没有任何危险，但可能会给你带来不便，小组访谈需要大家安排相同的时间。

益　　处：虽然参与本研究可能没有直接的个人益处，但你提供的信息对提高护理临床带教老师的教学能力，提高未来护理学生的学习质量和改善学习经历具有重要作用。

如何保密：整个研究将严格匿名和保密。任何个人信息都不会被泄露。在发表论文时可能会引用您的话，但任何个人的名字或者其他能辨认的信息绝对不会出现在任何地方。您提供的原始数据只有导师和研究者能接触到。

如果你对研究感兴趣，或者有任何问题请联系：

研究者：陈燕华（QQ ID707655508, email: y.chen@2012.hull.ac.uk）

自愿参与:

参与科研完全是自愿选择，如果你决定不参加研究，这不会对你带来认为负面影响。我真诚的希望你能参与研究，相信你的参与会对本研究作出有意义的贡献。该访谈不会涉及到个人隐私信息，如果你对任何问题有顾忌，可以不回答，任何时候，你都可以退出讨论。

我申明：我是在仔细阅读并完全理解本研究相关信息的基础上，同意加入该研究并签订知情同意书。

研究者姓名: 陈燕华　　签订日期: 2014-4-13

参加者姓名: _____　　签订日期:

签名方式：可以用电子签名，或者在 WORD，画图程序做一个别致的名字粘贴上来，也可以打印出来，手签，在照相传到 QQ 邮箱。

Appendix 15. Informed consent form

UNIVERSITY OF Hull

Informed consent form for students

Title of Project: Assessment of mentors behaviour----A scale development and validation in nursing

I am inviting you to participate in my research project. Please read the information bellow and the cover letter and information sheet about the project sent to you. If you would like to participate, please sign in the box below.

Purpose of the project: To develop and validate a scale to measure mentors' behaviour in nursing.

If you participate, you will be asked to: join in a group of 6-8 nursing students to talk about clinical learning experience with your mentors via QQ chat room with typing text. During the talk you can interact with other participants or ask them questions, or respond to their questions.

Time required for participation: The session will last approximately 1 hour.

Potential Risks of Study: No risk is seen at this moment

Benefits: Although there may be no personal benefits to your participation in this study, the information you provide can improve mentors' performance and the future students' learning experience in clinical placement.

How confidentiality will be maintained: You attend this focus group with your pseudonym instead of your real name. And your words may be quoted in the research paper, but no mention of your name or any other identifiable information. All the data will be protected with encryption. Only the researcher and the supervisors can access to the data.

If you have any questions about this study, feel free to contact:

Researcher: Chen Yanhua

Email: y.chen@2012.hull.ac.uk

QQ: 707655508

Voluntary Participation:

Participation in this study is entirely voluntary. If you decide not to take part in, it will not cause you any negative impact. Please be aware that if you determine to participate in the study when you are not comfortable with any question being asked at any time by any participants you are not obliged to answer, and you can withdraw at any time during the talk without any reason.

By signing this form I hereby confirm that I have read and understand all the information about the research and I freely give my consent to participate.

Name of Research: Chen Yanhua Date Signed:

Participant Signature: Date Signed:

Appendix 16. Informed consent statement for experts

UNIVERSITY OF Hull

Informed consent statement

Title of Project: Assessment of mentors' behaviour----A scale development and validation in nursing

I am inviting you to participate in my research project. Please read the information bellow and the cover letter and information sheet about the project sent to you.

Purpose of the project: To develop and validate a scale to measure mentors' behaviour in nursing.

If you participate, you will be asked to: rate each item for relevance to the mentors' behavior using the labels of 'not relevant', little relevance, highly relevant' three levels (1-3). You simply tick the boxes according to your perspective. Also you can add some new items according to your experience and knowledge.

Time required for participation: This may take you half hour to rate all the items in this new scale.

Potential Risks of Study: No risk is seen at this moment

Benefits: Although there may be no personal benefits to your participation in this study, the information you provide can improve mentors' performance and the future students' learning experience in clinical placement.

How confidentiality will be maintained: rating of the new scale will not gather any personal information and so it is anonymous. All the data from expert panel will be protected with encryption. The data will only be used for publication and policy recommendation and only my supervisors and I will have access to the information you provide to

us. Your words may be quoted in the research paper, but no mention of your name or any other identifiable information.

Voluntary Participation:

Participation in this study is entirely voluntary. If you decide not to take part in, it will not cause you any negative impact. Please be aware that if you rate the items in the new scale, the data cannot be withdrawn later but they are anonymous.

Participation will be considered as consent

If you rate the items in the new scale, it implies that you understand the information and consent to attend this research.

If you have any questions about this study, feel free to contact:

Researcher: Chen Yanhua

Email: y.chen@2012.hull.ac.uk

Appendix 17. Informed consent statement for student survey

UNIVERSITY OF Hull

Informed consent statement

Title of Project: Assessment of mentors' behaviour----A scale development and validation in nursing

I am inviting you to participate in my research project. Please read the information bellow and the cover letter and information sheet about the project sent to you.

Purpose of the project: To develop and validate a scale to measure mentors' behaviour in nursing.

If you participate, you will be asked to: fill in a questionnaire online through the web link (https://www.bristol.com/MySurvey). The questionnaire includes three parts: Part 1 is about students' general information such as how long have you been in placement; your study program etc. The items in Parts 2 and 3 are same, but the options are different; one assesses the importance of the mentors' behavior, and the other assesses your satisfaction with the actual mentoring experience. Depending on your actual situation you can choose Part 2 or Part 3 to complete or you can complete both parts. You simply need to tick the boxes which apply to you.

