Validation of data using RASCH analysis in a tool measuring changes in critical thinking in nursing students

Elisabeth R. Jacob[a,*], Christine Duffield[a,b], Alycia M. Jacob[a]

[a] School of Nursing & Midwifery, Edith Cowan University, 270 Joondalup Drive, Joondalup, WA 6027, Australia
[b] Centre for Health Services Management, Faculty of Health, University of Technology Sydney, Australia

A B S T R A C T

Background: In Australia there is no national examination to assess student
competency with respect to critical thinking skills on graduation from undergraduate registered nurse program.
Objectives: To describe the use of Rasch analysis for the validation of a tool that measures critical thinking in nursing students in Australia.
Design: A cross sectional survey was used to assess students' critical thinking.
Settings: One Australian university.
Participants: Seventy-four first year and third year nursing students participated in the project.
Methods: Nursing students at one university were requested to undertake a multiple choice question survey to determine their level of critical thinking. The survey tool was validated using Rasch analysis software.
Results: Rasch analysis enabled identification of the difficulty of each question, enabling the assessment to identify difference between students and for each student cohort. The use of Rasch analysis demonstrated marked differences in the students' critical thinking abilities, showing higher levels of critical thinking for third year students with greater ability to answer the more difficult questions.
Conclusion: Rasch analysis enables the matching of test questions to the cognitive skill level of students, and is a useful tool for validating multiple choice questionnaires. Use of this method of analysis can provide a robust assessment in which questions are matched to the level of the learner and enable differentiation between high and low achieving students.

1. Background

In Australia completion of an accredited undergraduate nursing program entitles a person to apply for registration as a nurse; a final national graduation exam is not required (Nursing and Midwifery Board of Australia, 2012). Pre-registration nursing programs in Australia include multiple ongoing assessments, completion of which assumes that nurses have obtained an acceptable level of knowledge, skills and critical thinking for safe practice. There is no uniform national measure used to ascertain the level of knowledge and critical thinking skills held by graduate nurses here (Jacob et al., 2017). In contrast, licensure exams to demonstrate appropriate knowledge and level of critical thinking to safely practice as a registered nurse are required in many other countries such as United States of America and Canada, or overseas educated nurses wanting to register in the United Kingdom (National Council of State Boards of Nursing, 2018; Nursing and Midwifery Council, 2017).

Critical thinking is a conscious and deliberate use of intelligent thought processes that looks to arrive at reasonable situationally aware judgements (Perez et al., 2015). It is associated with complex tasks,

requiring a degree of intelligence to complete, and draws on a base of existing knowledge which is applied to a cognitive process (Brudvig et al., 2013; Willis, 2004). Critical thinking is an essential component of competent nursing practice. It is essential to a nurse's role to improve patient outcomes in a timely and efficient manner (Carter et al., 2015; Castledine, 2010; Kutney-Lee et al., 2009; Perez et al., 2015). As nurses care for patients over extended periods, they are often able to detect early signs of deterioration or complications before other health care personnel do so (Clarke & Aiken, 2003; Jacob et al., 2017; Kutney-Lee et al., 2009). However, there is a lack of evidence around how, if at all, critical thinking skills are evaluated (Perez et al., 2015). This paper will describe the use of Rasch analysis to validate a tool that measures critical thinking in nursing students.

1.1. Rasch analysis

Rasch analysis is a method of evaluating internal consistency of a test (Coombes, 2013). It evaluates whether high scores consistently include correct answers to easy questions and low scores consistently do not include correct answers for harder questions. Rasch analysis looks

---

* Corresponding author.
E-mail address: e.jacob@ecu.edu.au (E.R. Jacob).

at a questionnaire as a whole, providing visual representations of the way that individual questions can be rescaled or removed in order to better fit the data and also, to target the specific criteria that the questionnaire seeks to measure (Hendriks et al., 2012).

