DR. ROBYN CANT (Orcid ID : 0000-0001-5430-3275)

DR. EMMA BALLARD (Orcid ID : 0000-0002-9430-2485)

PROF. SIMON COOPER (Orcid ID : 0000-0001-5561-3099)

PROF. TRACY LEVETT-JONES (Orcid ID : 0000-0003-4279-8957)

PROF. LISA MCKENNA (Orcid ID : 0000-0002-0437-6449)

DR. PHILIPPA SEATON (Orcid ID : 0000-0003-0136-4175)

DR. LINDA NG (Orcid ID : 0000-0002-6155-7625)

Article type : Review

Manuscript

TITLE PAGE

Locating 'gold standard' evidence for simulation as a substitute for clinical practice in pre-licensure health professional education: A systematic review.

AUTHORS:

Fiona E Bogossian, PhD – Professor [a,g,h]

Robyn P Cant, PhD – Associate Professor [b]

Emma L Ballard, PhD – Senior Statistician [f,g]

Simon J Cooper, PhD – Professor [b]

Tracy L Levett-Jones, PhD – Professor [c, h]

Lisa G McKenna, PhD – Professor [b,d]

Linda C Ng, PhD – Lecturer [g]

Philippa C Seaton, PhD – Associate Professor [e,i]

[a] University of the Sunshine Coast, Sippy Downs, Queensland, Australia

[b] Federation University Australia, Victoria, Australia

[c] University of Technology Sydney, New South Wales, Australia

[d] La Trobe University, Melbourne, Victoria, Australia

[e] University of Otago, Christchurch, New Zealand

[f] QIMR Berghofer, Medical Research Institute, Brisbane, Queensland, Australia

[g] The University of Queensland, St Lucia, Queensland, Australia

[h] Monash University, Melbourne, Victoria, Australia

[i] The University of Melbourne, Melbourne, Victoria Australia

CORRESPONDING AUTHOR:

Associate Professor Robyn Cant
Nursing and Healthcare Professions
Federation University Australia
Northways Rd, Churchill
VIC 3842, Australia
E: r.cant@federation.edu.au
M: +61 438 363634

ABSTRACT:

Aims and objectives: To extract, examine and report the highest available levels of evidence from healthcare disciplines in the use of simulation-based education as substitution for clinical placement in pre-licensure programs.

Background: Simulation is widely employed across pre-licensure health professional education to create safe, realistic clinical learning experiences for students. Whether simulation can be employed to substitute for actual clinical placement, and if so, in what proportion, replacement ratio and duration, is unclear.

Methods: A systematic review and quality appraisal of primary studies related to pre-licensure students in all health disciplines, guided by the PRISMA checklist.

Results: Ten primary studies were included, representing 2,370 students from three health disciplines in four countries. Eight studies were experimental and quasi-experimental and methodological quality was assessed as moderate to high with good to very good interrater agreement. Direct substitution of simulation for clinical practice ranged from 5-50%. With one exception, replacement ratios were 1:1 and duration of replacement ranged from 21 hours to two years. Levels of evaluation included measures of reaction, knowledge and behaviour transfer; no negative outcomes were reported. We appraised practicalities for design of substitution, design limitations and knowledge transfer to accreditation standards for pre-licensure programs.

Conclusions: This review synthesised highest levels and quality of evidence for substitution of simulation for clinical placement in health professional education. Included studies were heterogenous in simulation interventions (proportion, ratio and duration) and in evaluation of outcomes. Future studies should incorporate standardized simulation curricula, widen the health professions represented and strengthen experimental designs.

Relevance to clinical practice: Current evidence for clinical educational preparation does not appear to be translated into program accreditation standards governing clinical practice experience for pre-licensure programs in relevant jurisdictions. Overall, a stronger evidence base is necessary to inform future curricula and policy development, to strengthen clinical practice in health.

KEYWORDS: clinical practice, clinical simulation, evidence translation, health occupations, practicum, simulation education, students, systematic review, workforce education

IMPACT STATEMENT

- Globally, recommendations are explicit that pre-licensure curricula have criteria in place that meet accreditation standards for clinical practice components of their programmes. Yet accreditation standards do not have a strong evidence base.

- This paper synthesises the highest levels and appraises the quality of evidence for substitution of simulation for clinical placement in health professional education.
- There is a need for a stronger evidence base to inform future curricula and policy development and enhance translation of evidence into accreditation standards governing clinical practice experience for pre-licensure programs.

INTRODUCTION

The fundamental purpose of pre-licensure education across the health disciplines is to produce a capable, competent workforce that can provide safe, high quality healthcare services. In past decades, the education of some professions in the health workforce has moved from an apprenticeship hospital training model with high levels of clinical exposure into the tertiary education sector. This transition is now well supported by good evidence that a more highly qualified workforce not only improves safety and quality of care, but improves patient outcomes (Aiken, Clarke, Cheung, Sloane, & Silber, 2003).

Globally, recommendations are explicit that pre-licensure curricula have criteria in place that meet accreditation standards for clinical practice components of their programmes. The required number of clinical placement hours are commonly mandated by professional regulatory bodies. However, while there is a global need to provide evidence-based educational programmes, there is a limited evidence base. Wide variation exists in mandated program hours globally (Coyle, 2007) and more specifically, in relation to clinical learning hours and how these are constituted. Nonetheless, challenges in health workforce education include the provision of sufficient high quality clinical learning experiences for pre-licensure students, in the face of demand for increasing numbers of graduates in response to the global workforce crises. These health workforce challenges require reappraisal of past strategies and a paradigm shift in how we educate health care workers (World Health Organisation, 2016).

2. BACKGROUND

Simulation-based education has emerged as an essential element of pre-licensure education for healthcare learners. It provides engaging and authentic learning opportunities during realistic 'life-like' simulated clinical experiences (Cantrell, Franklin, Leighton, & Carlson, 2017). Learners often work in small teams of 3-6 students in interactive role-plays that may include low fidelity task trainers (models), programmable mannequins, simulated patients (actors), or virtual reality and computerized on-screen simulations (Cant & Cooper, 2017).

The literature cites numerous advantages of simulation for learning, including the benefit of enabling repeated practice of technical and non-technical skills as preparation for clinical practice (Motola, Devine, Chung, Sullivan, & Issenberg, 2013). It can offer exposure to uncommon clinical situations that, if encountered in real life, learners could only passively observe (Levett-Jones et al., 2010). A fundamental learning opportunity is the provision of formative and summative feedback, with learners able to reflect on practice, assisting the development of competence (INACSL., 2016; Motola et al., 2013). With the rapid increase in simulation scholarship, we sought to substantiate current knowledge for substitution of clinical practice hours with simulation-based education across the health care disciplines, using the best available or gold standard evidence.

Malina (2016) provides an historical account of the development of the randomised controlled trial (RCT) as a method to reduce bias and enhance the accuracy of experimentation in clinical research. She argues that by the 21st century RCTs had achieved the gold standard for therapeutic evidence, but that this method is not a single source of evidence nor a stable technique. RCTs generally aim to determine whether one intervention is better than another, however, inadequate reporting can lead to bias in effects (Piaggio, Elbourne, Altman, Pocock, Evans, & Consort Group, 2006). Despite the limitations of the method, when ideally performed the double-blind RCT is accepted as an objective scientific methodology that produces knowledge untainted by bias (Kaptchuk, 2001). Although having its origins in scientific clinical trials, the RCT is now widely regarded as the highest level of primary evidence, but it may not suit all disciplines; in which case, quasi-experimental and observational methods may need to be considered.

3. AIMS AND METHODS

The aim of this study is to extract, examine and report the highest levels and quality of evidence from healthcare disciplines in the use of simulation-based education as substitution for clinical placement in pre-licensure programs. The following research questions will be answered:

(i) What is the level of research evidence and the quality of primary studies investigating simulation - based education as a substitute for a proportion of pre-licensure clinical placement hours in health care disciplines?

(ii) What are the measures and outcomes used and does the evidence demonstrate statistically or clinically significant differences or equivalence for evaluation outcomes when simulation-based education is substituted for a proportion of pre-licensure learners' clinical placement?

(iii) If evidence supports the use of simulation-based education as a substitute for clinical placement, what is the optimal proportion of simulation hours versus clinical placement hours, in what replacement ratio, for what durations and, how is this evidence translated into pre-licensure professional education standards in health care disciplines?

3.1. Procedure

A systematic review of literature was conducted to determine the status of current evidence. Systematic reviews can provide a high level of evidence based on a summary of identified carefully designed trials (Cochrane ref). We used the systematic review process to critically appraise the evidence using a clearly documented methodology, to answer a research question. The PRISMA checklist (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was used to guide the study and reporting (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009) (See Supplementary File 1).

3.2. Search strategy

We conducted multiple database searches were conducted in a stepwise fashion. To enable an overview of relevant publications, open text searches were conducted using Google Scholar. Via the US National Library of Medicine, 'MeSH on Demand' was used to identify relevant Medical Subject Headings terms from collected abstracts and this provided direct links to some relevant articles. Second, a systematic search strategy was established in order to conduct electronic searches of six databases. These were the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Medical Literature Analysis and Retrieval System Online (Medline), SCOPUS, PubMed, Education Resources Information Center (ERIC), and The Cochrane Library. In a third stage, resources such as the Australian and New Zealand Clinical Trials Register, Joanna Briggs Institute (JBI) Database of Systematic Reviews and Implementation Reports, the journal 'Trials' and websites of professional bodies such as Nursing and Midwifery Councils, were examined for relevant information. Hand searches of relevant journals (for example Clinical Simulation in Nursing, Journal of Nursing Regulation) were conducted and already acknowledged ancestral articles were all gathered in an electronic library.