Time required for participation: It should not take you more than 10 minutes to complete the questionnaire.

Potential Risks of Study: No risk is seen at this moment

Benefits: Although there may be no personal benefits to your participation in this study, the information you provide can improve mentors' performance and the future students' learning experience in clinical placement.

How confidentiality will be maintained: the online survey will not gather any personal information and so it is anonymous. All the data from survey will be protected with encryption. Only the researcher and the supervisors can access to the data during this research and the data may be used by other researchers in the future, but it will be totally anonymity.

Voluntary Participation:

Participation in this study is entirely voluntary. If you decide not to take part in, it will not cause you any negative impact. Please be aware that if you fill in the questionnaire online, the data cannot be withdrawn later but they are anonymous.

Participation will be considered as consent

If you fill in the question, it implies that you understand the information and consent to attend this research.

If you have any questions about this study, feel free to contact:

Researcher: Chen Yanhua

Email: y.chen@2012.hull.ac.uk

QQ: 707655508

Appendix 18. Questions for focus group

Question route and time arrangement

For a beginner of focus group moderator, a question route is in need to guide the moderator and to make sure the right questions are asked in right order and probing correctly and control the time properly.

The question route includes opening question, introduction question, transition question, key question and ending question. And the wording of questions need reflect students or mentors language for easy understanding and provoking discussion. These questions were appraised and discussed within the research team.

Open questions

Open questions are about facts other than attitude, perspectives and it can be answered in less than 30seconds (Krueger & Casey, 2009 p.39). The aim of the open question is not to get information or invite discussion but get people to talk and make people feel comfortable. Regard to this research, the start question was: Would you please introduce yourself, such as where do you live and what is you hobby? This session will last about 5 munities.

Introduction questions

These questions will make the participants begin to think about the main topics of the focus group. Question as ‘tell us about your experience of clinical placement learning?’ was asked. In this part the moderator can get some themes or language the participants used. This session will last 10 munities.

Transition questions

Transition question serves as a logic link between the introduction and the key questions. For this research the transition question was ‘when you are in placement which mentor do you feel is the best one for you, and how does she mentor you? Or how do you expect she mentor you?’ which will last about 10 minutes.

Key questions

Three to six key question are needed, and they play a predominant role in solving the research questions, for this study aims to find out the mentors’ proper behaviour, based

on the purported three dimensional mentoring theoretical framework, the following 6 key questions were asked:

- How is your relationship with your mentors?
- How does your mentor support and encourage you or how do you expect your mentor support you and encourage you?
- How do your mentors facilitate your learning in placement?
- How do your mentors assess your study?
- How do your mentors promote your professional development?
- How do your mentors perform as a role model?

Each question needs about 15 minutes talking, so 90 minutes are required together.

Ending questions

In this session, participants will be asked to reflect about what they have commented and stated and what is the most important thing about this topic. The moderator need to summarise and criticize, to clarify and emphasis the main themes about the research which is also important for analysis.

In this research, students and mentors were asked to summarize their ideas about the most important mentors' behaviours, such as 'Among all of the behaviours we talked which are the top five behaviours you think the most important?' Then the moderator summarized the discussion using 2-3 minutes after which participants were asked about: 'is my summary complete or correct? Have I missed anything? Proximately 10 minutes are required for this session.

The questions were discussed by mentoring experts and research team; the following questions were addressed (Krueger & Casey, 2009 p.59):

- Are these the right questions?
- Are these the words that people use in the group?
- Do you understand the questions?
- Do any questions seem redundant?
- Do the questions seem to flow from one topic to another?
- What have we missed?

- What can we delete?

Prompts for each question were also prepared. Questions for mentors are identical to those for students which can allow comparison and contrast. Questions for asynchronous group were consistent with those of synchronous groups, but they just were put in the QQ blogs (each question formed a thread of discussion, and this was also a private environment only participants were allowed to access) and the questions were described in more detail as instant interaction was not available. And the questions were open to change, such as one of questions, 'how would you facilitate students to learn' was found not being addressed by anybody and participants reported that it was too broad, then it was divided into three more specific questions about facilitating learning (assess students, arrange learning activities and establish and maintain learning environment).

Appendix 19. Items from literature

No.	Item	Resource	Dimension
1	Helps students examine career interests	Chow & Suen, 2001	Professional development
2	Shares own professional experiences with students.	Chow & Suen, 2001, Winstanley & White, 2011	Professional development
3	sets a role model for students on good communication	2 items from Chow & Suen, 2001 , Altmann , 2006, Hallin & Danielson, 2010	Professional development
4	Happy to instruct and teach students	Chow & Suen, 2001 ; Saarikoski et al, 2008 ; Knox & Mogan, 1985 Altmann , 2006 ; Hou et al, 2010 ; Lee et al, 2009	Professional development
5	Commitment to the Preceptor role	Altmann , 2006	Professional development
6	Guides students to perform the future role of a registered nurse.	Chow & Suen, 2001, Adapt from Winstanley & White, 2011	Professional development
7	Promotes student independence	Knox & Mogan, 1985	Professional development
8	Show students clinical decision and judgment	Knox & Mogan, 1985 Löfmark et al, 2012	Professional development
9	demonstrated professional integrity ,	Berk et al , 2005	Professional development
10	Show Clinical competence	2 items from Altmann , 2006 ; Hou et al, 2011 ;	Professional development
11	Solving problem based on information and evidence	Hou et al,2011 ; Altmann 2006 From NMC guideline	Professional development
12	Improves the quality of care students give	Adapt from Winstanley & White, 2011	Professional development
13	Show a leadership on team work	NMC,	Professional development

