Nurse Education Today 79 (2019) 198–203

Contents lists available at ScienceDirect

Nurse Education Today

journal homepage: www.elsevier.com/locate/nedt

Development and psychometric testing of the Research Competency Scale for Nursing Students: An instrument design study

Chujin Qiu[a,f], Xianqiong Feng[b,*], Jan D. Reinhardt[c,d,e], Jialing Li[a]

[a] West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China
[b] West China School of Nursing, West China Hospital, Sichuan University, Chengdu, China
[c] Institute for Disaster Management and Reconstruction, Sichuan University and Hong Kong Polytechnic University, Chengdu, China
[d] Swiss Paraplegic Research, Nottwil, Switzerland
[e] Department of Health Sciences, University of Lucerne, Switzerland
[f] First Affiliated Hospital of Shantou University Medical College, Shantou, China

ARTICLE INFO

Keywords:
Nursing student
Research competency
Scale development
Psychometrics

ABSTRACT

Background: Undergraduate students make up the largest pool for the recruitment of future nurses. The fundament for developing research competency has to be laid at this level of education.
Objectives: To evaluate the construct validity, reliability, sensitivity to change and convergent validity of the Research Competency Scale for Nursing Students (RCS-N).
Design: A psychometric study with repeated measurements.
Setting: Nursing school from Sichuan, China.
Participants: 146 undergraduate nursing students.
Methods: Construct validity was evaluated with exploratory factor analysis, confirmatory factor analysis (CFA) and item response theory (IRT) analysis. Internal consistency reliability was determined with Cronbach's alpha, and IRT based reliability estimates were also provided. Sensitivity to change was evaluated with repeated measures ANOVA. Convergent validity was analyzed based on the correlation of the RCS-N scores with nursing research class examination scores.
Results: Exploratory factor analysis yielded one dominant factor explaining 72.59% of the items' variance. CFA confirmed unidimensionality of the scale and produced appropriate goodness of fit indices after accounting for local dependency by collapsing several items to testlets. IRT analysis demonstrated a good capability of the scale to differentiate between persons with different abilities. Internal consistency was excellent ($\alpha = 0.98$) and IRT based reliability was good to excellent. The participants' RCS-N scores increased significantly from 29.3 before participation in the research methodology class to 62.6 after the class. There was no correlation ($r = 0.16$, $p = 0.051$) between the RCS-N result and class test scores.
Conclusions: RCS-N is a promising, valid and reliable tool for evaluating the research competency of undergraduate nursing students. The scale can however not replace actual examination of students' knowledge acquired in research methodology courses.

1. Introduction

Nursing research is the systematic inquiry designed to develop knowledge about issues related to the profession of nursing, including nursing practice, education, management, and informatics (Polit and Beck, 2004; Parahoo, 2014). Strengthening research competency has been considered an important approach to promote the development of the nursing profession and to improve the quality of nursing practice. Nursing research competency, in turn, refers to the ability 1) to systematically summarize clinical experiences as well as literature in order to find researchable problems and formulate innovative research questions of relevance to nursing, 2) to collect, analyze and explain data related to those questions, and 3) to apply knowledge accumulated in this way to solve problems innovatively (Gerrish and Lacey, 2010; Pan et al., 2011).

Since undergraduate students make up the largest pool for the recruitment of future nurses, the fundament for the development of nursing research competency consequently has to be laid at this level of

* Corresponding author at: No. 37 Guo Xue Xiang, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China.
E-mail address: fengxianqiong66@126.com (X. Feng).

https://doi.org/10.1016/j.nedt.2019.05.039
Received 11 August 2018; Received in revised form 6 April 2019; Accepted 8 May 2019

education. Several studies, however, revealed that Chinese nursing undergraduates have relatively low research competency (Zhao et al., 2011; Zhou et al., 2001). These results suggest that it is necessary to enhance nursing research education within undergraduate programs.