We used a methodical search strategy based on 'participants', 'interventions', 'comparisons', 'outcomes', and 'study design' (PICOS). This asked: for healthcare students, does substitution of clinical practice hours with simulation-based education affect learning outcomes in experimental or quasi-experimental studies? The key search terms were based on medical subject headings (MeSH) and, for example, 11 single terms were used: clinical practicum; clinical practice; substitution; students, health professions; patient simulation; simulation training; experimental design, case-control studies; randomized controlled trial; medical education; students, nursing; and their Boolean combinations.

3.3. Inclusion and exclusion criteria

We determined that articles published in English up until February 2018 were eligible for inclusion. Studies that reported interactive simulation incorporating goal-based role-play to enable healthcare students to practice technical or non-technical skills, were selected. Three authors confirmed the criteria used were appropriate. All modalities of simulation - high, medium and low fidelity - were eligible. While quantitative experimental studies were the

main focus, studies of all designs were screened for eligibility as even non-research sources (such as policy documents) could form secondary sources of information.

3.4. Selection process

Citations and abstracts were downloaded into an Endnote database and were screened by title and abstract based on the criteria established for inclusion. The study variables of interest were tabulated including study origin, design, sample, participants, measures, validity of measures used, educational findings, and outcomes. Full text articles were read initially by one author who tabulated the study characteristics. Because the focus of the review was on substitution of clinical practice with simulation, four simulation studies which did not address direct substitution

3.5. Synthesis

All authors confirmed the 10 primary studies to be included in the review. Studies were grouped according to the applicable JBI Levels of Evidence - Effectiveness (Joanna Briggs Institute, 2014) where Level 1 is Experimental designs, Level 2 Quasi-experimental designs, Level 3 Observational-analytic designs and Level 4 Observational-descriptive studies. All RCTs (JBI Level 1.c) were also evaluated using the Consolidated Standards of Reporting Trials (CONSORT) Statement checklists for non-inferiority and equivalence trials (Piaggio, Elbourne, Altman, Pocock, Evans, & Consort Group, 2006), pilot or feasibility trials (Eldridge et al., 2016) or parallel group randomised trials (Moher et al., 2009) where applicable.

Elements of interest were carefully extracted in order to respond to the research questions. This required extrication of information and data from each study relating to: the simulation intervention substitution proportion, ratio and duration; determination of rates of recruitment, completion and losses to follow up; identification of the evaluation measures used and their reported validity and reliability and; classification of outcomes of evaluation. These extraction procedures were conducted initially by two authors (RC & FB) and details clarified, confirmed and extrapolated by a third (EB).

In order to classify outcomes, we used Kirkpatrick's four-level training evaluation scheme which had been adapted for higher education assessment (Praslova, 2010) and made slight descriptive modifications to clearly align with education outcomes for health disciplines. This

model includes four levels of evaluation: Level 1: *Reaction* (students' affective reactions and utility judgements e.g. degree of satisfaction); Level 2: *Learning* (direct measures of learning outcomes, knowledge tests, performance tasks or other graded work e.g. changes in knowledge and skills); Level 3: *Behaviour/transfer* (evidence of student transfer of knowledge and skills in the clinical context/situation), and Level 4: *Results* (improvements in patient outcomes and/or organisational change).

3.6. Methodological quality appraisal

To limit bias, five authors participated in objectively assessing the quality of studies using the Medical Education Research Study Quality Instrument (MERSQI). The MERSQI was designed to measure the methodological quality of experimental, quasi-experimental, and observational studies (Reed, Beckman, & Wright, 2009). The 10-item scale comprises six domains: study design; sampling; type of data; validity of measurement instruments; data analysis, and outcomes. The maximum domain score is 3, maximum total score is 18, with a potential range of 5-18. The instrument has been found reliable, with strong intra-class correlation coefficients for inter-rater (0.72 to 0.98) and intra-rater (0.78 to 0.99) agreement; scores were associated with journal impact factor, amount of study funding, and journal editorial decisions (Reed et al., 2009). The MERSQI has been used to evaluate the methodological quality of healthcare research, particularly in reviews of medical research (DeCoste-Lopez, Madhok, & Harman, 2015; Hsieh et al., 2016)

Inter-rater agreement for the independently derived quality scores was computed using a Kappa Measure of Agreement based on two scores for each study. Five researchers, between them, provided assessments for 10 studies. The Kappa statistic accounts for the proportion of agreement between two raters that could have occurred by chance. A value of .7 represents good agreement and a value above .8 represents very good agreement. With the aim of reaching strong agreement between raters ($k = \geq .7$), any differences were discussed but variations were permitted to remain as in some studies, the reporting of detail (design, instruments) could be unclear. The overall reliability of ratings was computed using the Intra-class Correlation Co-efficient. The scores are reported based on the average scores of two raters.

4. RESULTS

4.1. Levels of Evidence

The literature reporting empirical studies of the substitution of clinical experience for pre-licensure health professions with simulation, is sparse. In this review, we identified 10 primary studies that present evidence (see Table 2). The levels of evidence of the studies ranged from Level 1 to Level 4 (Joanna Briggs Institute, 2014). Six studies were randomized controlled trials (1c, 1d) (Blackstock et al., 2013; Hayden, Smiley, Alexander, Kardong-Edgren, & Jeffries, 2014; Kimhi, Cohen, Friger, Hurvitz, & Avraham, 2016; Schlairet & Pollock, 2010; Soccio, 2017; Watson et al., 2012) three were Level 2 quasi-experimental designs (Baillie & Curzio, 2009; Curl, Smith, Chisholm, McGee, & Das, 2016; Meyer, Connors, Hou, & Gajewski, 2011), and the remaining study was a Level 3 observational-analytic design (Giblett, Rathore, & Carruthers, 2017).

4.2. Quality of evidence

The methodological quality of the 10 primary studies assessed using MERSQI was moderate to high (M= 13.5 (75% on a scale of 100)), with a range 9.0-16.5 of a possible 18 points based on the average scores from two raters (Table 3). All studies were rated in the upper range quartiles, ≥50%, with four rated above the 75th percentile (≥13.5 points). Inter-rater agreement on the quality of each primary study was good to very good, with a Kappa measure (k) for each study being greater than 0.7 and agreement in nine studies very good (k >0.8). In addition, a significant Intraclass Correlation Co-efficient representing reliability across all raters and across the 10 studies was large: 0.977 for average measures across raters, using a one-way random effects model (F=44.186, df 9, 190 [CI .951- .993], p= <0.001). Data thus confirmed the overall consistency of ratings and the reliability of this tool in rating methodological quality.

There were, however, variations in scoring of studies using MERSQI that were linked to the variability in research designs. It was noted that higher ratings were achieved if a study recruited random samples, achieved high response rates, and used objective measurements. All of the studies (n=10) provided at least one or more of rates of recruitment, completion

and/or losses to follow up. Recruitment rate was only reported by four studies and ranged from 29-100%. Completion rates were reported by, or able to be calculated for, most studies and ranged from 53-100% (median 88%) and losses to follow up ranged from 0-47% (median 5%).

Eight studies used objective measures rather than self-reported measures, for example, competence observed and rated by a trained observer. The types and quality of measures used varied across studies with some (n=5) reporting the use of internally devised measures particular to the study, while the majority used established measures or a combination of both validated and purposely developed tools. Reporting of the validity and reliability of measures was also inconsistent across studies, with studies that used internally devised measures largely failing to report on either their validation or the reliability. Reliability was reported for just over half the measures used (52%, n=12) and inter-rater reliability and/or intra-class correlation coefficients demonstrated reliability and ranged from 0.72 – 0.99. In studies where established measures were used, the validity and reliability were inconsistently reported i.e. both validity and reliability were reported, or reliability was reported but validity was presumed, or neither were reported. However, the study methodological quality limitations most often included a failure to provide detail of the validity of measurement instruments (content, internal structure, relationship to other variables).

The outcomes assessment in the MERSQI achieved higher scores when behavioural measures were actions with real patients in a clinical context (or substitutes); and highest of all when patient/healthcare outcomes were actual effects on real patients, programs, or society (Reed et al., 2009). No study reported this latter level of outcome.

There were no superiority trials among the six included RCTs, however two studies were identified as non-inferiority studies (Blackstock et al., 2013; Watson et al., 2012) and one study was initially identified as an equivalence trial (Hayden et al., 2014). The study by Hayden et al. initially appeared to be a parallel group trial with some reference to equivalence, however was confirmed as equivalence in design by author S. Kardong-Edgren (personal communication, 6th May 2018).

Two of the RCTs appeared to be sufficiently powered: Hayden et al. (2014) and Watson et al. (2012), two were potentially underpowered based on information given (Blackstock et al., 2013; Schlairet & Pollack, 2010) and two RCTs were samples of convenience (Kimhi et al., 2016; Soccio, 2017). Three of these studies used validated tools for their primary outcome.

Enrolment in RCTs was voluntary and the response rate was reported in only half of the studies. Bias arising from participant sample choice is poorly understood, however. Estimates of effect size are comparable between groups within studies due to appropriate study designs and data analysis. We found variability within studies was generally reported for each group. Both of these estimates will be useful for sample size calculations in future studies. The interpretation of results was appropriate. The two non-inferiority studies (Blackstock et al. (2013) and Watson, et al. (2012) and the equivalence study by Hayden et al. (2014), are generalisable, with strengths in being multi-site studies, their use of validated tools and also well described, appropriate and reproducible interventions.