Rasch analysis also provides a method of validating a questionnaire to match the difficulty level of items with the individual level of students undertaking the test (Hendriks et al., 2012; Tractenberg et al., 2013). Questions can be placed in an order where easy questions can be undertaken first to enable students to build confidence prior to undertaking more difficult questions (Tractenberg et al., 2013). It is possible to measure the difficulty of individual tasks (test items) using this technique, and compare both individuals and large cohorts in accordance with their performance on tasks of varying difficulties (Hagquist et al., 2009). When an assessment contains questions with a wide range of levels of difficulty, students who have a higher level of proficiency are able to demonstrate their advanced skills more clearly (Jacob et al., 2017).

2. Method

The project was conducted using an electronic survey. The survey was created with input from nursing experts who contributed via participation in a Delphi study. Development of the survey has been previously reported (Jacob et al., 2018).

The survey consisted of demographic data, and five unfolding scenarios covering a variety of common clinical situations which incorporated 25 multiple choice questions relating to these scenarios. The scenario questions were designed to test decision-making and judgement skills. The survey was constructed in Qualtrics. Ethical approval for the study was obtained from Edith Cowan University.

A power analysis was undertaken through the G-Power program to determine the required number of participants. Power analysis at alpha level of 0.1 and power level of 80% required a sample size of 82 participants as the effect size was considered moderate.

Participants were students in a registered nurse program at an Australian higher education institution. Surveys were sent electronically to students in the first and final semester (end of year three) of their nursing program via student email accounts. No personal details were recorded so individual responses were anonymous.

Data were analysed using Rasch analysis through the program WINSTEPS Version 4.0.1. This statistical program facilitates data from individual students to be entered to enable the production of Wright maps and Mean Square Errors. Wright maps are a visual representation of survey data which enable comparison of students which scored high on the score with how difficult the questions were using two vertical histograms with an integrated scale. The placement of students' results relevant to the placement of the question items can provide information about the appropriateness of the questions in the test to the student's skill level. The closer the spread of students is to the spread of responses, the better the test is able to accurately measure the student's level, and the more closely the levels of difficulty of questions are matched to the cohort. A mean square error score demonstrates how predictable the assessment is to the Rasch model: It determines the difference between the model and the data. The aim is for a mean-squares error score of 1.0 (Wright and Linacre, 1994).

3. Results

A total of 128 students commenced the survey. After removal of incomplete results, 74 surveys were analysed. Of these completions, 50 surveys were from first year nursing students and 24 were from third year students. The majority of surveys (71, 95%) were completed by local students (i.e. Australian residents) and the remainder by international students.

The age of participants varied with 17 people aged 17–20; 18 aged 21-25 yrs.; 27 aged 26–40 and 12 people over 40+ years. A large
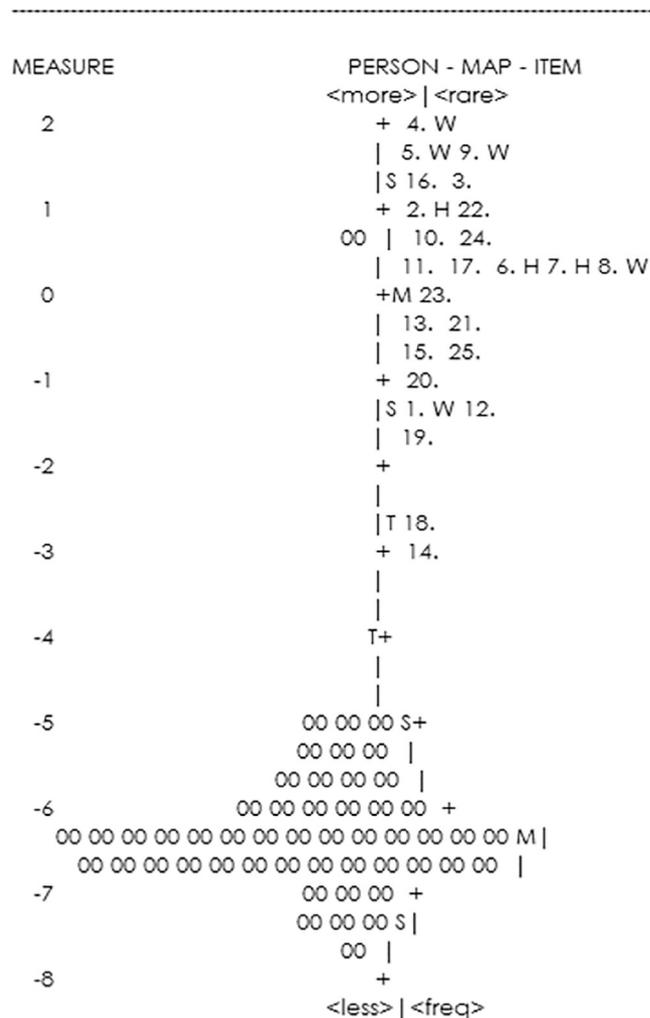