A key to continuously improve the quality of undergraduate curricula and teaching practice aiming at generating students' research skills is the systematic, regular, and standardized assessment of these skills. Most instruments aiming at the assessment of nursing research competency which are currently used in China (Pan and Cheng, 2011) and elsewhere were developed with the goal to evaluate clinical nursing staffs' research competency (Gething et al., 2001; Flinkman et al., 2017; Zhang et al., 2015). While their suitability for undergraduate nursing students, who have not yet come into contact with clinical work, is limited. Moreover, modern test theory has rarely been applied to evaluate construct validity and reliability of those scales.

The objectives of this study were to develop a scale to measure undergraduate nursing students' research competency, and to evaluate its psychometric properties including reliability and validity, so as to offer an effective tool to continuously monitor and improve the quality of research education in Chinese undergraduate nursing programs. The specific aims were to evaluate 1) the content validity, 2) the construct validity, 3) reliability, 4) sensitivity to change, and 5) convergent validity of the Research Competency Scale for Nursing (RCS-N), employing both classical as well as modern test theory.

2. Methods

2.1. Design

2.1.1. Psychometric study

This study was approved by the Medical Ethics Committee of Sichuan University (No. K2018058).

2.2. Instrument

A first draft of the RCS-N was initially developed following a critical review of the literature. Four types of literature were consulted: 1) literature related to research procedures including: research questions/hypotheses, literature review, research design, pilot-study, data collection, data analysis and research manuscript writing (Gething et al., 2001; Parahoo, 2014), 2) literature related to research skills, mainly including: creative thinking ability, problem identification and problem solving ability, research designing ability, research implementation ability, data collection, analysis and result expression ability (Polit and Beck, 2004), 3) literature related to nursing students' research competency and factors affecting the research competency, 4) literature associated with existing nursing research competency scales.

Based on the above literature review and the researchers' experiences, a scale containing 20 items was constructed. Responses to items were graded on a five-point Likert scale ranging from 1 (don't know at all) to 5 (very familiar). Afterward, the researchers supplemented and revised the contents of the scale by discussing with the experts involved in the teaching of nursing research courses: three items were added (i.e. “ethical requirements for research”, “concept of evidence-based nursing”, and “steps of evidence-based nursing”), and the item “types and components of research design” was divided into two items (i.e. “types of research design” and “components of research design”). The Delphi method was then used to further evaluate and revise the scale content. According to the experts' opinions, several modifications were made: 1) item “ways of writing research paper” was changed into “writing requirements for research papers”. 2) item “methods of statistical analysis” was changed into “analytical method for research data”. 3) the rating option “don't know at all” was changed into “very unfamiliar”.

2.3. Assessment of content validity

To ensure the content validity of the instrument, five experienced teaching and clinical nursing staffs (four associate/full professors and one clinical nurse specialist) were invited to rate each item of the scale based on the relevance of the item content to nursing research competency. The scoring method is as follows: 1 = not relevant, 2 = somewhat relevant, 3 = relevant, and 4 = very relevant. The item-level content validity index (I-CVI) was calculated by dividing the number of experts with scores > 3 by the total number of experts. The scale-level content validity index (S-CVI) is the average of the I-CVI for all entries.

2.4. Psychometric analysis

2.4.1. Participants

The RCS-N was administered to a convenience sample of 153 undergraduate students from a top five nursing school in China (www.chinadegrees.cn, 2018). Students were eligible if they were enrolled in the nursing research course. All participants were full-time students. Participants were informed about the purpose of the study and signed the informed consent form. Participation was voluntary, and they could withdraw from the study at any point in time.

2.4.2. Data collection

Data were collected between 2016 and 2017. In addition to RCS-N, information on the participants' age, gender, and grade were also collected. Participants were asked to fill the RCS-N at the beginning and the end of the nursing research course.

2.4.3. Data analysis

Statistical analyses were performed using Stata/SE 15.0 (Stata Corporation, Texas, USA). Construct validity was evaluated with factor analysis and item response theory (IRT) analysis. Internal consistency reliability was determined with Cronbach's α and IRT based reliability estimates. Sensitivity to change was evaluated with repeated measures ANOVA. Convergent validity was analyzed based on the correlation of the RCS-N scores with nursing research class test scores.