4.3. Evidence of statistical difference or equivalence for evaluation outcomes

Statistically equivalent levels of performance between the intervention and control groups regarding general nursing competence were notable findings of most studies. Having examined the level and quality of evidence we now turn attention to whether the evidence demonstrates statistically significant differences or equivalence for evaluation outcomes when simulation-based education is substituted for a proportion of pre-licensure learners' clinical placement hours.

In all, 2,370 health care professional students (sample range: n = 48 to n = 847) were recruited into included studies, and 1,972 students (sample range: n = 48 to n = 666) participated; yielding an overall participation rate of 83%. The majority of studies reported nursing research (n = 7) and others described physiotherapy (n = 2), or medicine (n = 1) studies. The studies were from the USA (n = 5), Australia (n = 2), the UK (n = 2) and Israel (n = 1) (see Table 2).

Hayden et al. (2014) conducted a national study in 10 nursing programs in USA with 666 students completing the study. This longitudinal, randomized controlled study replacing clinical hours with simulation in pre-licensure nursing education investigated replacement hours at the levels of 25% and 50% simulation substitution. The 10% simulation cohort was regarded as the traditional education control group. Knowledge, clinical competency, critical thinking and readiness for practice were assessed at end of the undergraduate nursing program and first-time National Council Licensure Examination or NCLEX pass rates were examined between groups. Educational outcomes were found to be equivalent on all

evaluation outcomes when up to 50% of traditional clinical experience in the nursing program was replaced by simulation.

Similarly, Curl et al. (2016) in USA conducted a quasi-experimental two-group study to test whether senior Associate Degree in Nursing students participating in high fidelity simulation and clinical experiences in healthcare settings would attain knowledge and skills equal to students participating exclusively in traditional clinical experiences. Students were recruited (n = 124) and 97 were assessed using valid objective assessments (including faculty observations), after completing 80 hours of simulation (4 hrs per module). Post-test knowledge in the intervention group was significantly higher in the medical-surgical test ($p = 0.05$) and in the exit exam ($p = 0.01$), than for the control group. The findings were that simulation experiences could be used in lieu of 50% of traditional clinical experiences in both block and integrated curricula. The study equated a ratio of one hour of simulation to two hours of clinical time.

A similar experiment in 2008 in the USA by Schlairet and Pollock (2010) utilizing a 2x2 crossover design found that novice nursing students completing the intervention (n = 71) gained statistically equivalent nursing knowledge from two weeks of simulation and two weeks of traditional clinical experience and remained so despite different sequencing. Simulated clinical experience was as effective as traditional clinical placement experience in promoting students' knowledge acquisition in a fundamentals of nursing course.

This pattern of findings of statistically equivalent levels of performance between the intervention and control groups regarding general nursing competence was seen in other nursing studies that were focussed on specific practice areas. Soccio et al. (2017) studied the impact of pre-clinical mental health simulation labs on USA nursing students, where three of 12 clinical weeks were substituted with simulation. They reported that students experiencing simulation as a replacement for 25% of traditional clinical hours had equivalent mental health knowledge and self-confidence to those who did not receive the simulation, recommending that simulation could be used as a replacement for 25% of traditional clinical hours in mental health nursing.

Other studies investigated slight variants of outcome measures, but also reported the benefits of simulation substitution in nursing. Kimhi et al. (2016) examined the impact of simulation and clinical experience on self-efficacy in 56 first-year nursing students who completed their study in Israel, also suggesting that self-efficacy can be regarded as a competence measure. In

a double crossover design, students' self-efficacy/ self-confidence increased significantly in both groups after simulation.

Meyer et al. (2011) conducted an observational study of USA junior nursing students (n = 116) who completed simulation and clinical experience in a paediatric course, finding no difference in final facilitators' ratings of students' placement performance between groups. They reported that time in simulation enhanced clinical performance, as simulation students achieved higher scores more quickly than those without simulation and they maintained high performance levels. An earlier study of 276 UK pre-licensure nursing students (Baillie, & Curzio, 2009) sought student and facilitator views on substitution of half days of simulation within five days of clinical placement experience. They found that for the 52.8% of students who completed the second questionnaire (n = 141), while there was no difference in evaluation feedback between the intervention and control groups, the intervention group were satisfied with the simulation program and nearly all (93%) felt it helped their skills ability and 89% felt it increased their confidence for placement.

Four studies of physiotherapy students have reported similar positive outcomes for simulation substitution in randomized trials. In Australia, Watson et al. (2012) developed a simulated learning programme as a replica for clinical education in musculoskeletal practice. Two single-blind, multicentre RCTs were conducted using different sequencing of the simulation component. There was no significant difference within trial groups in observed physiotherapy competence, indicating that simulation can in part replace clinical time with real patients without compromising students' attainment of the professional competencies required to practise.

Also in physiotherapy and in Australia, Blackstock et al. (2013) substituted specifically focused clinical education in two randomized controlled trials, using different sequencing. They reported there was no significant difference in observed competency of students within both trials, between simulation and control groups; concluding that a simulation learning experience could replace clinical time in cardio-respiratory physiotherapy practice to the extent of 25%.

Finally, Giblett et al. (2017) reported significantly higher student satisfaction, self-confidence and self-evaluation of knowledge (all $p<0.001$) following simulated patient pathway surgical experiences in first year medical students.

4.4. Levels of educational evaluation

The levels of educational evaluation outcomes in the studies were assessed to determine the translational impact of the findings (Table 2). Our assessment indicates that various combinations of educational assessment between Level 1 and Level 3 were evaluated. Seven studies assessed students' *Reaction*, seven studies assessed *Learning* in various ways and five studies assessed *Behaviour/transfer*. The studies that evaluated *Learning* did so via an objective measure such as a knowledge test or written examination with some examinations pertaining to national entry-tests of competency standards. However, some studies used students' self-reports of self-efficacy or self-confidence and we allowed self-efficacy as evidence of learning if a validated scale was used. Other studies assessed *Behaviour/transfer* through subsequent administration of an objective structured clinical examination by trained experts using a standardized checklist, or else observational ratings made during or at the end of a clinical practicum.

4.5. Evidence for proportional substitution, replacement ratios and duration and evidence translation into pre-licensure professional education standards

The evidence seems clear that simulation is beneficial and can provide a proportion of clinical experience hours in pre-licensure health professional education in nursing, physiotherapy and medicine. For the studies in which direct substitution occurred the proportional substitution seemed to be arbitrarily determined in trial design and ranged from approximately 5-50%. Australian physiotherapy research recommended that 25% of clinical hours be substituted, while international nursing research studies recommended 25% and up to 50% be replaced. One study (Curl et al., 2016) reported a substitution ratio 1:2 for simulation to clinical placement hours, the remaining studies assumed 1:1 ratio. There is no clear finding from this review as to the application of simulation hours equivalency to clinical hours. The duration of the simulation replacement also varied from 21 hours to a program-wide approach of two years. Likewise, there is no clear conclusion as to the optimal duration of simulation substitution.

Further evidence of simulation substitution may be gained by examining translational education outcomes and the training policy documents of relevant professions, and whether they mandate, or specify, such training. Several professions in the jurisdictions relevant to the studies included in this review have mandated proportional substitution of clinical hours with simulation.

In nursing, global standards identify that schools should have access to clinical simulation laboratories, and that programmes demonstrate the use of recognized approaches to teaching and learning including clinical simulation. Relevant to this, the nursing studies in this review were conducted in the UK, USA and Israel. In the UK, the Nursing and Midwifery Council (NMC) prescribes standards for pre-licensure nursing programs. In their 2009 policy, 300 hours of the required 2,300 hours (13%) of clinical practice can be replaced with simulation practice. In a review undertaken from June 2017, the NMC proposed that simulation could be used for up to half of the 2,300 practice hours required to register as a nurse. Recently revised standards simply state that educational institutions must 'ensure technology and simulation-based learning opportunities are used effectively and proportionately to support learning and assessment' (Nursing and Midwifery Council, 2018). There is no further guidance provided as to the proportion or maximum permitted levels of simulation.

In the USA, an expert panel of the National Council for State Boards of Nursing developed National Simulation Guidelines for Prelicensure Nursing Programs (Alexander et al., 2015) to guide Boards of Nursing in evaluation of readiness of programs to substitute simulation for clinical experience and in establishing evidence based simulation programs. In part, the guidelines have translated evidence from the NCSBN National Simulation Study (Hayden et al., 2014), advising that the amount of simulation should be increased slowly and steadily with the acquisition of expertise in simulation. Although the NCSBN study tested proportional replacement in programs with a minimum 600 hours clinical, no universal recommendation was made regarding substitution of simulation hours. Primarily this occurred because there is no evidence for programs with less than 600 hours and quality of the experience is deemed the most important. The recommendation is that the overall number of program hours, pass rates of students, clinical site availability, turnover of faculty/program director and complaints from students should be considerations in the amount of simulation that can be substituted for traditional clinical hours.

We were unable to locate information regarding program accreditation standards in Israel. Publicly available information from the Ministry of Health simply identifies the proportion of theory to clinical credits in programs leading to registration (State of Israel, 2018).