Fig. 1. First year nursing students' Wright map. Average student score: 14.48/25; High student score — 20; low student score — 8.

number of participants indicated that they had previously worked in the healthcare sector (30 students, 41%) as either Enrolled Nurses [also known as licensed practical nurses or division two nurses] ($n = 2$), personal care assistants [unregulated health professionals] ($n = 14$) or other healthcare workers such as dental assistant, ambulance officer, laboratory technician or support worker ($n = 14$).

Participants' educational level varied. Two participants held postgraduate qualifications, eight had previously completed an undergraduate degree, 34 participants had a certificate level qualification and 28 participants had only high school qualifications.

Data from questionnaires were manually entered into Rasch analysis software (Winsteps Version 4.1.0). Analysis of the data involved comparison of the results of two student cohorts using Wright Maps and Mean square error.

Wright maps were produced (Figs. 1 and 2) to demonstrate how well the level of difficulty of the questions matched up with the skill level of the students. Results were recorded for the first and third year students separately. The results were then analysed using Rasch modelling. The individual student's results (shown as OO) are on the left of the vertical axis and the questions as numbers (1–25) on the right of the axis. Results are ranked along a vertical axis according to difficulty. More difficult items and higher scoring students are positioned higher up the axis, while lower scoring students and less difficult questions are found towards the bottom. Students who are positioned level with a question are considered to be 50% likely to get that question correct.
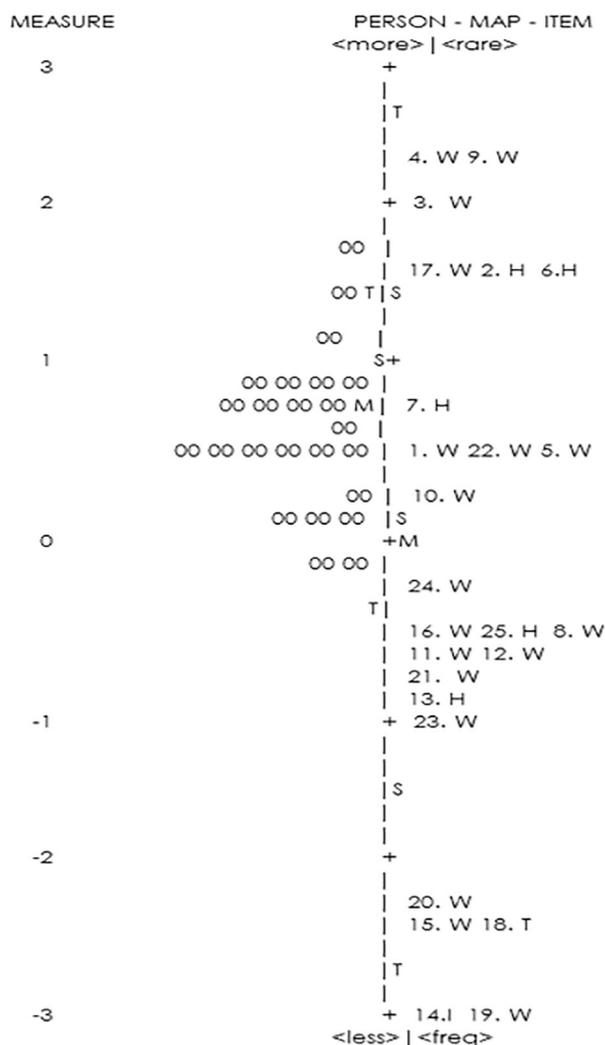
Fig. 2. 3rd year nursing students' Wright map. Average student score: 15.77/25; High student score — 20; low student score — 13.