- IRT analysis:

IRT analysis is an approach that is used for the design, analysis, and scoring of scales. It is based on the assumption that a latent ability of a person on a certain trait can be measured with a set of items each of which may have a different difficulty and thus representing the ability in a different way. IRT analysis is based on several assumptions which are iteratively tested for, including unidimensionality of the trait, local independence of items, and ordering of thresholds (ordinal scales items) (Nering and Ostini, 2011).

- Construct validity:

To avoid clustering of the data due to repeated measurements, two random samples of students were drawn in a way that each student was represented only once in each sample while time points (before vs. after nursing research class) were equally represented. In this way, a development (n = 146, 78 students' pre-test results and 78 students' post-test results) and a validation sample (n = 146, the other 78 students' pre-test results and the other 78 students' post-test results) each covering the full response spectrum were available. To assess the dimensionality of the scale, exploratory factor analysis was then performed on the development sample. Dimensionality was assessed based on Eigenvalues, factor loadings, and the scree plot. Based on the results of exploratory factor analysis and confirmatory factor analysis (CFA), a structural equation model (Acock 2013) was performed on the validation sample. Modification indices were evaluated to detect local

dependency. Locally dependent items were collapsed to testlets and rescaled so that the original five response levels were preserved for each testlet. At each step the following goodness of fit indices were evaluated: 1) Root Mean Squared Error of Approximation (RMSEA), a measurement comparing our model to a fully saturated model with values smaller than 0.08 indicating acceptable fit and values smaller than 0.05 indicating excellent fit; 2) Comparative Fit Index (CFI), a statistic comparing the model with a baseline model with values above 0.95 indicative of good fit; 3) Standardized Root Mean Squared Residual (SRMR), with values below 0.08 indicating good fit; 4) coefficient of determination (R^2), with values closer to 1 indicating good fit (Hu and Bentler 1999; Barret, 2007). The resulting scale was then subjected to IRT analysis on both samples in order to determine the ordering of thresholds and representation of the full spectrum of abilities. The fit of a rating scale model against a partial credit model was assessed with the likelihood ratio test. Based on the final model, a metric score for the RCS-N was derived and transformed into a user-friendly total score ranging from 0 to 100.

- Reliability:

Internal Consistency Reliability was evaluated by counting Cronbach's alpha coefficient for the development sample with $\alpha \geq 0.70$ considered acceptable (Nunnally et al. 1967). Reliability estimates conceptualized as the consistency with which the scale could differentiate between two persons with different true abilities were calculated for the validation sample at different ability thresholds (theta −2, −1, 0, 1, 2) from the IRT model. The reliability coefficients ρ are thereby defined as the variance of the fitted model (1.0) minus the error variance at the respective ability data point (Thissen 2000).

- Sensitivity to change:

Sensitivity to change was analyzed with repeated measures ANOVA with two measurement points per subject (n = 146, 292 observations), i.e. before and after participation in the nursing research course held at the nursing school. Eta^2 was calculated to determine the effect size.

- Convergent validity:

Pearson correlation of the total score that students received based on several tests (quiz, paper analysis, research plan, and final test) conducted in the nursing research course with the transformed post-test RCS-N total score was analyzed to determine convergent validity (n = 146). In addition, the scatterplot of the research methods class total score and the RCS-N transformed total score was analyzed based on quadrants defined by the means of the two variables.

3. Results

3.1. Demographic Characteristics

149 of 153 students volunteered to participate in this study, and 146 (97.99%) of them completed both the pre and post-course survey. The rest 3 students only completed the pre-course survey, but not the post-course survey. Demographic characteristics of the participants are summarized in Table 1.

3.2. Content validity

The overall S-CVI of the RCS-N was 0.975, and the I-CVIs ranged from 0.80 to 1.0. In addition, some items were reworded and modified based on the experts' suggestions and comments.

Table 1
Demographic characteristics of the participants (N = 146).