Two of the studies included in this review were in the discipline of physiotherapy and both were conducted in Australia. The Australian Physiotherapy Council regulates course accreditation and the accreditation standards do not prescribe the amount of simulation which may be included in programmes. Rather, they refer to the quality and quantity of clinical

education being sufficient to produce a competent graduate and that learning and teaching methods are intentionally designed to ensure that the required learning outcomes are achieved (Australian Physiotherapy Council, 2017).

The final study included in this review related to medicine in the UK. The General Medical Council identifies that learners must have access to technology-enhanced and simulation-based learning opportunities within their training programme as required by their curriculum. Experiential learning can be undertaken in clinical settings, both real and simulated, and should increase in complexity in line with the curriculum (General Medical Council, 2015). There are no specific requirements for the amount of simulation or for clinical practice hours overall.

5. DISCUSSION

The published literature reporting primary studies of the substitution of simulation based education for clinical placement in health care disciplines is sparse. However, of the 10 primary studies identified, the majority provided experimental and quasi-experimental evidence, ranging in methodological quality from moderate to high (M=13.3 of 18 points) with high inter-rater reliability. In comparison, other reviews of medical education literature have reported lower mean MERSQI scores of 9.9 (DeCoste-Lopez, Madhok, & Harman, 2015) over 38 studies.

Of the six RCTs included in this review, none were double-blinded as the pragmatic considerations of educational research are likely to have restricted the ability to blind participants to their assigned groups (Sullivan, 2011). However, of these trials Blackstock et al. (2013) and Watson et al. (2012) used single-blinding in order to remove potential bias in ascertainment of outcomes. Although blinding of participants may not be feasible, future simulation education research should, where possible, maximise blinding of data collectors, outcome assessors and data analysts.

None of the RCTs were superiority trials, rather they set out to test whether simulation was equivalent to, or no worse than, clinical practice, or whether there was a statistically significantly difference in outcomes. However, there are questions about superiority and equivalence data. The interpretation of superiority trials as noninferiority trials and *vice versa* is complicated and is best approached by expressing the results as a confidence interval for the difference between the intervention and control groups (Committee for Proprietary

Medicinal Products, 2001). Alternatively, an effect size based on the standardized mean difference between groups with an outcome of ≥ 0.2 standard deviations is seen as clinically significant (Polit, & Tatano-Beck, 2012, p. 478). Unless a primary outcome measure shows a reasonable clinically significant difference between these two groups, a superiority trial does not appear to be achievable. The tools used to assess the primary outcome measures in the reviewed studies showed small non-clinically significant differences between simulation and clinical placement groups and this was argued as being 'equivalent' or 'not different'. Many of the studies appeared to use standard assessment tools for their discipline with students generally fulfilling the requirements.

These somewhat conservative approaches used in trials to date are surprising given the dynamic growth of simulation technologies, scientific studies, peer reviewed dissemination, simulation learned societies and interest groups, educational resources, curricula and policies; all evidence that 'clinical simulation science is thus past its early developmental stages' (Sevdalis, Nestel, Kardong-Edgren, & Gaba, 2016) and that it has 'matured over the past 40 years on substantive and methodological grounds'(McGaghie, Issenberg, Petrusa, & Scalese, 2010). A demonstration of superior outcomes could justify the investment required to overcome the identified physical and human infrastructure barriers (Bogossian et al., 2017) to providing sustainable financial support and dedicated simulation specialists to advance simulation in health care education (Qayumi et al., 2014).

In this review, the evidence from the range of studies in three health care professions suggests that there is a statistically equivalent level of performance when simulation-based education is substituted for a proportion of clinical placement hours in pre-licensure programs. It is important to note that there is no evidence in any of the studies of any negative impact on learner performance for the almost 2000 participating students, regardless of health profession, level of program of study, level of student seniority or simulation exposure when substituting simulation for a proportion of clinical placement.

5.1. Practicalities for design of substitution

The timing and duration of substitution of simulation for clinical placements varied across included studies, as did the level of detail given in reporting the instructional design of simulation interventions - for example, critical faculty training, simulation modalities,

debriefing and feedback methods. Furthermore, the majority of studies minimally reported on, or demonstrated compliance with, standards for best practice in simulation. These shortcomings preclude cross study comparisons and raise issues of potential confounding within studies. While it is likely that publication word limit restrictions will continue to restrict full reporting of all elements of simulation design this could be overcome by publishing detailed study protocols. Alternately reporting could be strengthened in line with recent recommendations by Cheng et al., (2016), who proposed simulation based research extensions to the CONSORT and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statements including, for example, clear description of simulation specific exposures, potential confounders and effect modifiers.

Most of the studies evaluated specific curricula components being taught over short placement equivalent periods, with the exception of Hayden et al. (2014) where a 'whole of program' simulation replacement was evaluated over a two-year period. Only one study reported an explicit ratio of substitution where one hour of simulation was equated to two hours of clinical placement (Curl et al., 2016), and the national survey of academics in USA (Breymier et al., 2015) found that the most frequently utilized ratio was 1: 1 hours. While there is currently no evidence to support either approach it would seem reasonable to propose that simulation (which can be both controlled and time compressed), would attract more than parity in terms of clinical practice hours. This is an area worthy of further investigation.

5.2. Design limitations

Although the within study evidence for the substitution of simulation for a proportion of clinical placement is strong, in our review of included studies potential limitations are evident. Six of the studies failed to report the rates of recruitment from the eligible population to participant sample. While recruitment rates might be important to inform judgements about the potential for coercion or selection bias within a study, they also can provide insight into the appeal of simulation interventions to pre-licensure students and inform planning of future studies. Of those studies for which recruitment rates were reported (or could be calculated from the information provided), the rates of recruitment varied from 29-100%.

Similarly, rates of completion of the intended intervention or control exposures and losses to follow up were inconsistently reported. While we were able to calculate completion rates and

losses to follow up for all of the included studies, we encourage future research to report these rates as a matter of course. Completion rates can assist with interpretation of findings as non-compliance with a protocol can result in over- or underestimating the effects and reporting of rates of losses to follow up also provides a valuable interpretive perspective.

While most of the studies used established measures to estimate outcomes of interest, the validity of these measures was also inconsistently reported. When studies use self-developed measures, reporting on the *accuracy* and *credibility* of measures by addressing relevant components of validity i.e. face, content, consequential and predictive validity (Cooper & Bogossian, 2018) becomes even more important, not only to inform interpretations but also to advance measurements. Perhaps even more important in simulation research, where there are many sources of potential bias in observational ratings and a relative lack of development of direct measures of performance (McGaghie et al., 2010), is the demonstration of reliability or the *stability* and *consistency* of measures (Cooper & Bogossian, 2018). Outcome measurement has been recognised as one of the greatest challenges now facing research in the field (McGaghie et al., 2010).

We assert that an even larger challenge lies in demonstrating the translational impact of simulation education research. In this review, we adopted Kirkpatrick's levels of evaluation and found that despite the underlying assumption that simulation education techniques enhance education processes and outcomes, (which in turn promote patient safety), none of the included studies measured Level 4 outcomes (improvements in patient outcomes and/or organisational change) (Praslova, 2010). While the measurement of higher-level outcomes might be ambitious in simulation education research (particularly with pre-licensure students), there is an ethical imperative to demonstrate this translation. McGaghie et al. (2010) illustrate this translation from T1 where the participant shows improvement in knowledge and skills in the simulated setting to T2 where improvements in knowledge and skills are used in patient care settings, to T3 where improvement is measured in health of individuals and populations. Simulation education research measures need to be extended to capture improved patient care practice and to inform our policy.

5.3. Lagging translational policies

The evidence from the included studies does not appear to have informed or the knowledge translated in the accreditation standards for pre-licensure programs in health care disciplines and in the jurisdictions represented. In nursing in the UK, the standards relating to simulation have recently become less prescriptive. While in US guidelines, a distinction is made between programs with greater or less than 600 hours clinical, there are other considerations which need to guide the substitution of clinical with simulated hours. For the remaining professions and jurisdictions there seems to have been no explicit translation of evidence into program accreditation requirements.

Given the cost of conducting this type of research, it would be reasonable to anticipate translational impact. However, we propose that there are three major reasons for the lack of evidence-informed policy. Firstly, there may be reticence to change standards of program accreditation in the absence of national and profession-specific evidence. This reticence might reflect a philosophical stance that nothing can replace learning in the clinical setting or - concern for the potential for evidence from other disciplines and jurisdictions to be generalised and imposed, or an accreditation focus on competency attainment rather than prescribed educational exposures. Secondly, in some practice-based disciplines there are concerns and issues related to the adequacy of students' clinical placement hours with wide variation globally in minimum required hours. For example, in nursing, in some countries there are no specified hours compared with 2,800 hours in South Africa (Miller & Cooper, 2016); moreover, there is an absence of evidence to inform these requirements. It is not unreasonable to ask whether we ought to be considering the development of a sound evidence base for clinical practice hours before exploring replacement with simulation.

Thirdly, there may be tacit recognition of the resource implications relating to transitioning components of clinical education away from the clinical setting (Bogossian et al., 2017) and moving costs from health services to tertiary providers. The adequacy of existing simulation resources and their access have been clearly identified as concerns. This is particularly salient if universal standards are proposed relating to replacement of clinical practice with simulation, rather than standards which recognise contextual differences and are conditional based on resourcing and access. Recent surveys have shown that professions may have adequate access to equipment-type simulation resources (task trainers, programmable manikins, simulated patients, equipment) but lack the faculty resources to use them in a

standardized way- in nursing (in Australia, New Zealand) (Bogossian et al., 2017), in paramedicine (USA) (McKenna et al., 2015), in speech pathology (Australia) (MacBean, Theodoros, Davidson, & Hill, 2013), and in radiology (Patel, 2017). Barriers experienced included staff time and lack of training and resource development. Given the profound resource implications it is imperative that future research should include measures of cost-effectiveness in evaluation of simulation substitution for clinical practice.

