University of Wollongong

Research Online

Faculty of Science, Medicine and Health -
Papers: Part B

Faculty of Science, Medicine and Health

2019

A systematic review of the effectiveness of empathy education for undergraduate nursing students

Tracy Levett-Jones

University of Technology, Sydney

Robyn Cant

University of Technology, Sydney, Robyn.Cant@uts.edu.au

Samuel Lapkin

University of Wollongong, slapkin@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/smhpapers1>

Publication Details Citation

Levett-Jones, T., Cant, R., & Lapkin, S. (2019). A systematic review of the effectiveness of empathy education for undergraduate nursing students. Faculty of Science, Medicine and Health - Papers: Part B. Retrieved from <https://ro.uow.edu.au/smhpapers1/728>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

A systematic review of the effectiveness of empathy education for undergraduate nursing students

Abstract

Objective

The objective of this systematic review was to identify, critically appraise and synthesize evidence for the effectiveness of empathy interventions in undergraduate nursing education.

Design

A systematic review of literature.

Data Sources

A three-stage systematic search of six electronic databases was conducted.

Review Methods

The Preferred Reporting Items for Systematic reviews and Meta-Analyses guided the review. English language articles published between 2000 and 2018 were eligible. Methodological rigour was examined using the Medical Education Research Study Quality Instrument. Changes in empathy were assessed using Cohen's effect size correlation (r) and reported as effective when the variance was >0.2 standard deviations ($r \geq 0.2$).

Results

Of 23 included studies, four were experimental and four were case-control studies. Of these, the mean effect size was $r = 0.45$ and three were regarded as effective empathy interventions. Although 10 of 13 single group studies demonstrated a significant change in empathy between pre-test and post-test ($p < 0.05$), effect sizes were often low (mean $r = 0.26$). Six single-group studies reported an intervention effect of $r > 0.2$. The most effective empathy education involved immersive and experiential simulation-based interventions. Simulation modalities ranged from role plays, manikin-based scenarios, to 3D e-simulations and point-of-view simulations where students wore a hemiparesis suit\

Conclusions

Nine of 23 empathy education studies in undergraduate nurse education demonstrated practical improvements in empathy. The most effective interventions were immersive and experiential simulations that focused on vulnerable patient groups and provided opportunities for guided reflection. We noted the research designs were limited in terms of levels of evidence and use of subjective measures. Larger experimental studies are required to provide higher levels of evidence to identify unequivocal outcomes in terms of empathy research. Future studies should consider transfer to practice and longer-term changes in empathy as study outcomes.

Keywords

review, effectiveness, empathy, education, undergraduate, nursing, systematic, students

Publication Details

Levett-Jones, T., Cant, R. & Lapkin, S. (2019). A systematic review of the effectiveness of empathy education for undergraduate nursing students. *Nurse Education Today*, 75 (1), 80-94.

A SYSTEMATIC REVIEW OF THE EFFECTIVENESS OF EMPATHY EDUCATION FOR UNDERGRADUATE NURSING STUDENTS

ABSTRACT

Objective

The objective of this systematic review was to identify, critically appraise and synthesize evidence for the effectiveness of empathy interventions in undergraduate nursing education.

Design

A systematic review of literature.

Data sources

A three-stage systematic search of six electronic databases was conducted.

Review methods

The Preferred Reporting Items for Systematic reviews and Meta-Analyses guided the review. English language articles published between 2000 and 2018 were eligible. Methodological rigour was examined using the Medical Education Research Study Quality Instrument. Changes in empathy were assessed using Cohen's effect size (r) and reported as effective when the variance was $\geq .2$ standard deviations ($r = > .2$).

Results

Of the 23 included studies, four were experimental and four were case-control studies. Of these, the mean effect size was $r = 0.45$ and three were regarded as effective studies. Although 10 of 13 single group studies demonstrated a significant change in empathy between pre-test and post-test ($p < 0.05$), effect sizes were often low (mean $r = 0.26$). Six single-group studies reported an intervention effect of ≥ 0.2 . The most effective empathy interventions included immersive and experiential simulation-based interventions. Simulation modalities ranged from role plays, manikin-based scenarios, to 3D e-simulations and point-of-view simulations where students wore hemiparesis suits.