Variables	Category	Number	%
Gender	Female	129	88.36
	Male	17	11.64
Age	≤20	58	39.73
	> 20	88	60.27
	Mean	20.79	–
Grade	Freshman	71	48.63
	Sophomore	75	51.37

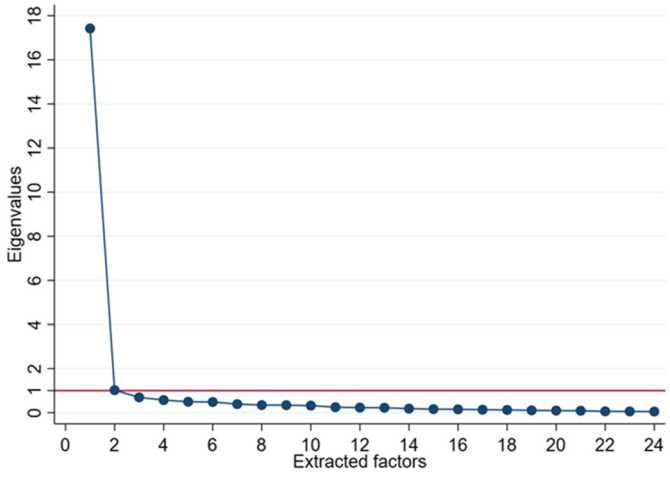

Fig. 1. Scree plot for exploratory factor analysis of the RCS-N (N = 78).

3.3. Construct validity

The exploratory factor analysis performed on the development sample yielded two factors with Eigenvalues > 1. With an Eigenvalue of 17.42 as compared to 1.02 for the second, there was however clearly a dominant first factor (see Fig. 1). The factor loading for exploratory and confirmatory factor analysis are shown in Table 2.

With Chi-squared (df = 252) of 1002.00 ($p < 0.001$), RMSEA of

Table 2
Factor loadings for exploratory (principal component = PCA) and confirmatory factor analysis (CFA) (N = 78).

Item	PCA	CFA
1. Basic steps of nursing research	0.86	0.85
2. Ways to formulate research questions	0.89	0.86
3. Statements of research questions	0.87	0.84
4. Databases for document retrieval	0.71	0.61
5. Types of databases for document retrieval	0.77	0.72
6. Ways and methods of document retrieval	0.79	0.74
7. Ethical requirements for research	0.73	0.74
8. Types of research design	0.90	0.89
9. Components of research design	0.92	0.90
10. Characteristics of different research designs	0.89	0.89
11. Concepts about population and sample	0.84	0.78
12. Sampling methods	0.86	0.83
13. Ways to estimate sample size	0.83	0.83
14. Data collection methods	0.86	0.86
15. Factors that affect research quality	0.86	0.86
16. Methods to improve research quality	0.88	0.87
17. Ways to measure instrument performance	0.90	0.86
18. Types of research data	0.88	0.90
19. Methods of statistical analysis	0.88	0.88
20. Formats of research articles	0.83	0.91
21. Ways to write research articles	0.85	0.92
22. Research article appraisal	0.88	0.88
23. Concept of evidence-based nursing	0.86	0.88
24. Steps of evidence-based nursing	0.87	0.86