5.4. Enabling highest levels of evidence in educational research

To our knowledge this review is the first to span the health care professional literature and in doing so to synthesise the highest levels and quality of evidence of the replacement of clinical practice with simulation. The 'gold standard' in scientific research is generally based on the randomized controlled trial. While there is some debate about the application of this standard in educational research (Norman, 2010; Sullivan, 2011) we support the assertion that RCTs have a role to play in education research when examining relatively standardised interventions such as clinical simulation, when they reflect the nature of the research questions asked and are amenable to experimentation (Norman, 2010). However, to date the use of RCTs in simulation education research is relatively sparse thus we opted to include other levels of evidence in this systematic review to more fully answer the research questions. While it might be argued that in a review of gold standard evidence these studies should have been excluded, we took the view that in education research alternative approaches need to be included, there are research questions and contexts in which randomisation and experimentation are inappropriate, and that these are not inferior (Sullivan, 2011). The quality appraisal revealed research scholarship that aligned with best practices in interventional educational research and quality ratings were consistent with those in the medical education literature. However, the quality and rigour of the simulation interventions in each study were not able to be assessed, thus assertions about the quality of studies are limited to reported components.

The nature of the research questions in this study restricted pragmatic systematic searches of databases, and it is inevitable that some studies such as non-English studies may have been missed. Systematic reviews of literature can be subject to reporting bias. With this awareness, we applied best practice guidelines (PRISMA) to develop the review and its reporting. Although we found published primary studies for the disciplines of nursing, physiotherapy,

and medicine, in Australia alone we are aware of primary studies currently underway in other disciplines e.g. occupational therapy (Imms, et al., 2017), speech pathology (HealthWorkforce Australia and The Speech Pathology Association of Australia, 2014) and secondary studies which did not meet the inclusion criteria for this review (See Supplementary Table A). All these studies will no doubt contribute to the body of evidence in due course.

6. CONCLUSION

Gold standard evidence from educational research provides conditional support for substitution of clinical practice with simulation-based education across three health care professions in four countries. There is no evidence of negative outcomes, rather outcomes are similar when simulation replaces clinical practice. However, the studies included in this review are notable in their heterogeneity, both in the proportion, ratio and duration of the simulation programs offered and in the evaluation outcomes incorporated. Thus, it is not possible to conclude how much clinical practice can be replaced with simulation at this point in time. Future studies, therefore, should incorporate standardized simulation curricula and consistent evaluation outcomes.

The field of simulation education research is rapidly advancing. Future research should: expand the representation to other health care professions, explore experimental approaches, employ blinding where feasible, report recruitment, compliance and loss to follow up rates, demonstrate validity and reliability of measures, extend measures to include translational outcomes and cost effectiveness and enhance the quality of reporting in simulation-based research. These recommendations will lead not only to a stronger evidence base, but will also bridge the gaps between education research, evidence informed healthcare professional curriculum renewal, program accreditation policy development and standard setting and, ultimately, better patient care.

6.1. Relevance to clinical practice

The quality of health care clinical practice is in part determined by the clinical educational preparation of health professionals. The current evidence for clinical educational preparation

does not appear to be translated into program accreditation standards governing clinical practice experience for pre-licensure programs in the relevant jurisdictions. Overall, there is a need for a stronger evidence base to inform future curricula and policy development to strengthen clinical practice in health.

References

Aiken, L. H., Clarke, S. P., Cheung, R. B., Sloane, D. M., & Silber, J. H. (2003). Educational levels of hospital nurses and surgical patient mortality. *JAMA, 290*(12), 1617-1623.

Alexander, M., Durham, C. F., Hooper, J. I., Jeffries, P. R., Goldman, N., Kesten, K. S., . . . Tillman, C. (2015). NCSBN simulation guidelines for prelicensure nursing programs. *Journal of Nursing Regulation, 6*(3), 39-42.

Australian Physiotherapy Council. (2017). *Accreditation standard for entry-level physiotherapy practitioner programs.* Retrieved from *https://physiocouncil.com.au/wp-content/uploads/2017/10/ACCREDITATION-STANDARD-V1.1-13112017.pdf*.

Baillie, L., Curzio, J. (2009). Students' and facilitators' perceptions of simulation in practice learning. *Nurse Education in Practice, 9*, 297–306.

Blackstock, F., Watson, K. M., Morris, N. R., Jones, A., Wright, A., McMeeken, J. M., … Watson, G. (2013). Simulation can contribute a part of cardiorespiratory physiotherapy clinical education: two randomized trials. *Simulation in Healthcare, 8*, 32-42.

Bogossian, F., Cooper, S., Kelly, M., Levett-Jones, T., McKenna, L., Slark, J., & Seaton, P. (2017). Best practice in clinical simulation education − are we there yet? A cross-sectional survey of simulation in Australian and New Zealand pre-registration nursing education *Collegian, 25* (3), 327-334. http://dx.doi.org/10.1016/j.colegn.2017.1009.1003.

Breymier, T. L., Rutherford-Hemming, T., Horsley, T. L., Atz, T., Smith, L. G., Badowski, D., & Connor, K. (2015). Substitution of clinical experience with simulation in prelicensure nursing programs: A national survey in the United States. *Clinical Simulation in Nursing, 11*(11), 472-478. doi: http://dx.doi.org/10.1016/j.ecns.2015.09.004

Cant, R., & Cooper, S.J,. (2017). Use of simulation-based learning in undergraduate nursing education: An umbrella systematic review. *Nurse Education Today, 49*, 63-71. Doi: http://dx.doi.org/10.1016/j.nedt.2016.1011.1015.

Cantrell, M. A., Franklin, A., Leighton, K., & Carlson, A. (2017). The evidence in simulation-based learning experiences in nursing education and practice: An umbrella review. *Clinical Simulation in Nursing, 13*(12), 634-667. doi:10.1016/j.ecns.2017.08.004

Cheng, A., Kessler, D., Mackinnon, R., Chang, T. P., Nadkarni, V. M., Hunt, E. A., . . . Pusic, M. (2016). Reporting guidelines for health care simulation research: extensions to the CONSORT and STROBE statements. *Advances in Simulation, 1*(1), 25.

Committee for Proprietary Medicinal Products. (2001). Points to consider on switching between superiority and non-inferiority. *British Journal of Clinical Pharmacology, 52*, 223–228. doi: 210.1046/j.0306-5251.2001.01397-01393.x.

Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Academic Medicine, 90*(8), 1067-1076. doi:10.1097/acm.0000000000000786

Cooper, S., & Bogossian, F. (2018). Enhancing the rigor of virtual simulation. In R. M. Gordon, D McConigle (Eds.), *Virtual simulation in nursing education*. NY: Springer Publishing Company.

Coyle, G., Mandated hours in NZ allied health degrees. *Focus on Health Professional Education*, 2007. **9**(1): p. 67.

Curl, E. D., Smith, S., Chisholm, L. A., McGee, L. A., & Das, K. (2016). Effectiveness of integrated simulation and clinical experiences compared to traditional clinical experiences for nursing students. *Nursing Education Perspectives*, *37*(2), 72-77.

DeCoste-Lopez, J., Madhok, J, & Harman, S. (2015). Curricular Innovations for medical students in palliative and end-of-life care: a systematic review and assessment of study quality. *Journal of Palliative Medicine, 18*, 338-349. https://doi.org/310.1089/jpm.2014.0270.

Eldridge, S., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G.A. (2016). CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *Pilot and Feasibility Studies, 2*, 64. https://doi.org/10.1186/s40814-40016-40105-40818

General Medical Council. (2015). *Promoting excellence: Standards for medical education and training.* Retrieved from *https://www.gmc-uk.org/-/media/documents/promoting-excellence-standards-for-medical-education-and-training-0715_pdf-61939165.pdf*

Giblett, N., Rathore, R., & Carruthers, D. (2017). Simulating the surgical patient pathway for undergraduates. *Journal of Surgical Education, 74*, 271-276. doi: 210.1016/j.jsurg.2016.1010.1003.

Hayden, J., Smiley, R.A., Alexander, M., Kardong-Edgren, S., & Jeffries, P. R. (2014). The NCSBN national simulation study: A longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. *Journal of Nursing Regulation, 5*, C1-S64.

HealthWorkforce Australia and The Speech Pathology Association of Australia. (2014). Project plan Phase 2- Embedding simulation in clinical training in speech pathology project.

Hsieh, M-J., Bhanji, F., Chiang, W-C., Yang, C. W., Chien, K. L., & Ma, M. H. M. (2016). Comparing the effect of self-instruction with that of traditional instruction in basic life support courses. A systematic review. *Resuscitation, 108*(Supplement C), 8-19. https://doi.org/10.1016/j.resuscitation.2016.08.021

Imms, C., Chu, E. M., Guinea, S., Sheppard, L, Froude, E, Carter, R, … Nicola-Richmond, K. (2017). Effectiveness and cost-effectiveness of embedded simulation in occupational therapy clinical practice education: Study protocol for a randomised controlled trial. *Trials, 18*(1), 345.