The results between the first and third year nursing students were markedly different demonstrating differences in student ability and the difficulty level of the questions (Figs. 1 & 2). This shows an expected difference in skill level between the groups.

Fig. 1 demonstrates that the majority of first year nursing students (*n* = 50) had similar skill levels demonstrated by the grouping of their results. The Wright map demonstrates that the first year students would have found the test quite difficult for their skill level as all of the students sat well below the difficulty level of the lowest test item. The mean ability of the students was seven Standard Deviations below the mean difficulty of the questions. No questions were at the same ability level as the majority of students.

One student, who may have had prior health experience, was an outlier with results significantly above those from other students. The highest scoring question in the survey was question 14 with 49 (98%) of first year students selecting the correct answer. This question asked students to assess whether a particular patient's blood pressure was of concern. The lowest scoring question was question 4, with only 10 students (20%) selecting the correct answer. Question 4 dealt with the factors that would require consideration prior to placing oxygen on a patient.

Fig. 2 demonstrates that in contrast to the first year students, the third year students (*n* = 24) found the test moderately easy for their

ability level. The majority of students scored slightly above the mean for the question. The spread of data for third year students demonstrates a high level of discrimination of results between students. The highest scoring question was the same as the first year students (Question 14) with all third year students selecting the correct answer. Similarly, the lowest scoring question (Question 4) was also the lowest scoring for first year students, with only four (17%) third year students selecting the correct answer. This demonstrates that the questions differentiated between the high achieving students and those with less ability in critical thinking.

Mean square error was calculated as 0.87 for the first year students and 0.86 for the third year students, indicating that the data were what was expected for the model. The range of 0.8–1.2 is recommended for multiple choice question assessments in which large marks are awarded (Wright and Linacre, 1994).

4. Discussion and conclusion

Rasch analysis is a useful tool when validating a multiple choice questionnaire for assessment of a specific cohort as it provides a platform for matching test questions to the skill level of students who are being assessed (Hendriks et al., 2012; Tractenberg et al., 2013). Identification of questions that were easy and difficult for students allows questions to be included based on the ability of students to complete them. Questions can be adjusted in assessments to enable differentiation between high achieving and average achieving students. Use of Rasch analysis demonstrated that the tool developed was able to distinguish between students' responses in the more advanced cohort and those in first year by identifying students for whom questions posed a greater challenge.

Using Rasch analysis enabled the creation of Wright maps that assisted in visualising the way that students' skill level matched with the level of difficulty of questions. Whilst previously a multiple choice format test was considered inadequate to detect change in development (Carter et al., 2015), the use of Rasch measurement validates the ability of the tool to demonstrate differences in student abilities, both within and between cohorts. The comparison highlights the impact of targeting an assessment to the appropriate cohort. It is clear that the critical thinking assessment was more suited to the skill level of a third rather than a first year student.

Despite the fact that critical thinking is a specific criterion in the accreditation of nursing programs (Zuriguel Perez et al., 2015) in this country, it is a difficult skill to measure (Brudvig et al., 2013). The analysis presented here, using Rasch analysis, demonstrated that the critical thinking tool developed differentiated between the critical thinking skills of first and third year nurses. This validation of the tool means that it can be used to determine the critical thinking ability of nursing students in Australia as they progress through their undergraduate program. Perhaps more importantly, the ability to reliably assess critical thinking skills in nursing students may increase the confidence of healthcare personnel in the new graduates' ability to provide appropriate and safe patient care (Jacob et al., 2017).