Conclusions

Nine of 23 studies of empathy in undergraduate nurse education demonstrated practical improvements in empathy ($r > 0.2$). The most effective empathy interventions were immersive and experiential simulations that focused on vulnerable patient groups and provided opportunities for guided reflection. We noted the research designs were limited in terms of levels of evidence and use

of subjective measures. Larger experimental studies are required to provide higher levels of evidence to identify unequivocal outcomes in terms of empathy research. Future studies should consider transfer to practice and longer-term changes in empathy as study outcomes.

KEYWORDS

Empathy; nursing education; effectiveness; systematic review; students, nursing.

40 **INTRODUCTION**

41 Empathy is a multidimensional construct and various definitions of the term are evident in the
42 literature. Generally speaking, empathy involves the cognitive ability to comprehend what another
43 person is feeling, an emotional resonance with those feelings, and the willingness to respond
44 appropriately to the person's needs (Hatfield 2011). In healthcare, empathy is considered a basic
45 component of therapeutic relationships and a critical factor in patients' definitions of quality care
46 (Hojat 2013). More than 200 studies have demonstrated the positive impact of empathic healthcare
47 interactions on patient outcomes (Trzeciak 2017). Empathetic encounters with healthcare
48 professionals result in: decreased levels of depression, anxiety, distress; and increased levels of
49 emotional wellbeing, satisfaction and adherence to treatment regimens (Hojat 2013). Examples of
50 physiological consequences of empathic healthcare interactions include: reduced rates of infection,
51 improved wound healing, higher cancer survival rates, and a reduction in diabetes complications,
52 and pain (Scott 2011, Trzeciak 2017). However, despite the increasing recognition of the impact of
53 empathy on patient outcomes, there is compelling research indicating that contemporary healthcare
54 is besieged by a generalised lack of empathy (Lown 2011), which manifests as apathy, indifference,
55 callousness, cruelty, and dehumanisation; and results in increased risk of harm to patients (West
56 2006).

57 To address these issues, and in recognition of the fact that students' empathy levels can decline
58 during the period of enrolment in a healthcare degree (Ward, Cody et al. 2012), nursing programs
59 are increasingly implementing educational interventions designed to enhance students' empathy
60 levels. However, to date, a limited number of studies have sought to examine the effectiveness of
61 empathy interventions.

62 Previous reviews have identified that communication skills training, experiential learning and
63 interprofessional educational interventions are effective strategies to improve medical students'
64 empathy levels (Batt-Rawden 2013), and that simulation-based education is an appropriate method
65 to teach empathy (Bearman 2015). However, these reviews are limited by the primary focus on
66 medical students (Batt-Rawden 2013) and specific educational approaches (Bearman 2015). Thus,
67 the current review aims to address this gap by focusing on nursing students and the impact of a
68 broad range of educational intervention on empathy levels.

REVIEW METHODS

Study Objective

A systematic review includes a critical appraisal of all related evidence using a clearly documented methodology to answer a specific research question (Higgins 2011). The objective of this review was to systematically identify, appraise and synthesize the best available evidence for the effectiveness of empathy interventions in undergraduate nursing education in comparison with an alternative or no intervention. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist was used as a guide to the search and reporting of the results (Moher 2009).

Eligibility criteria

Types of studies - The review included primary studies that used experimental and quasi-experimental designs published in English between January 2000 and May 2018.

Types of participants - The review considered studies of undergraduate (or pre-licensure) nursing students regardless of gender, age or country.

Types of interventions/phenomena of interest - The review included studies of educational interventions designed to enhance nursing students' empathy levels. Studies of similar concepts such as 'caring', 'compassion' or 'emotional intelligence' were excluded.

Types of outcome measures - The review considered any objectively measured or self-reported quantitative data reporting on empathy outcomes and assessed using validated instruments.