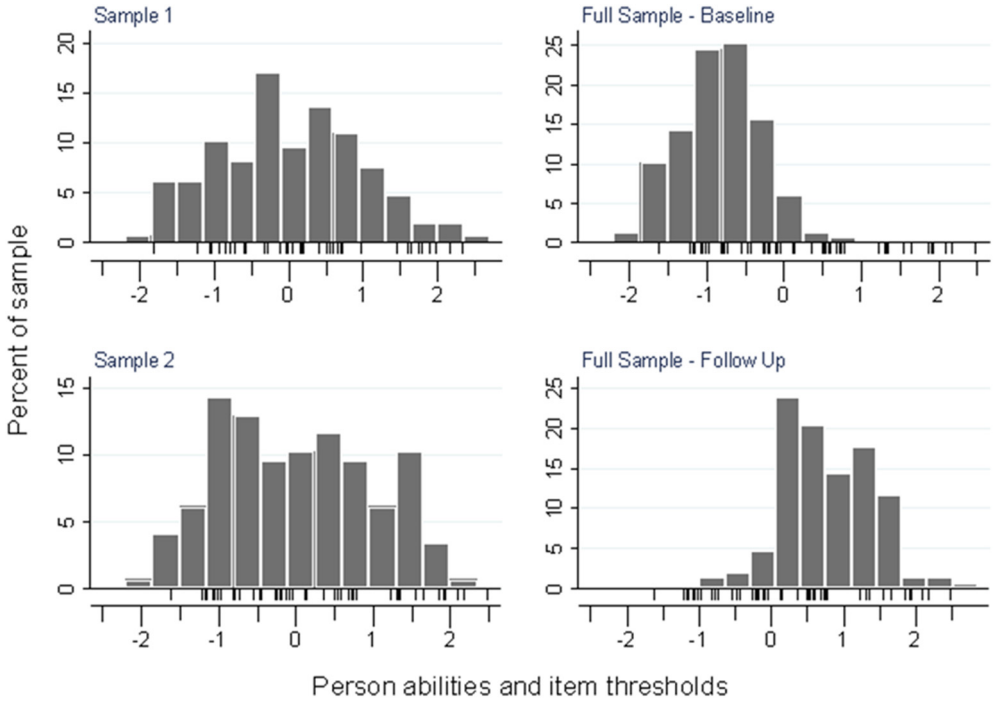

Fig. 2. Person-item maps for development sample (sample 1), validation sample (sample 2), and full sample at baseline and follow up (N = 146)

0.143, CFI of 0.84, and SRMR of 0.057 model fit was however suboptimal, indicating local dependency.

Modification indices for the validation sample suggested high correlations of the error variances of the following items: a) 1, 2, 3; b) 4, 5, 6; c) 8, 9; d) 11, 12, 13; e) 14, 15; f) 7, 16; g) 20, 21, 22; h) 23, 24 indicating local dependency. After collapsing the respective items to testlets, model fit was acceptable: Chi-squared (df = 54) = 99.91 ($p < 0.001$), RMSEA = 0.076; CFI = 0.978; SRMR = 0.021, R^2 = 0.977. In IRT analysis of the scale with testlets a partial credit model showed superior fit as compared to a rating scale model (LR $chi^2(33) = 128.39$, $p < 0.001$). All thresholds for the non-collapsed items were well ordered (threshold ordering is not expected for testlets) and the scale differentiated well between persons with different abilities (Fig. 2). For further analysis, the resulting logit scale was transformed into a user-friendly score ranging from 0 to 100. A transformation table for raw scores was created (Table 3).

3.4. Reliability

With a Cronbach's α of 0.98 reliability analysis showed good internal consistency of the RCS-N for the development sample. Reliability based on estimations from the IRT analysis was good for persons with very low ability (ρ = 0.87 for theta = −2) and excellent for all other ability data points (ρ > 0.95, for theta = −1,0,1,2).

3.5. Sensitivity to change

Scores increased significantly from 29.3 (95% CI 27.48–31.12) before the nursing research class to 62.6 (95% CI 60.8–64.5) after participation in the nursing research class ($F(1,145) = 6.02$, $p < 0.001$). With an eta^2 of 0.82, the effect size was large.

3.6. Correlations with nursing research class scores

There was no significant correlation between the RCS-N transformed total score and the class test score ($r = 0.16$, $p = 0.051$). Quadrant analysis of the scatterplot (Fig. 3) revealed that this was mainly due to 22 overly confident students in terms of RCS-N who did not achieve over average test results in the class. If these overconfident students were excluded, the correlation would have increased to $r = 0.45$ ($p < 0.001$).

4. Discussion

In this study, we evaluated the content validity and psychometric properties of the RCS-N, a tool to assess research competency of undergraduate nursing students. Content validity was confirmed by an expert panel and psychometric properties were adequate. Unidimensionality of the RCS-N was confirmed with both exploratory factor analysis as well as CFA after accounting for local dependency between sets of items by collapsing them to testlets. The result of the data could be fitted to a partial credit model and a metric score was created differentiating well between persons with various abilities. Reliability was good according to both classical (internal consistency) as well as modern test theory criteria. The RCS-N was moreover sensitive to change indicating significantly higher research competency after the students had participated in the research methods class. Convergence with the total test scores students achieved in the research methods class was however low, which may be due to some students who scored high on RCS-N but achieved below average test scores.