Use of Rasch analysis is recommended for evaluation of multiple choice questions in nursing assessments. This approach to questionnaire design enables a more robust assessment in which questions can be matched to the difficulty level of the learner and enable differentiation between high and low achieving students.

References

Brudvig, T.J., Dirkes, A., Dutta, P., Rane, K., 2013. Critical thinking skills in health care

professional students: a systematic review. Journal of Physical Therapy Education 27 (3), 12–25.

Carter, A.G., Creedy, D.K., Sidebotham, M., 2015. Evaluation of tools used to measure critical thinking development in nursing and midwifery undergraduate students: a systematic review. Nurse Educ. Today 35 (7), 864–874. https://doi.org/10.1016/j.nedt.2015.02.023.

Castledine, G., 2010. Critical thinking is crucial. Br. J. Nurs. 19 (4), 271. https://doi.org/10.12968/bjon.2010.19.4.46792.

Clarke, S., Aiken, L., 2003. Failure to rescue. Am. J. Nurs. 103 (1), 42–47.

Coombes, J. (2013) Practice based competency development – A study of resource geologists and the JORC code system, PhD Thesis, Edith Cowan University.

Hagquist, C., Bruce, M., Gustavsson, J.P., 2009. Using the Rasch model in nursing research: an introduction and illustrative example. Int. J. Nurs. Stud. 46 (3), 380–393. https://doi.org/10.1016/j.ijnurstu.2008.10.007.

Hendriks, J., Fyfe, S., Styles, I., Skinner, S., Merriman, G., 2012. Scale construction utilising the Rasch unidimensional measurement model: a measurement of adolescent attitudes towards abortion. Australas Med J 5 (5), 251–261. https://doi.org/10.4066/AMJ.2012.952.

Jacob, E., Duffield, C., Jacob, D., 2017. A protocol for the development of a critical thinking assessment tool for nurses using a Delphi technique. J. Adv. Nurs. 73 (8), 1982–1988. https://doi.org/10.1111/jan.13306.

Jacob, E., Duffield, C., Jacob, D., 2018. Development of an Australian nursing critical thinking tool using a Delphi process. J. Adv. Nurs. 74 (9), 2241–2247 (doi:10.111/jan.13732).

Kutney-Lee, A., Lake, E.T., Aiken, L.H., 2009. Development of the hospital nurse surveillance capacity profile. Research in Nursing & Health 32 (2), 217–228. https://doi.org/10.1002/nur.20316.

National Council of State Boards of Nursing. (2018) NCLEX & Other Exams Accessed 28th March 2018 from https://www.ncsbn.org/nclex.htm

Nursing and Midwifery Board of Australia. (2012) National Board strategic plan 2013 – 2015, Accessed 28th March, 2018 from http://www.nursingmidwiferyboard.gov.au/About/Strategic-plan.aspx

Nursing and Midwifery Council. (2017). Registering as a Nurse or Midwife in the UK. Accessed 28th March, 2018 from https://www.nmc.org.uk/globalassets/sitedocuments/registration/registering-nurse-or-midwife-outside-eu.pdf

Perez, E.Z., Canut, M.T.L., Pergueroles, A.F., Llobet, M.P., Arroyo, C.M., Merino, J.R., 2015. Critical thinking in nursing: scoping review of the literature. Int. J. Nurs. Pract. 21, 820–830. https://doi.org/10.1111/ijn.12347.

Tractenberg, R.E., Gushta, M.M., Mulroney, S.E., Weissinger, P.A., 2013. Multiple choice questions can be designed or revised to challenge learners' critical thinking. Adv. Health Sci. Educ. 18 (5), 945–961. https://doi.org/10.1007/s10459-012-9434-4.

Willis, T. J. (2004). Affective dispositions and cognitive skills in critical thinking: Implications for measurement, training and team performance. http://scholarcommons.usf.edu/etd/1304.

Wright, B.D., Linacre, J.M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8(3) Accessed 2nd May 2018 from https://www.rasch.org/rmt/rmt83b.htm