Search strategy

A three-step search strategy was utilised for this systematic review. The initial search involved free text searches of Google Scholar undertaken to gain information about the breadth of publications. A search was then conducted of the PubMed database (see Box 1) via the US National Library of Medicine to identify Medical Subject Headings. It was noted, however, that the MeSH term 'empathy' was not linked to any educational subgroups in the search tree. This was followed by conducting searches in CINAHL Plus, SCOPUS, PubMed (Ovid Medline), and Education Resources Information Center (ERIC). We used a systematic strategy based on derivatives of 'nursing students' (Population); 'education' or 'teaching' and 'empathy' (Intervention); 'undergraduate' or 'prelicensure' or 'university' (Context); and 'effectiveness' or 'impact' (Outcome). Lastly, the reference lists of recent studies and reviews were hand searched for eligible papers that may have been previously missed.

INSERT: Box 1 about here

Study selection

All citations and abstracts identified by the search strategy were downloaded to an Endnote database and duplicates were identified and removed. Potential articles were screened for eligibility by title and abstract. Articles that met the inclusion criteria were independently reviewed by two authors (RC and TLJ).

Critical appraisal

The methodological quality of each study was assessed using the 10-item Medical Education Research Study Quality Instrument (MERSQI) for experimental, quasi-experimental, and observational studies (Reed, Beckman et al. 2009). The MERSQI comprises six domains: study design, sampling, type of data collected, validity of instruments, data analysis, and level of outcomes, with a potential score range of 5-18. Substantial reliability of the MERSQI (Cronbach's alpha > 0.60 - >0.72) has previously been reported in medical research (Reed 2007, Cook and Reed 2015). In recent nursing research, interrater agreement using the MERSQI across 26 studies was high (ICC= 0.988) (Cant 2018).

Data collection process

The study variables of interest (study origin, design, aim, sample, intervention, evaluation measures, validity and reliability of measurement instruments and results) were extracted from included papers and tabulated using a standardised form (see Table 1). In addition, relevant information from included studies was discussed in the results section of the review.

Results synthesis

Descriptive and summary statistics (means, standard deviations) were used to report results using IBM-SPSS Statistics for Windows Vs 25 (Armonk, NY: IBM Corp.) to collate the extracted and tabulated data. We intended to pool the quantitative study results and to present a statistical meta-analysis to improve estimates of the size of the intervention effects. However, there was substantial heterogeneity with regard to the interventions and outcome measures in the included studies and reporting using this approach was not regarded as being valid (Field and Gillett 2010). The findings are therefore presented as a narrative summary.

To aid data presentation, the impact of each educational intervention was examined by testing the comparative strength of study outcomes using effect sizes. Standardized effect sizes for the primary outcome (change in empathy level) were computed from study instrument scores and standard deviations using an online calculator (<https://www.uccs.edu/lbecker/>). They were reported using Cohen's *d* statistic presented as effect size *r* (a Pearson correlation co-efficient). The *d* is the

standardized mean difference between two groups and r describes the proportion of the variance that is explained. A small effect for the r statistic is generally interpreted as 0.10, a medium effect 0.30, and a large effect 0.50. In this study the change in mean empathy scores was considered to be negligible when r was less than 0.2, as this indicates that the means of two groups differ by less than 0.2 standard deviations and the difference, even when statistically significant, is trivial (Polit 2012 p. 478). An effect size of more than $r = 0.20$ was regarded as a valid measure of change in empathy that was both practical and likely to be clinically meaningful.

RESULTS

Study selection

The search produced a total 1176 articles. After the study selection process (see Figure 1) 23 studies were included in the review.

INSERT: Figure. 1. PRISMA flow chart about here.

Characteristics of included studies

The included studies were published between 2004 and 2018 and were conducted in eight countries: 11 in the USA, four in Turkey, three in Australia, and one each from Canada, Italy, Spain, Sweden and Taiwan (see Table 1). Twelve of the studies used single group pre-test post-test designs with outcomes generally assessed immediately after the intervention: Beddoe and Murphy (2004), Mete (2007), Ozcan, Bilgin et al. (2011), Ward, Cody et al. (2012), Sheehan, Perrin et al. (2013), Nosek, Gifford et al. (2014), Everson, Levett-Jones et al. (2015), Fleming, Thomas et al. (2015), Ward (2016), Bas-Sarmiento, Fernández-Gutiérrez et al. (2017), and Levett-Jones (2017). Two further studies used post-test only designs: Lobchuk, Halas et al (2016) and Streklova, Kreiger et al (2017).