INACSL. (2016). INACSL Standards of best practice: simulation: Debriefing. *Clinical Simulation in Nursing, 12*, S21-S25. doi:10.1016/j.ecns.2016.09.008

Joanna Briggs Institute. (2014). *New JBI levels of evidence*. Adelaide: Joanna Briggs Institute.

Kaptchuk, T. J. (2001). The double-blind, randomized, placebo controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology, 54*, 541-549.

Kimhi, E., Reishtein, J. L., Cohen, M., Friger, M., Hurvitz, N., & Avraham, R. (2016). Impact of simulation and clinical experience on self-efficacy in nursing students: intervention study. *Nurse Educator*, *41*(1), E1-E4.

Levett-Jones, T., Hoffman, K., Dempsey, J., Jeong, S. Y. S., Noble, D., Norton, C. A., ... & Hickey, N. (2010). The 'five rights' of clinical reasoning: An educational model to enhance nursing students' ability to identify and manage clinically 'at risk'patients. *Nurse Education Today*, *30*(6), 515-520. doi.org/510.1016/j.nedt.2009.1010.1020

MacBean, N., Theodoros, D., Davidson, B., & Hill, A. E. (2013). Simulated learning environments in speech-language pathology: An Australian response. *International Journal of Speech-Language Pathology, 15*(3), 345-357. doi:10.3109/17549507.2013.779024

Malina, D. (2016). Assessing the Gold Standard – Lessons from the history of RCTs. *New England Journal of Medicine, 374*(22), 2175-2181.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation- based medical education research: 2003–2009. *Medical Education, 44*(1), 50-63.

McKenna, K. D., Carhart, E., Bercher, D., Spain, A., Todaro, J., & Freel, J. (2015). Simulation use in paramedic education research (SUPER): A descriptive study. *Prehospital Emergency Care, 19*(3), 432-440. doi:10.3109/10903127.2014.995845

Meyer, M. N., Connors, H., Hou, Q., & Gajewski, B. (2011). The effect of simulation on clinical performance: A junior nursing student clinical comparison study. *Simulation in Healthcare, 6*(5), 269-277. doi: 210.1097/SIH.1090b1013e318223a318048

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine, 6*, e1000097.

Motola, I., Devine, L., Chung, H., Sullivan, J., & Issenberg, B. (2013). Simulation in healthcare education: A best evidence practical guide. AMEE Guide No. 82. *Medical Teacher, 35*, e1511–e1530.

Norman, G. (2010). Is experimental research passe? *Advances in Health Science Education, 15*, 297-301.

Nursing and Midwifery Council. (2018). *Realising professionalism: Part 3: Standards for pre-registration nursing programmes.* Retrieved from *https://www.nmc.org.uk/globalassets/sitedocuments/education-standards/programme-standards-nursing.pdf*

Pai, H. (2016). An integrated model for the effects of self-reflection and clinical experiential learning on clinical nursing performance in nursing students: A longitudinal study. *Nurse Education Today, 45*, 156-162. doi: 110.1016/j.nedt.2016.1007.1011

Patel, R., & Dennick, R. (2017). Simulation based teaching in interventional radiology training: is it effective? *Clinical Radiology, 72*(266), e267-e214. doi: 210.1016/j.crad.2016.1010.1014

Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., Evans, S. J., & Consort Group. (2006). Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA, 295*(10), 1152-1160.

Polit, D., & Tatano-Beck, C. (2012). *Nursing research: Generating and assessing evidence for nursing practice.* Sydney: Wolters Klewer.

Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational Assessment, Evaluation and Accountability, 22*(3), 215-225.

Qayumi, K., Pachev, G., Zheng, B., Ziv, A., Koval, V., Badiei, S., & Cheng, A. (2014). Status of simulation in health care education: an international survey. *Advances in medical education and practice, 5*, 457.

Reed, D. A., Beckman, T. J., & Wright, S. M. (2009). An assessment of the methodologic quality of medical education research studies published in The American Journal of Surgery. *The American Journal of Surgery, 198*(3), 442-444. doi:https://doi.org/10.1016/j.amjsurg.2009.01.024

Roh, Y. S., Lim, E. J., & Issenberg, B. (2014). Effects of an integrated simulation-based resuscitation skills training with clinical practicum on mastery learning and self-efficacy in nursing students. *Collegian, 23*(1), 53-59. doi:10.1016/j.colegn.2014.10.002

Schlairet, M. C., & Pollock, J. (2010). Equivalence testing of traditional and simulated clinical experiences: Undergraduate nursing students' knowledge acquisition. *Journal of Nursing Education, 49*(1), 43-47.

Sevdalis, N., Nestel, D., Kardong-Edgren, S., & Gaba, D. M. (2016). A joint leap into a future of high-quality simulation research—standardizing the reporting of simulation science. *Advances in Simulation,* **1**:24. https://doi.org/10.1186/s41077-016-0026-x

Soccio, D. A. (2017). Effectiveness of mental health simulation in replacing traditional clinical hours in baccalaureate nursing education. *Journal of Psychosocial Nursing & Mental Health Services, 55,* 36-43. http://dx.doi.org.ezproxy.lib.monash.edu.au/10.3928/02793695-20170905-02793603.

State of Israel, Minstry of Health Studies. (2018). *Ministry of Health Studies Program for a Registered Nurse Degree.* Retrieved from *https://www.health.gov.il/English/Services/MedicalAndHealthProfessions/nursing/nursingStudies/Pages/programs.aspx*

Sullivan, G. M. (2011). Getting off the "gold standard": randomized controlled trials and education research. *Journal of Graduate Medical Education, 3*(3), 285-289.

Veltri, L. M., Rowe, J. M., Bell, K. J., Arwood, E. L., & Kindler, L. L. (2014). The Maternal- Newborn Assessment Study: Can simulation replicate the clinical learning experience in undergraduate nursing education? *Journal of Obstetric, Gynecologic & Neonatal Nursing, 43,* S84. doi: https://doi.org/10.1111/1552-6909.12442

Watson, K., Wright, A., Morris, N., McMeeken, J., Rivett, D., Blackstock, F., ... & Peterson, R. (2012). Can simulation replace part of clinical time? Two parallel randomised controlled trials. *Medical Education*, *46*(7), 657-667.

World Health Organisation. (2016). *Global strategy on human resources for health: workforce 2030.* Retrieved from *http://who.int/hrh/resources/globstrathrh-2030/en/.*

Table 1 Inclusion and exclusion criteria

Inclusion criterion	Exclusion criterion
• Pre-licensure students in all health-related disciplines. • Primary experimental/quasi-experimental controlled studies that report the effectiveness of simulation-based education as a **substitute** for traditional clinical placement. • Publication year – any (includes up to 1 February, 2018). • Studies published in English.	• Studies of single clinical proficiencies • Interprofessional simulation • Postgraduate training • Postgraduate team training

Table 2 Characteristics of included studies of simulation substitution for clinical practice in healthcare disciplines

Study/origin	Design	Discipline/ Sample	Simulation[1] Intervention (INT) DURATION SUBSTITUTION PROPORTION (%) RATIO (sim: clinical)	Control (CON)	Rates of Recruitment (RR) Completion (CR) & losses to follow up (LTFU)	Evaluation/ Measures used	Validity (V) and Reliability (R) of measures	Evaluation Outcomes/ Kirkpatrick's Levels/ Findings
Level 1: Experimental Designs								
Blackstock et al. (2013). (Australia)	Two independent single-blind multi-institutional RCTs conducted using a **non-inferiority** design. Students were stratified on academic grade and randomly allocated.	Physiotherapy students, entry level, acute care cardio-respiratory clinical placements. RCT 1 (n=176) RCT 2 (n=173)	RCT 1: 1 week in simulation before 3 weeks of clinical immersion(n=88) RCT 2: 2 weeks of interspersed simulation (equivalent total 1 week) during 4 weeks clinical immersion (n=88) **DURATION 4 wk.** **SUBSTITUTION 25%.**	Clinical immersion 4 weeks placement RCT1 (n=88) RCT2 (n=85)	RR 29.0% (349/1200) CR – 91.1% (318/349) LTFU Overall: 6.3% (22/349) RCT 1: 6.8% (12/176) RCT 2: 5.8% (10/173)	Clinical Examinations x 2 using the Assessment of Physiotherapy Practice (APP) measure. Secondary outcomes: student perception of experience (13-item scale on confidence and clinical educator and patients' rating of student performance.	APP measure: V- established during scale development, & noted to discriminate 4 competency levels[2] R-high inter-examiner reliability (ICC = 0.92, 95% CI= 0.84-0.96) during development[2] Student confidence scale V – not reported R - Cronbach alpha 0.72- 0.90 in current study.	Level 1, Level 2, Level 3. No significant differences in competency between simulation and control groups in either RCT. RCT 1: Mean APP score for 2 clinical exams was 2.56 (SE 0.05) for SLE group and 2.61 (SE 0.05) for controls (CI of mean difference: -0.09 to 0.17), indicating no difference as upper bound of CI was less than the margin (0.4) (and no difference on any of 7 standards (all: p=>0.05). RCT 2: Mean APP score was 3.02 (SE: 0.05) for the SLE and 2.80 (SE: 0.05) for controls (CI of mean difference: -0.36 to--0.09, upper bound of the CI < 0.4) indicating no difference.