The latter result may be interpreted in the following way: A tool used to quickly assess students' awareness of several subjects relevant to conducting nursing research such as the RCS-N can of course not replace comprehensive testing of students' skills in terms of correctly answering questions about research methodology and applying knowledge to conduct clinical research. In this respect, the RCS-N rather measures confidence in one's research competency than actual knowledge of research methods. Overconfident students are particularly problematic in this respect because they may not invest enough time in learning for the research methods class tests. Therefore, in the teaching of nursing research methods in the future, teachers may increase continuous process evaluation, for example through introducing simulated examination at mid-term, in order to point students to their own shortcomings, to help them avoid being overconfident and to consolidate what they have learned. Those tests in combination with the RCS-N may also be used to identify the problematic group of

Table 3
Transformation table for converting raw score to metric score based on the estimates from the item response theory (IRT) model.

Raw score	IRT ability	0–100 score
12	−2.21	0
13	−1.82	8
14	−1.59	13
15	−1.43	17
16	−1.30	19
17	−1.19	22
18	−1.09	24
19	−1.00	26
20	−0.92	27
21	−0.84	29
22	−0.76	31
23	−0.69	32
24	−0.62	34
25	−0.55	35
26	−0.48	37
27	−0.41	38
28	−0.35	39
29	−0.28	41
30	−0.21	42
31	−0.15	44
32	−0.08	45
33	−0.02	46
34	0.05	48
35	0.12	49
36	0.19	51
37	0.26	52
38	0.33	54
39	0.40	55
40	0.47	57
41	0.55	58
42	0.63	60
43	0.71	62
44	0.79	63
45	0.88	65
46	0.96	67
47	1.05	69
48	1.14	71
49	1.23	73
50	1.33	75
51	1.43	77
52	1.53	79
53	1.64	81
54	1.75	83
56	1.99	89
58	2.31	95
59	2.53	100

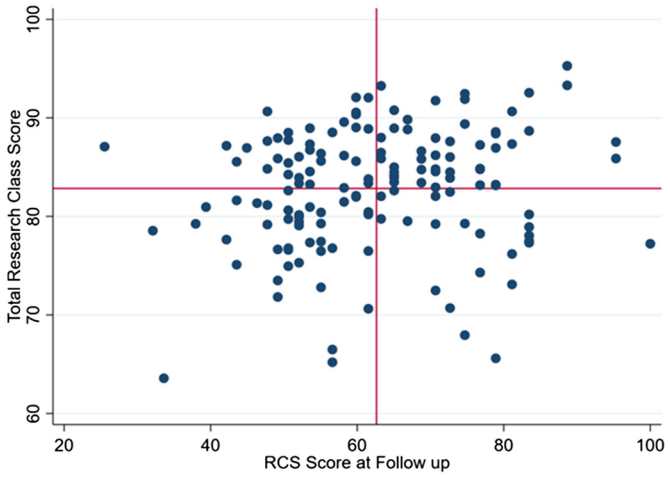

Fig. 3. Scatterplot of transformed RCS total scores and total research methods class test scores, quadrants are defined by means of the two variables (N = 146).

overconfident students.

As compared to other scales measuring nursing research competency, the RCS-N achieved similar internal consistency as the Research Capacity Scale ($\alpha = 0.99$) developed by Gething et al. (2001) and better internal consistency than the Research Ability Self-evaluation Rating Scales ($\alpha = 0.861$) developed by Pan et al. (2011).

While the RCS-N was unidimensional, other scales were considered multi-dimensional and could be divided into several sub-scales, such as theoretical learning, literature retrieval, data collection, data analysis, article writing etc. (Pan and Cheng 2011; Gething et al. 2001). This may be related to the fact that our participants were undergraduate nursing students, with their main task being curriculum learning.

To our knowledge, no other scale measuring nursing research competency has been evaluated with modern test theory, nor has their sensitivity to change and convergent validity been evaluated, so that there are no other studies we could compare our respective findings with.