14	Help students to become part of the team	Lee et al, 2009, Jakubic,2012	Professional development
15	Made students aware of the legal implications of treatment decisions	Löfmark et al, 2012	Professional development
16	I tried to adapt my mentor's behavior.	Weng et al, 2010	Professional development
17	widen student's clinical knowledge base	Adapt from Winstanley & White, 2011；Chow & Suen, 2001 , Lee et al, 2009；Knox & Mogan, 1985	Facilitating learning
18	helps students develop psychomotor skills	Adapt from Winstanley & White, 2011；Chow & Suen, 2001 , Lee et al, 2009；Knox & Mogan, 1985	acilitating learning
19	It is difficult for mentor to find the time to mentor students	2 items adapted from Winstanley & White, 2011; Knox & Mogan, 1985; Berk et al, 2005	Facilitating learning
20	Motivates students to reflect on clinical learning	2 items adapted from Winstanley & White, 2011, 2 Chow & Suen, 2001 , Löfmark et al, 2012; Hou et al, 2011, Hallin & Danielson, 2010	Facilitating learning
21	Helps students understand the rationales behind the way of practice	3 items Chow & Suen, 2001, Saarikoski et al, 2008 , Löfmark et al, 2012, Knox & Morgan, 1985, Hallin & Danielson, 2010	Facilitating learning
22	Asks students alternative ways of performing a task	Chow & Suen, 2001, Löfmark et al, 2012, Lee et al, 2009	Facilitating learning
23	Recommends sources of relevant references to students	Chow & Suen, 2001, Berk et al, 2005	Facilitating learning
24	Answered students questions satisfactorily (e.g.,timely response, clear, comprehensive)	Lee et al, 2009; Berk et al, 2005	Facilitating learning
25	Gives hints and guidance whenever necessary	Chow & Suen, 2001; Jakubic,2012; 2 items from Winstanley & White, 2011	Facilitating learning
26	Encouragement of questioning	Lee et al, 2009	Facilitating learning

27	Listens attentively	Knox & Morgan,1985; Adapt from Jakubic,2012	Facilitating learning
28	Orientates students to the clinical environment.	Chow & Suen, 2001	Facilitating learning
29	Gives students appropriate learning opportunities	2 items from Chow & Suen, 2001 , Saarikoski et al, 2008; Knox & Mogan, 1985	Facilitating learning
30	Gives useful feedback	Saarikoski et al, 2008 , Hou et al, 2010, 4 items from Löfmark et al, 2012, Lee et al, 2009, 2 items from Chow & Suen, 2001 , 2 items from Knox & Mogan, 1985, Berk et al, 2005	Facilitating learning
31	Aid with learning objectives and goals	Saarikoski et al, 2008 ; Löfmark et al, 2012, Lee et al, 2009	Facilitating learning
32	Organize clinical teaching activities from observing to independent practice	2 items from Löfmark et al, 2012, Hou et al, 2010, Mårtensson et al.2012	Facilitating learning
33	Ongoing assessment of achievements	Jokelainen et al, 2013; Altmann , 2006; Hou et al; 2010 Löfmark et al, 2012; Knox & Morgan,1985	Facilitating learning
34	Create a supportive practice environment	Hou et al, 2010; Adapt from Jakubic,2012	Facilitating learning
35	motivate students to extend nursing abilities	2 items from Berk et al, 2005	Facilitating learning
36	Co-working	Jokelainen et al,2013	psychosocial support
37	Give students support and encouragement	Winstanley & White, 2011; Berk et al, 2005; Löfmark et al, 2012; Knox & Morgan,1985, Chow & Suen, 200111,18, P, WW, S, 7,3 ,10, 1, 8, 24, 4, 29, 22, 6.20, SWM, 26 WA	psychosocial support
38	Discuss sensitive (personal) issues	Winstanley & White, 2011	psychosocial support
39	Reduce students stress	Winstanley & White, 2011; Chow & Suen, 2001	psychosocial support
40	Acts in a superior manner	Winstanley & White, 2011; Saarikoski et al, 2008; Adapt from	psychosocial

	towards students	Winstanley & White, 2011	support
41	Being approachable.	Berk et al , 2005	psychosocial support
42	Cares about students study	Löfmark et al, 2012; Knox& Morgan, 1985; Adapt from Jakubic,2012; Chow & Suen, 2001	psychosocial support
43	Respect students	Saarikoski et al, 2008, 2 items from Chow & Suen, 2001	psychosocial support
44	Trust students	Adapted from Saarikoski et al, 2008	psychosocial support
45	Has a warm and friendly attitude.	Chow & Suen, 2001	psychosocial support
46	Concerned about students occupational safety during placement	Chow & Suen, 2001	psychosocial support
47	Solve problem that students meet on wards	From research?	psychosocial support
48	Establish a good relationship with students	NMC	psychosocial support
49	Enhances students' confidence on practice	Chow & Suen, 2001	psychosocial support