								Students in the interspersed group (RCT 2) achieved a higher score in 5 of 7 APP standards (P <0.05). Clinical educators and patients reported comparability between groups.
Hayden et al. (2014) (USA)	Longitudinal, randomized, unblinded, controlled **equivalence trial** *(comparing maximum effect size using group means)* replacing clinical hours with simulation in pre-licensure nursing education, studied over 24 months.	Nursing students in 10 nursing programs across the United States (n=847)	Two simulation study interventions Group 1-25% n =293 Group 2-50% n= 286 formed from 7 core pre-licensure courses from new students with a two-year window to graduation. Proportion of required clinical hours were spent in the simulation laboratory plus usual course participation and proportional clinical experience. **DURATION 2 yr.** **SUBSTITUTION 25% & 50%.**	Traditional nursing course content and placement n=268 (simulation limited to 10% of course)	RR – not reported CR - 78.6% (n=666/847) Withdrew or Withdrawn LTFU - Overall: 21.4% (181/847), INT 25%: 19.5% (57/293), INT 50%: 25.9% (74/286), CON: 18.7% (50/268) Course Failure	12 stepwise measurement components (instructors =3, students=5, new graduate nurses =3, managers =1): Instructors: Creighton Competency Evaluation Instrument (CCEI)[3] New Graduate Nurse Performance Survey (NGNPS) [4] Global Assessment of Clinical Competency and Readiness for Practice (GACCRP) (instructor feedback); Critical Thinking Diagnostic tool[5] Students: Knowledge at course end tested by ATI RN	CCEI: V – 70% or better for 20 of 23 items R – Cronbach's alpha 0.974 – 0.979 NGNPS: V – not reported R – Cronbach's alpha =0.972; split-half reliability 0.916. GACCRP: V –reported as not established R – intra-rater reliability .80 during development, 0.81 agreement ATI Content Mastery Series®	Level 1, Level 3. There were no statistically significant differences in knowledge (p=0.478), clinical competency (p=0.688), critical thinking and readiness for practice (NCLEX) (p=0.737) for students undertaking traditional placements versus students substituting 25% and 50% of clinical placement time with simulation.

					LTFU – INT 25%: 7.5% (22/293), INT 50%: 6.6% (19/286) CON: 9.3% (25/293)	Comprehensive Predictor®[6] Specialty knowledge tested using ATI Content Mastery Series®[7] (CMS) exams for fundamentals of nursing, Adult Medical-Surgical Nursing, Maternal-Newborn, Nursing Care of Children, Mental Health, and Community Health. First-time NCLEX[7] pass rates. Longitudinal follow-up surveys were conducted of graduates to 6 months. Clinical instructor feedback and a Manager survey were conducted.	(CMS) exam: V – not reported R –Cronbach's alpha in current survey 0.976. ATI RN Comprehensive Predictor® V – not reported R – not reported Critical Thinking Diagnostic© V – not reported R – Cronbach's alpha.976 during development NCLEX: V – not reported R – not reported	
Kimhi et al, 2016 (Israel)	Randomized double crossover design to investigate student outcomes	Nursing Students, first year, second semester (n=67)	Manikin-based simulations program of 3 days (18 hrs) delivered either before or following clinical experience of 5 days for all	Each participant acted as their own control. Clinical experience	RR – 95.7% (67/70) CR – 83.6% (56/67) LTFU - 16.4%	Self-confidence/ Self-efficacy for Nursing Process Scale (SSNPS) Short Form[8] of 7 items measuring students' self-	SSNPS (Short form) V – full 21-item scale had acceptable content validity during development; short	Level 1, Level 2: Students' self-confidence/self-efficacy at time 2 was significantly higher than at baseline (t = -9.02, P < .01; effect size =

	(self-confidence/ self-efficacy for nursing process) after simulation compared with clinical experience.		students. **DURATION 8 days.** **SUBSTITUTION 37.5%.**	of five days was experienced by all students.	(11/67).	confidence and efficacy for nursing skills was used to measure 'capability' at 3 time points.	form V not known. R - Cronbach's alpha .80 in a pilot study of 40 1st-year nursing students. (In present study, alpha was also .80).	0.54). The difference in students' self-confidence/self-efficacy following the clinical scenario (t=-3.37, p<0.01, effect size = 0.39). Order of experiments not significant.
Schlairet & Pollack 2010 (USA)	A 2×2 crossover design, random allocation, unblinded and **equivalence testing** to explore effects of simulated clinical experiences on undergraduate students' knowledge acquisition in a fundamentals of nursing course.	Nursing students, novice (n=74)	Students participated in 2-week laboratory-based simulated clinical experiences with HFS and 2 weeks of traditional clinical experiences. (n=74) **DURATION 4 wk.** **SUBSTITUTION 50%.**	Each participant acted as their own control. All students received the intervention in a crossover design.	RR not reported CR - 95.9% (71/74) LTFU - 4.1% (3/74)	Knowledge pre-tests and post-tests (25 questions from NCLEX-RN® study set). (A priori equivalence bounds around the difference between the groups were set at ±5 points).	NCLEX-RN®[8] V – Previously validated but not reported R – Internal consistency reliability coefficients (KR-20) within acceptable range across all administrations of the knowledge test.	Level 2. Significant knowledge was gained in both simulated and traditional clinical experiences: pre (M = 60.05, SD 9.30) to post 1 (M = 62.68, SD 8.54, p = 0.015); post 1 (M = 62.68, SD 8.54) to post 2 (M = 64.78, SD 9.35, p = 0.028); and pre (M = 60.11, SD 9.32) to post 2 (M = 64.61, SD 9.39, p = 0.001). Both groups' knowledge scores were statistically equivalent (mean difference 0.49 (95% CI -3.58 – 4.56)) The scores for simulated-traditional and traditional-simulated were also statistically equivalent (mean difference -0.33 (95% CI -4.77 – 4.11).
Soccio 2017 (USA)	RCT (pilot study) including quantitative and qualitative program	Nursing students, juniors undertaking a mental	Three mental health simulation laboratories (labs) were designed and implemented (2 scenarios in each lab)	The control group was assigned to an inpatient psychiatric unit for 12	RR – not reported CR – 100% (48/48)	ATI RN Mental Health Mastery Examination (MHME)[10] 2013 was used to measure mental health	MHME: V- a valid nationally used exam developed in a national standard	Level 1, Level 2. Difference in ATI scores between groups did not reach significance (p =

	evaluation of effectiveness of mental health simulation in replacing traditional clinical hours in baccalaureate nursing education.	health unit (n=48)	followed by 9 weeks in inpatient psychiatric unit. (n=24). **DURATION 12 wk.** **SUBSTITUTION 25%.**	weeks. (n=24)	LTFU – 0% (0/48)	knowledge. Mental health self-confidence was measured using Mental Health Nursing Clinical Confidence Scale (MHNCCS) [11] Qualitative questionnaire	setting study-validated by ATI through qualitative and criterion referenced research R – not reported. MHNCCS: V – not reported R - Cronbach's alpha of 0.93. Test-retest overall correlation 0.859	0.590); ATI scores were higher in the experimental group (67% passed) compared to the control group (50% passed). Confidence: Pre–post clinical MHNCCS scores were significantly improved ($p < 0.0001$) with no effect of the group ($p = 0.646$). Qualitative data indicated students found the simulation helpful in learning how to manage patient behaviours.
Watson et al, 2012 (Australia)	Two parallel, randomised, single blind, controlled trials using a **non-inferiority** design, examined whether simulation can replace part of Physiotherapy clinical time. Students were stratified on academic grade and randomly	Physiotherapy students from six Australian universities undertaking clinical education in an ambulatory care setting with patients with musculoskeletal disorders. RCT 1 (n=192)	RCT 1: 1 week in simulation before 3 weeks of clinical immersion(n=96) RCT 2: 2 weeks of simulation in parallel during first 2 of 4 weeks clinical placement (n=89) **DURATION 4 wk.** **SUBSTITUTION**	Clinical placement in ward (4 weeks of traditional clinical immersion). RCT 1: n=96 RCT 2: n=89	RR – 30.5% (370/1200) CR - Overall: 94.1% (348/370) RCT1 INT: 94.8% (91/96), CON:93.8% (90/96) RCT2 INT: 93.3% (83/89), CON: 94.4%	Primary outcome blinded assessment of student competency conducted over two clinical examinations at week 4 using the Assessment of Physiotherapy Practice (APP)[2]measure. Secondary outcomes were student perceptions of experience (a 13-item scale on confidence). Clinical educator and	APP V- a validated tool used nationwide in PT practice, shown to discriminate four levels of competence[2]. R- high inter-examiner reliability ratings by 30 pairs of examiners (ICC = 0.92, 95% CI 0.84–0.96). Student	Level 3: RCT 1: mean APP score for two clinical examinations was 2.73 (SE 0.04) in the SLE group and 2.68 (SE 0.04) in the traditional group (CI of mean difference -0.07 to 0.17). Upper bound is less than the margin of 0.4, indicating no significant difference between groups. There were no between-group differences in scores for the seven competencies.