The main limitations of this study include the small sample size and the concentration on a particular nursing school which limiting the generalizability of our findings. Further testing with larger cohorts and in other contexts is thus warranted in order to examine the applicability of the RCS-N to other nursing schools. Besides, the CFA model set up here based on the results of the exploratory factor analysis could be one of many models that may fit the data. Future research may test different models and compare fit indices with the one used here. Because of the small and relatively homogenous sample we also refrained from testing for differential item functioning. Moreover, test-retest reliability for the RCS-N still needs to be established.

In conclusion, the RCS-N developed in this study is a promising, valid and reliable tool for evaluating the research competency of undergraduate nursing students. The RCS-N can however not replace actual written and oral examination of students' knowledge acquired in research methodology courses, though it may be helpful to identify the students' subjective confidence in their research competencies.

Acknowledgements

This research is supported in part by New Century Higher Education Teaching Reform Project (Phase 8) of Sichuan University under grant SCU8364.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nedt.2019.05.039.

References

Acock, A.C., 2013. Discovering Structural Equation Modeling Using Stata: Revised Edition. (Stata press).
Barrett, P., 2007. Structural equation modeling: adjudging model fit. Personal. Individ. Differ. 42 (5), 815–824. https://doi.org/10.1016/j.paid.2006.09.018.
Flinkman, M., Leino-Kilpi, H., Numminen, O., Jeon, Y., Kuokkanen, L., Meretoja, R., 2017. Nurse competence scale: a systematic and psychometric review. J. Adv. Nurs. 73 (5), 1035–1050.
Gerrish, K., Lacey, A., 2010. The Research Process in Nursing. John Wiley & Sons.
Gething, L., Leelarthaepin, B., Burr, G., Sommerville, A., 2001. Fostering nursing research among nurse clinicians in an Australian area health service. J. Contin. Educ. Nurs. 32 (5), 228–237.
Hu, L.-t., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. 6 (1), 1–55.
Nering, M.L., Ostini, R. (Eds.), 2011. Handbook of Polytomous Item Response Theory Models. Taylor & Francis.
Nunnally, J.C., Bernstein, I.H., Berge, J.M.T., 1967. Psychometric Theory. vol. 226 McGraw-Hill, New York.
Pan, Y. H., Cheng, J. L. (2011). Revise of self-evaluation research competency scale for nursing staffs. Chinese Nursing Research, (13):1205–1208. doi: https://doi.org/10.3969/j.issn.1009-6493.2011.13.040.
Pan, Y.H., Cheng, J.L., Kou, L.H., Sun, Y.M., 2011. Investigation on scientific research capacity of nursing staff in Shanxi province. Chin. Nurs. Res. 25 (22), 1989–1992. https://doi.org/10.3969/j.issn.1009-6493.2011.22.009.

Parahoo, K., 2014. Nursing Research: Principles, Process and Issues. Palgrave Macmillan.
Polit, D.F., Beck, C.T., 2004. Nursing research: Principles and methods. Lippincott Williams & Wilkins, Philadelphia, London.
Thissen, D., 2000. Reliability and Measurement Precision. Lawrence Erlbaum Associates Publishers.
www.chinadegrees.cn, 2018. China university subject rankings (cusr). http://www.chinadegrees.cn/xwyyjsjyxx/xkpgjg/index.shtml April.
Zhang, J., Ye, W., Fan, F., 2015. Development of a self-assessment tool for measuring competences of obstetric nurses in rooming-in wards in China. International Journal of Clinical & Experimental Medicine 8 (10), 18548.
Zhao, L., Han, S., Lu, B., Wang, M., Zhao, X., 2011. Study on nursing research capacity of nursing undergraduates and its influencing factors. J. Nurs. Sci. 26 (4), 64–66. https://doi.org/10.3870/hlxzz.2011.04.064.
Zhou, Q., Zhang, L., Zhang, G., 2001. Comparative analysis of nursing staff's nursing scientific research awareness and practice. International Journal of Nursing 30 (6), 917–920. https://doi.org/10.3760/cma.j.issn.1673-4351.2011.06.063.