Appendix 20. Item pool construction after focus group

	Label	Source	Dimension
1	Helps students examine career interests	Lit	Professional development
2	Shares own professional experiences with students.	Lit	Professional development
3	sets a role model for students on good communication	Lit	Professional development
4	Happy to instruct and teach students	Lit	Professional development
5	Commitment to the Preceptor role	Lit	Professional development
6	Guides students to perform the future role of a registered nurse.	Lit	Professional development
7	Promotes student independence	Lit	Professional development
8	Show students clinical decision and judgment	Lit	Professional development
9	demonstrated professional integrity ,	Lit	Professional development
10	Show clinical competence (deal with emergency)	Lit and focus	Professional development
11	Solving problem based on information and evidence	Lit	Professional development
12	Improves the quality of care students give	Lit	Professional development
13	Show a leadership on team work	Lit	Professional development
14	Help students to become part of the team	Lit	Professional development
15	Made students aware of the legal implications of treatment decisions	Lit	Professional development
16	I tried to adapt my mentor’s behavior.	Lit	Professional development

17	Acts as a good role model for nurse	Focus group	Professional development
18	Discusses nursing professional prospects with students	Focus group	Professional development
19	Give students career guidance	Focus group	Professional development
20	Fosters professional growth of students	Focus group	Professional development
21	Show prioritising task and working methods	Focus group	Professional development
22	Show clinical competence (deal with emergency)	Focus group	Professional development
23	Maintains standards and principles of nursing procedure	Focus group	Professional development
24	Transmits negative information of nursing profession	Focus group	Professional development
25	Instils positive professional attitude	Focus group	Professional development
26	Promote students to acquire nursing identity	Focus group	Professional development
27	Foster clinical and critical thinking	Focus group	Professional development
28	Shows team work cooperation	Focus group	Professional development
29	widen student's clinical knowledge base	Lit	Facilitating learning
30	It is difficult for mentor to find the time to mentor students	Lit	Facilitating learning
31	Motivates students to reflect on clinical learning	Lit	Facilitating learning
32	Helps students understand the rationales behind the way of practice	Lit	Facilitating learning
33	Asks students alternative ways of performing a task	Lit	Facilitating learning
34	Recommends sources of relevant	Lit	Facilitating learning

	references to students		
35	Answered students questions satisfactorily (e.g.,timely response, clear, comprehensive)	Lit	Facilitating learning
36	Gives hints and guidance whenever necessary	Lit	Facilitating learning
37	Encouragement of questioning	Lit	Facilitating learning
38	Listens attentively	Lit	Facilitating learning
39	Orientates students to the clinical environment.	Lit	Facilitating learning
40	Gives students appropriate learning opportunities	Lit	Facilitating learning
41	Gives useful feedback	Lit	Facilitating learning
42	Aid with learning objectives and goals	Lit	Facilitating learning
43	Organize clinical teaching activities from observing to independent practice	Lit	Facilitating learning
44	Ongoing assessment of achievements	Lit	Facilitating learning
45	Create a supportive practice environment	Lit	Facilitating learning
46	motivate students to extend nursing abilities	Lit	Facilitating learning
47	Gives students objective and comprehensive assessment	Focus group	Facilitating learning
48	Never gives any negative feedback in front of others,	Focus group	Facilitating learning
49	Corrects mistakes without embarrassing students	Focus group	Facilitating learning
50	Supervises students	Focus group	Facilitating learning
51	Asks students questions to facilitate and assess learning	Focus group	Facilitating learning
52	Arranges interdisciplinary learning activities	Focus group	Facilitating learning
53	Provide peer students sharing opportunity on wards	Focus group	Facilitating learning

54	Has a clear plan for students' learning	Focus group	Facilitating learning
55	Follows the clinic learning plan strictly	Focus group	Facilitating learning
56	Discuss learning objectives	Focus group	Facilitating learning
57	Accept student's individual difference	Focus group	Facilitating learning
58	Be accountable and responsible for students' learning	Focus group	Facilitating learning
59	Be active in teaching and instruction	Focus group	Facilitating learning
60	Help students to linking theory to practice	Focus group	Facilitating learning
61	Widens and deepens students' clinical knowledge base	Focus group	Facilitating learning
62	Widens and deepens students' clinical skills	Lit and focus group	Facilitating learning
63	Does work together with students	Lit and focus group	psychosocial support
64	Give students support and encouragement	Lit	psychosocial support
65	Discuss sensitive (personal) issues	Lit	psychosocial support
66	Reduce students stress	Lit	psychosocial support
67	Acts in a superior manner towards students	Lit	psychosocial support
68	Being approachable.	Lit	psychosocial support
69	Cares about students study	Lit	psychosocial support
70	Respect students	Lit	psychosocial support
71	Trust students	Lit	psychosocial support
72	Has a warm and friendly attitude.	Lit	psychosocial support
73	Concerned about students occupational safety during placement	Lit	psychosocial support
74	Solve problem that students meet on wards	Lit	psychosocial support

75	Establish a good relationship with students	Lit	psychosocial support
76	Enhances students' confidence on practice	Lit	psychosocial support
77	Listens to students' ideas and suggestions.	Focus group	psychosocial support
78	Confirms and affirms students	Focus group	psychosocial support
79	Establishes a teacher-colleague-friend relationship with students	Focus group	psychosocial support
80	Takes students as learners, not a pair of hands	Focus group	psychosocial support
81	Be caring	Focus group	psychosocial support
82	Demonstrates empathy	Focus group	psychosocial support
83	Call students their preferred title	Focus group	psychosocial support
84	Get to know and understand students	Focus group	psychosocial support