	allocated.	RCT 2 (n=178)	**25%.**		(84/89) LTFU - Overall 1.1% (4/370) RCT1 both INT and CON 1.4% (1/96, 1/96) RCT2 both INT and CON 1.1% (1/89, 1/89)	patients' ratings of student performance were employed.	Perceptions: V- not reported R- Cronbach's alpha ranged from 0.77 to 0.9 in current study Clinical educator and patients' ratings of student performance: V- not reported R- not reported	RCT 2: mean overall APP score was 2.61 (SE 0.05) in the SLE group and 2.58 (SE 0.05) in the traditional group.(CI of the mean difference -0.11 to 0.16). The upper bound of this CI was < 0.4, indicating no difference between groups. There were no between-group differences in scores for the seven competencies. Confidence: Students in both RCTs showed significant change in all measures over time ($p < 0.01$ in all cases), with confidence increasing at by the end of the placement.
Level 2: Quasi-Experimental Designs								
Baillie and Curzio 2009. (UK)	Quasi-experimental three-group design to test students' and facilitators' perceptions of simulation and practice learning	Nursing students, Yr 1, 2 and 3 (n=267)	8 simulation subgroups, undertaking full or half- day simulation program before, or during for a total of five clinical placement days. (n= 179) **DURATION 5 days.**	4 traditional clinical placement experience groups, over five days (n=88)	RR – not reported CR – Simulation or Clinical Placement 100% each group. Evaluation surveys Questionnaire	Pre-test and post-test evaluation tools were developed to elicit information for the Nursing & Midwifery Council's common evaluation tool; additional questions were devised re the simulation experiences.	Evaluation tools: V – Not reported. R – Not reported.	Level 1. The SBE intervention was seen by 93% (130/140) of students as increasing ability and 89% (125/141) for confidence in placement skill performance. No significant difference in confidence (p=0.364) or in preparedness (p=212) for placement skills

			SUBSTITUTION PROPORTION UNCLEAR.		1: INT: 94.4% (169/179) Questionnaire 2: Overall: 52.8% (141/267) INT: 40.8% (73/179) CON: 77.3% (68/88) LTFU Questionnaire 2: Overall: 47.2% (126/267)			programme between groups.
Curl et al 2016 (Texas, USA)	Quasi-experimental two-group study to test whether students participating in High Fidelity Simulation and clinical experiences in healthcare settings would attain	Nursing students in 3 different USA universities' Associate Degree in Nursing programs (LVN to RN) (senior students) (n=124)	Intervention comprised 20 simulation modules, 5 modules for each clinical specialty area, 4 hrs each (n=59) **DURATION 80 hr.** **SUBSTITUTION**	Traditional clinical experiences in obstetrics, paediatrics, mental health or critical care specialties over 160 hrs (n=65)	RR – not reported CR – 78.2% (97/124) completed the course and passed the unit INT: 84.7% (50/59) CON: 72.3%	Pre-test and post-test knowledge: tested by *Evolve-Reach* (Health Education Systems) (HESI) [12] Medical Surgical national exam. Post-test HESI Clinical Specialty exam related to the placement experience; satisfaction rated by	HESImedical-surgical and specialty exams: V- not reported R- not reported (authors stated the exam is nationally standardized with pass based on standardized scores)	Level 1, Level 2, Level 3. High fidelity simulations are an effective educational strategy. Post-test knowledge in the intervention group was significantly higher in the HESI medical-surgical post exam (p =0.05) and exit exam (p= 0.01), than for the control group. No significant difference in HESI clinical specialty

	knowledge/ skills equal to students participating exclusively in traditional clinical experiences.		**50%.** **RATIO 1:2.**		(47/65) LTFU – Overall: 4.0% (5/124) INT: 3.4% (2/59) CON: 4.6% (3/65) Course Failure INT: 11.9% (7/59) CON : 23.1% (15/65)	purposely developed Student Evaluations of Clinical Simulation Effectiveness Tool (SECET), clinical performance and simulation performance objectively evaluated by teachers.	SECET: V – not reported R – not reported Objective evaluations V – not reported R – not reported	standard exam scores.
Meyer 2011 (USA)	Quasi-experimental study to evaluate effects of a theory-driven pediatric simulation curriculum on nursing students' clinical performance using a staggered timing model; two of three	Nursing Students, junior, enrolled in a pediatric clinicial course (n=120)	Students attended simulation in groups of 10 for 2 weeks (24 hrs of simulation and 72 hrs of clinical) of an 8-week clinical semester (4 simulation sessions, 2 scenarios in each. **DURATION 2 wk.** **SUBSTITUTION 25%.**	The staggered timing model allowed for comparison with students who had not yet completed the stimulation at weeks 2, 4 and 6.	RR – 100.0% (120/120) CR – 96.7% (116/120) LTFU – 3.3% (4/120)	Students' clinical performance was assessed using a Likert-style tool by the clinical faculty at 2-week intervals, using an adapted validated performance evaluation tool.	Adapted tool V –validated through continual use by the school. R - Cronbach's alpha 0.992 among four raters, to 0.996 for nine raters.	Level 2, Level 3. Those students attending the simulation before clinical scored higher than those who had not yet attended simulation (mean: 1.74 Std error ±0.75, p= 0.02). On item-level analysis, therapeutic skills were positively impacted by simulation (p= 0.02). The timing of the simulation during clinical rotation had no significant effect on student overall performance (p=0.244)

	blocks had random assignment.							
Level 3 Observational –Analytic Designs								
Giblett et al 2017 (UK)	Prospective 2-group post-test study to examine student awareness and knowledge of safe surgical practices using a simulated surgical patient pathway compared with a control group receiving traditional surgical education programs.	Medical students during first clinical year (n=104)	Medical students participated in simulated surgical patient pathway, comprising 7 half-day sessions using multiple modes of simulation along with clinical experience. (n=50) **DURATION 21 hr.** **SUBSTITUTION PROPORTION UNCLEAR.**	Traditional surgical education program, students assigned to a surgical team in alternative hospitals These students went on to complete the simulation in semester 2 (n=54)	RR – not reported CR – training not reported Knowledge based assessment (semester 1) Overall: 85.5% (89/104) INT: 78.0% (39/50) CON: 92.6% (50/54) Evaluation Overall: 81.7% (85/104) INT: 70.0% (35/50) CON: 92.6%	Post-test surveys were completed by both groups at the end of semester 1. A clinical knowledge test mapped to students' learning outcomes was conducted (CON reassessed at the end of semester 2); perceptions of teaching methods used and perceived confidence in the assessment of surgical patients were surveyed.	Clinical knowledge test: V – not reported R – not reported Evaluation V – not reported R – not reported	Level 1, Level 2. All domains of the student satisfaction survey scored higher in group 1 (INT) compared to group 2 (CON) (all: $p < 0.001$), as did all domains of student confidence and self-evaluation of understanding surgical principles ($p < 0.001$). Students in simulation pathway were significantly more knowledgeable than the control group ($p<0.001$). Groups 2 showed a significantly improved subjective experience of surgical teaching, with greater awareness and confidence of safe surgical principles after the simulation.

					(50/54) LTFU - Knowledge based assessment Overall 14.4% (15/104) INT: 22.0% (11/50) CON: 7.4% (4/54) Evaluation Overall: 18.3% (19/104) INT: 30.0% (15/50) CON: 7.4% (4/54)			

1. Footnotes: authors' use of Simulated learning experiences SLE or simulation based education SBE terminology is referred to as 'simulation' in the table.
2. Dalton, Davidson & Keating (2011)
3. Hayden, Keegan, Kardong-Edgren & Smiley (2014)
4. Berkow, Virkstis, Stewart & Conway (2008)
5. Berkow, Virkstis, Stewart, Aronson & Donohue (2011)
6. Assessment Technologies Institute, LLC
7. ATI RN Comprehensive Predictor® 2010 (Assessment Technologies Institute, LLC), At: https://www.atitesting.com/Solutions/DuringNursingSchool/ComprehensiveAssessmentAndReviewProgram.aspx
8. https://www.ncsbn.org/nclex.ht

9. Levitin and Kushnir (2012)
10. ATI Mental Health Mastery website: (https://quizlet.com/246802539/mental-health-ati-mastery-flash-cards/)-Exams/Quizlets. At: https://quizlet.com/246802539/test
11. Bell, Horsfall & Goodin (1998)
12. HESI for Nursing. At: https://evolve.elsevier.com/education/nursing-review-and-testing/

Table 3 Methodological quality of included studies based on MERSQI domain scores (Mean value scores from two raters)

Domain	Element [possible score]	Study first author									
		Baillie (2009)	Blackstock (2013)	Curl (2016)	Giblet (2017)	Hayden (2014)	Kimhi (2016)	Meyer (2011)	Schlairet (2010)	Soccio (2017)	Watson (2012)
Study design	1. Study design [3]	2.0	3.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0	3.0
Sampling	2. Sampling [1.5]	1.0	1.5	1.5	0.5	1.5	1.5	1.5	1.5	1.5	1.5
	3. Response rate [1.5]	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	0.5	0.5
Type of data	4. Type of data [3]	1.0	2.5	1.0	3.0	3.0	1.0	3.0	3.0	3.0	3.0
Validity of evaluation instrument	5. Internal structure [1]	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0
	6. Content [1]	0.0	1.0	0.0	0.0	1.0	0.5	0.0	1.0	1.0	1.0
	7. Relationships to other variables [1]	0.0	1.0	0.5	1.0	0.5	0.0	0.5	0.5	0.5	0.5
Data analysis	8. Appropriateness of data analysis [1]	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	9. Complexity of analysis [2]	1.5	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
Outcomes	10. Outcomes [3]	1.0	2.0	1.5	1.5	1.5	1.0	1.5	1.5	1.5	2.0
	Mean total score/18	9.0	16.5	11.0	12.5	16.0	11.5	13.0	15.0	15.0	15.5
	Kappa Measure of agreement	.85	.86	.87	1.00	.87	.86	.75	.87	.87	.87