Appendix 21. Scale for mentor experts review and content validity

No.	Item	No relevant	Somewhat relevant	Quite relevant	Highly relevant	I-CVI
1	Enjoys teaching			2	7	100
2	Takes responsibility for my learning		1	5	3	88.9
3	Orientates me to the clinical environment.			1	8	100
4	Provides a supportive practice environment		1	3	5	88.9
5	Has a clear plan for my learning			2	6	100
6	Discusses learning objectives with me				9	100
7	Helps me achieve learning objectives and goals			1	8	100
8	Asks me questions to facilitate and assess learning				8	100
9	Actively instructs me		1	4	4	88.9
10	Encourages me to reflect on my learning			1	8	100
11	Helps me to link theory to practice				9	100
12	Answers my questions satisfactorily			2	7	100
13	Treats me as individual				9	100
14	Arranges interdisciplinary learning activities			7	2	100
15	Organizes learning activities from observing to independent practice			2	7	100
16	Provide peer students sharing opportunity on wards		4	2	3	55.6
17	Recommends sources of relevant references to me		2	3	4	77.8
18	Assesses my achievements continuously			1	8	100
19	Gives me objective and comprehensive assessment			2	7	100

20	Never gives any negative feedback in front of others			2	6	100
21	Corrects mistakes without embarrassing me			4	5	100
22	Gives me constructive feedback				9	100
23	Helps me examine my career prospects	1	5	1	2	33.3
24	Shares professional experiences with me		2	1	6	77.8
25	Facilitates good communication skills with staff and patients			3	5	100
26	Guides me to become a registered nurse.		2		7	77.8
27	Shows me how to make decisions about patient care			2	7	100
28	Shows me how to prioritise tasks			2	7	100
29	Displays clinical competence				9	100
30	Demonstrates professional integrity				9	100
31	Transmits a positive image of the nursing profession			2	7	100
32	Fosters critical thinking in me			4	5	100
33	Makes me feel part of the team			2	7	100
34	Makes me aware of the legal implications of nursing care			1	8	100
35	Makes me think ‘I aim to be like my mentor’		1	4	4	88.9
36	Encourages the use of evidence-based practice				9	100
37	Motivates me to give the best possible care				9	100
38	Encourages in-depth learning about clinical practice			1	8	100
39	Includes me in all clinical activities			4	5	100
40	Adheres to standards of practice				9	100

41	Always makes time to teach me			6	3	100
42	Works the same shifts as me		1	5	3	88.9
43	Works with me while on the same shift			5	4	100
44	Supports and encourages me			2	7	100
45	Instils confidence in me			3	6	100
46	Shows respect to me				9	100
47	Call me by my preferred title or name			2	7	100
48	Has a warm and friendly attitude			3	6	100
49	Listens to my ideas and suggestions			2	7	100
50	Treats me as a learner, not a pair of hands			1	8	100
51	Guides my personal development			4	5	100
52	Protects me from stress		4	5		55.6
S-CVI/UA		0.81				
S-CVI/Ave		0.95				

Appendix 22. Translation and back translation

No.	Item
1	看起来喜欢教学　修改为（喜欢教学） My mentor enjoys teaching.
2	对我的学习负责 My mentor is responsible for my learning.
3	**引**导我适应临床环境 My mentor leads me to adapt to the clinical environment.
4	**提供支持性的**实习环境 My mentor provides a supportive practice environment.
5	对我的学习有清晰的计划 My mentor holds a clear plan for my learning.
6	**与我一起**讨论学习目标 My mentor discusses learning goals with me.
7	**帮助我**实现学习目标 My mentor helps me achieve my learning goals.
8	**通**过提问促进和评估我的学习 My mentor promotes and assesses my learning by asking questions.
9	积极指导我学习 My mentor actively guides me to learn.
10	**鼓励我反思自己的学**习 My mentor encourages me to reflect on my own learning.
11	**帮助我理**论联系实践 My mentor helps me integrate theory with practice.
12	**回答**问题让我满意 My mentor answers questions to my satisfaction.
13	**把我看作独立个体** My mentor treats me as a separate entity.
14	**安排跨学科的学**习活动 My mentor makes arrangements for interdisciplinary learning activities.

15	为**我安排从**观察到独**立**实践的教学活动 My mentor arranges for me the teaching practice, from mere observation to independent practice.
16	**持**续评估我的学习成效 My mentor conducts an ongoing assessment for my learning outcome .
17	给我客观、综合的评价 My mentor gives me an objective and comprehensive evaluation.
18	**从不在其他人面前**给予我负面反馈 My mentor never gives negative feedback in front of others.
19	纠正错误时**不**让我尴尬 My mentor never puts me in embarrassment when correcting my errors.
20	给予我建设性反馈 My mentor gives me constructive feedback.
21	**促**进我与**工作人**员**及病人**间的良好交流 My mentor promotes my effective communication with the staff and patients as well.
22	**向我展示如何**对患者的护理做出决策 My mentor shows me how to make decisions on patient care.
23	**向我展示如何**优先处理工作 My mentor shows me how to prioritize work.
24	引导我成为一名注册护士 Guides me to become a registered nurse.
25	**展示**临床能力 My mentor demonstrates the clinical ability
26	**展**现职业操守 My mentor demonstrates professional conduct.
27	传递积极的护理专业形象 My mentor passes a positive image of the nursing profession.
28	**培养我的批判性思**维 My mentor develops my critical thinking ability.

29	让我感觉自己是团队的一份子 My mentor makes me feel that I am part of the team.
30	让我意识到护理所涉及**的法律**问题 My mentor makes me aware of the legal issues invovled in nursing.
31	**我想成**为导师那样的护士 I want to be a nurse as my mentor.
32	**鼓励使用循**证实践 My mentor encourages the use of evidence-based practice.
33	**激**发我提供尽可能最优的护理 My mentor stimulates me to provide the best possible care.
34	**鼓励我**进行深入临床实践学习 My mentor encourages me to carry out in-depth study of clinical practice.
35	让我参与所有相关的临床实践活动 My mentor lets me participate in all activities related to clinical practice.
36	坚持公认的实践标准 My mentor adheres to generally accepted standards of practice.
37	总是抽出时间指导我 My mentor always takes time to instruct me.
38	**和我上一**样的班次 My mentor is on the same shift with me.
39	**在同一班次**时和我一起工作 My mentor works with me when we are on the same shift.
40	**支持鼓励我** My mentor supports and encourages me.
41	**逐**渐树立我的信心 My mentor gradually builds my confidence.
42	**尊重我** My mentor respects me.
43	**以我喜**欢的称谓或名字称呼我 My mentor addresses me with my favorite title or name.

44	对我热情、友好 My mentor is warm and friendly to me.
45	倾听我的观点和建议 My mentor listens to my ideas and suggestions.
46	把我当做学生而不是劳动力 My mentor treats me as a student rather than a labour.
47	引导我的个人发展 My mentor guides my personal development.

Appendix 23. Test-retest reliability and stability

item	Mean		T	p	ICC
	Test1	Test2			
v2	4.53±0.62	4.47±0.87	0.81	0.43	0.84
v3	4.41±0.62	4.41±0.94	0.29	0.77	0.69
v4	4.12±0.86	4.18±0.95	0.00	1.00	0.93
v5	4.24±0.97	4.29±0.99	-0.29	0.77	0.83
v6	4.29±0.99	4.24±0.97	0.90	0.38	0.83
v7	4.18±0.88	3.76±1.03	2.28	0.04	0.73
v8	4.41±0.71	4.18±0.95	1.07	0.30	0.64
v9	4.53±0.72	4.06±1. 03	2.73	0.01	0.69
v10	4.24±0.83	4.06±1.09	1.68	0.11	0.77
v11	4.35±0.93	4.18±0.95	1.68	0.11	0.76
v14	4.12±0.86	3.71±0.92	2.96	0.01	0.85
v16	4.47±0.80	4.00±1.00	3.64	0.00	0.84
v17	4.59±0.62	4.00±0.87	4.03	0.00	0.8
v20	4.53±0.62	4.29±0.85	1.79	0.09	0.49
v21	4.35±0.86	4.35±0.93	0.44	0.67	0.9
v22	4.47±0.94	4.47±0.80	0.77	0.45	0.68
v23	4.59±0.87	4.29±0.85	2.28	0.04	0.68
v24	3.88±1.05	3.71±1.26	1.64	0.12	0.7
v25	4.29±0.77	4.12±1.05	1.42	0.17	0.75
v26	4.41±0.80	4.59±0.80	0.00	1.00	0.66
v27	4.59±0.62	4.41±0.94	1.56	0.14	0.73

v28	4.24 ±0.83	4.24 ±0.97	0.89	0.39	0.58
v29	4.41 ±0.80	4.47 ±0.87	0.52	0.61	0.7
v30	4.29 ±0.85	4.24 ±0.90	1.46	0.16	0.66
v32	4.29 ±0.85	4.29 ±0.92	0.70	0.49	0.83
v33	4.47 ±0.80	4.18 ±0.95	2.39	0.03	0.76
v34	4.47 ±0.80	4.12 ±0.99	2.19	0.04	0.74
v37	4.18 ±0.88	4.06 ±0.83	1.42	0.17	0.81
v38	3.82 ±0.81	3.59 ±1.18	1.42	0.17	0.8
v39	4.00 ±0.71	3.71 ±1.05	2.11	0.05	0.8
v40	4.41 ±0.87	4.12 ±0.99	1.84	0.08	0.81
v41	4.35 ±0.86	4.29 ±0.99	0.62	0.54	0.68
v42	4.59 ±0.62	4.47 ±0.80	1.29	0.22	0.68
v44	4.53 ±0.80	4.12 ±0.86	2.73	0.01	0.65
v45	4.47 ±0.62	4.12 ±0.99	2.96	0.01	0.81
v46	4.71 ±0.77	4.71 ±0.77	0.57	0.58	0.93
v47	4.53 ±0.80	4.12 ±1.17	2.38	0.03	0.67
PS	43.59 ± 6.53	41.29 ±7.24	2.216	.042	0.85
PD	70.35 ±11.04	67.76 ±12.09	1.639	.121	0.91
FL	47.41 ±7.75	45.53 ±8.68	1.912	.074	0.93

Appendix 24. Online survey format

临床护理导师行为评价z

1 请评价以下每项导师行为对提高护理学生实习质量和实习感受的重
实际行为表现。（评价包括两部分：左边是重要性评价，右边是导

Having trouble with the format of this question? View in tableless mode

	重要性评价1一点也不重要；2.不重要；3.不确定；4.重要；5.非常重要					请回忆你最近的为）她具备以下同意；2.不同意	
	1	2	3	4	5	1	2
我的导师喜欢教学							
我的导师对我的学习负责							
我的导师引导我适应临床环境							
我的导师提供支持性的实习环境							
我的导师对我的学习有清晰的计划							

Appendix 25. Syntax of statistical strategies used

Syntax of exploratory factor analysis

DATASET ACTIVATE DataSet1.

FACTOR

/VARIABLES v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v14 v16 v17 v20 v21 v22 v23 v24 v25 v26 v27 v28 v29

v30 v32 v33 v34 v37 v38 v39 v40 v41 v42 v44 v45 v46 v47

/MISSING LISTWISE

/ANALYSIS v2 v3 v4 v5 v6 v7 v8 v9 v10 v11 v14 v16 v17 v20 v21 v22 v23 v24 v25 v26 v27 v28 v29 v30

v32 v33 v34 v37 v38 v39 v40 v41 v42 v44 v45 v46 v47

/PRINT INITIAL KMO EXTRACTION ROTATION

/FORMAT SORT

/PLOT EIGEN

/CRITERIA FACTORS(3) ITERATE(25)

/EXTRACTION PAF

/CRITERIA ITERATE(25) DELTA(0)

/ROTATION OBLIMIN

/METHOD=CORRELATION.

Syntax of MSA

Using MSA

Open R and type:

> library(mokken)

Converting an SPSS file for use in R

> library(foreign)

> FileR <- data.frame(read.spss("C:/FileSPSS.sav"))

You may get some errors or warnings at this stage which may have to be fixed before proceeding, then:

> fix(FileR)

This will show you the data as they appear in R, then:

> save(FileR, file = "C:/FileR.Rdata")

Once you have create an R file is can be uploaded again by:

> load("C:/FileR.Rdata")

Generating scales
To partition items in the FileR database into Mokken scales type:

> aisp(FileR)

Scalability coefficients
To produce scalability coefficients for items and the overall scale(s) type:
> coefH (FileR)

Mean item scores
To produce the mean values for all of the items in the scale type:
> apply(FileR,2,mean)

Monotonicity
To check monotonicity type:
> summary(check.monotonicity(FileR))

Plotting item step response functions
To plot item step response functions type:
> plot(check.monotonicity(FileR))
NB: if this does not work and you get:
*****Error in est - qnorm(1 - alpha.ci/2) * se : non-conformable arrays
In addition: Warning messages:
1: In (x - x^2)/n :
longer object length is not a multiple of shorter object length****
this is a problem in R and you should use:
> plot(check.monotonicity(FileR), plot.ci = FALSE

Invariant item ordering
To check invariant item ordering type:
> summary(check.iio(FileR))

OR

> iio.results <- check.iio(WEMWBSM14)

> summary(check.iio(WEMWBSM14, item.selection = FALSE))

Generating pair plots
To generate pair plots:
> plot(check.iio(FileR))
The confidence intervals can be omitted by:
> plot(check.iio(FileR), plot.ci = FALSE)
To select item pairs, eg 1st, 3rd & 7th:
> plot(check.iio(FileR), item.pair = c(1, 3, 7))

Saving plots
To save plots in a file (eg as pdf) in eg drive C:\
> NameOfFigure = "FileR.pdf"
> setwd("C:")
> pdf(NameOfFigure)
> plot(check.iio(FileR), ask = FALSE)
> dev.off()

Without confidence intervals
> NameOfFigure = "FileR.pdf"
> setwd("C:")
> pdf(NameOfFigure)
> plot(check.iio(FileR), plot.ci = FALSE, ask = FALSE)
> dev.off()

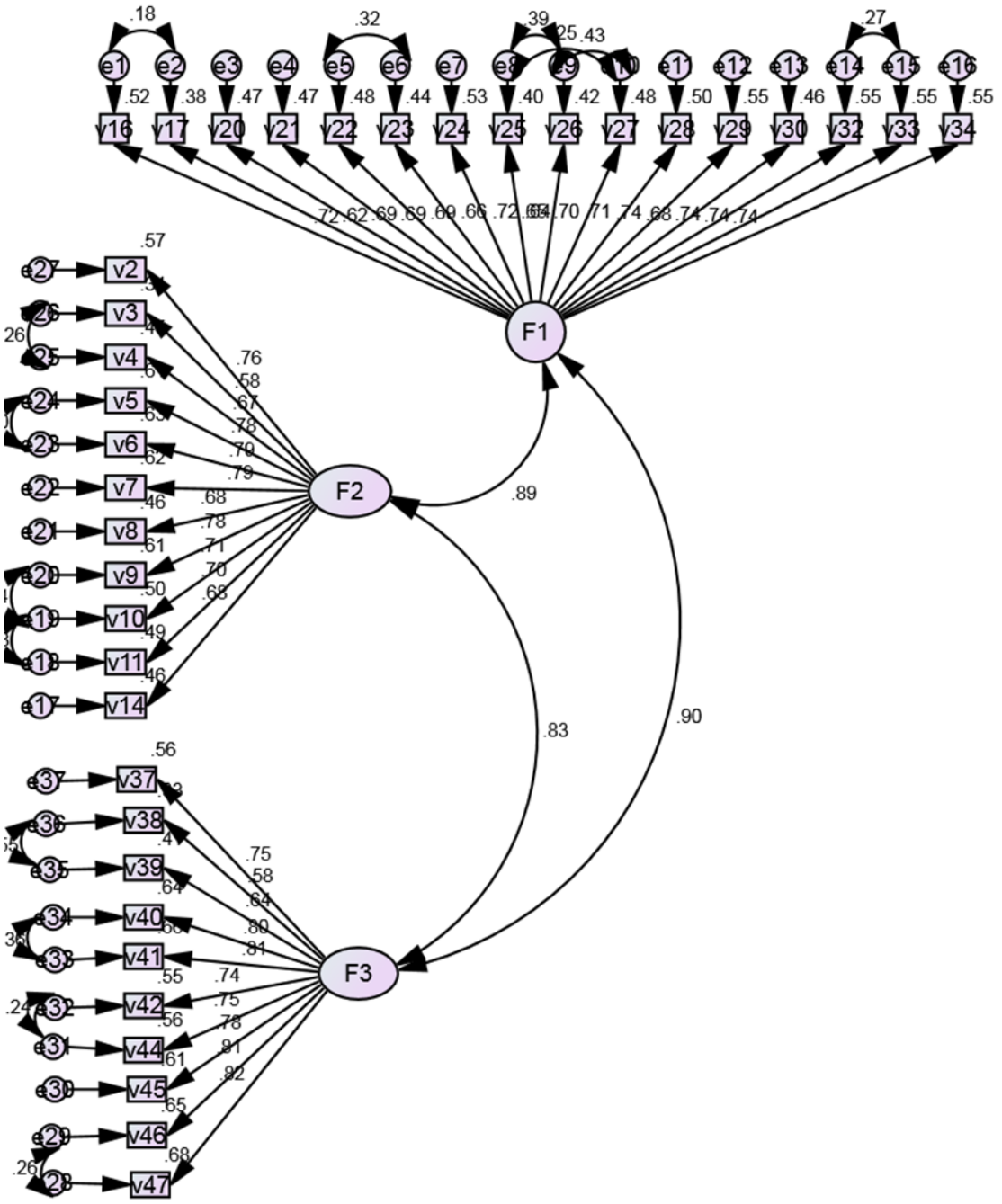
Reliability
To check reliability type:
> check.reliability(FileR)

Selecting items to analyse
To select specific items you need to create a new file as follows, type:
> FileRy <- FileR[,c(1,2,3,4)] - this will select items 1, 2, 3 & 4
> FileRy <- FileR[,c(1,2,3:10)] - this will select items 1, 2, 3, 4, 5, 6, 7, 8, 9 & 10
In both cases you analyse FileRy

Selecting individual for analysis
To select specific individuals you need to create a new file as follows, type:
> FileRx <- FileR[c(1,2:5)] - this will select individuals 1, 2, 3, 4 & 5

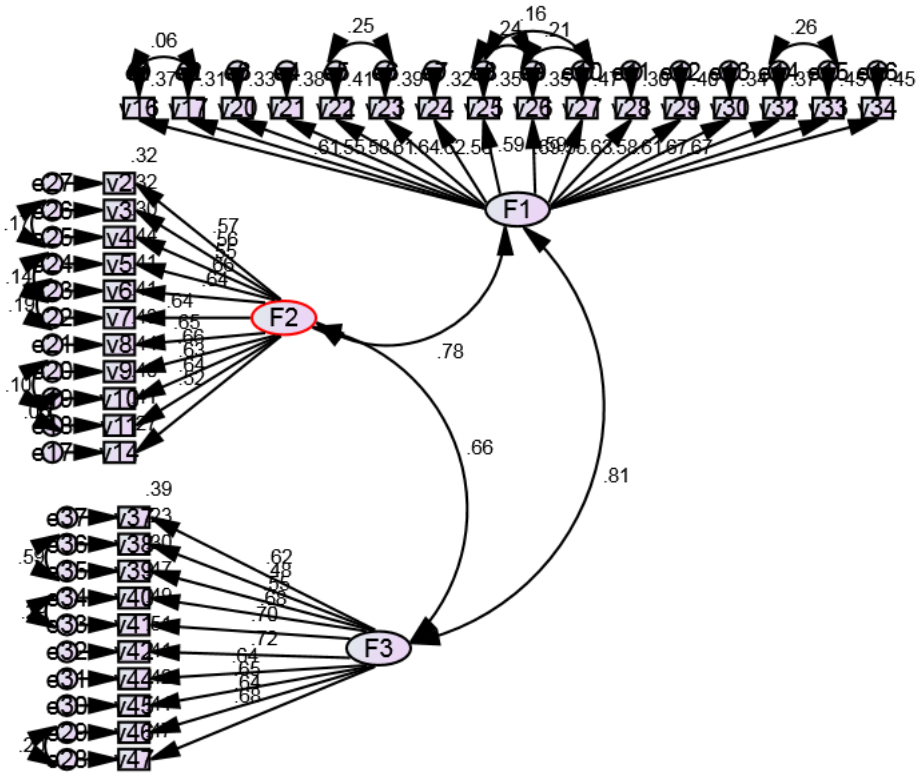
Roger Watson 10 March 2014

Appendix 26. Equivalent model in the assessment data set



*F1 represents professional development; F2 reprsents facilitating learning; F3 represents psychosocial support.

Appendix 27 Modelling in the importance data set



*F1 represents professional development; F2 reprsents facilitating learning; F3 represents psychosocial support.