Insert Table 1 about here Characteristics of included studies

Of the eight studies that assessed two-groups of students, four were RCTs and four were case-control studies with each of them testing the change in empathy after an educational intervention and comparing empathy outcomes with those of a control group: Wikstrom (2003), Webster (2010), Henry, Ozier et al. (2011), Cunico (2012), Mennenga, Bassett et al. (2016), Geçkil, Kaleci et al. (2017), Haley, Heo et al. (2017), and Lee, Yu et al. (2018). Four of these studies randomized students into an intervention or control group: Henry, Ozier et al. (2011), Mennenga, Bassett et al. (2016), Geçkil,

Kaleci et al. (2017), and Haley, Heo et al. (2017). Two longitudinal studies reassessed empathy levels at four weeks (Haley, Heo et al. 2017) or six months (Geçkil, Kaleci et al. 2017) after the intervention.

Participant characteristics

All the studies recruited students using convenience sampling. Overall, n = 3250 participants were involved in the empathy interventions and the average sample size was n = 141 (range: 11 - 460). Nursing students enrolled in all years of their respective nursing programs were represented, however, most were in their second or third year of study.

Types of outcome measures

In all, the included studies reported the use of 21 different empathy instruments, including translations of several English-language instruments into Italian, Spanish, Turkish or Chinese. All instruments were described as having been previously developed and authors generally provided prior evidence of validity and reliability, often (but not always) reporting reliability outcomes from their current studies also. Details of the empathy instruments used are provided in the Table 1. The most frequently used instruments were:

- Jefferson Scale of Empathy-Health Professions Students (JSE-HP-S): 20 items, 7-point Likert-type items; measuring Perspective Taking, Compassionate Care and Standing in the Patient's Shoes (used in three studies); and the Jefferson Scale of Empathy for Health Professionals (JSE-HP) (used in two studies).
- Interpersonal Reactivity Index (IRI): 20 items, 7-point Likert-type response scale; subscales of Perspective Taking, Fantasy, Empathic Concern, Personal Distress (used in 3 studies).
- Kiersma-Chen Empathy Scale (KCES): 15 items, 7-point Likert-type response scale and a Modified Kiersma-Chen Empathy Scale (MKCES) of 8 items (each were used in one study).

Four studies used objective independent measures to test empathy application/performance rather than subjective self-assessments of empathy levels. Lobchuk, Halas et al (2016) utilized students' electronic responses to video excerpts as assessments of empathetic communications, which were then rated by faculty staff. Objective Structured Clinical Examinations of empathy performance were conducted in the study by Lee, Yu et al (2018), in addition to participant surveys, and in the study by Bas-Sarmiento et al (2017). Ward et al (2016) developed a checklist so that standardized patients and simulated family members could rate students' empathy.

Methodological quality components

The studies were assessed for methodological quality. Following MERSQI assessment, one study (Chen, Kiersma et al. 2015) was removed because of a low score of less than 50%. For the 23

included studies the scores ranged from 9.0 (50%) to 12.5 (69%) out of a possible 18 points, with a mean of 10.9 (60.7% on a scale of 100) (see Table 1). This range was interpreted as moderate in quality.

The main factors that impacted the scores was the use of self-report surveys and a lack of objective instruments. In the MERSQI the use of self-report surveys receives 1 point, whereas 3 points are awarded for use of objective approaches (for example, observation by a trained observer, feedback from a standardized patient, or computer-based ratings of performance). Only three studies used objective measures and consequently achieved higher scores.

Higher MERSQI scores were achieved by the eight studies that used case-control designs or randomized two-group studies, along with those that reported a response rate or that achieved >75% sampling. Levels of evidence in the designs were often low, with only four experimental studies. Most studies (15/23) examined interventions in a single group of students without a comparison group. Each of these criteria limited the quality scores.

Empathy outcomes for single group study designs

Thirteen single group studies reported pre- and post-test empathy comparisons, often using test-retest measurements, with 12 identifying a significant change in overall or an empathy subscale after the intervention (see Table 1 and Table 2). Mean empathy scores improved significantly post intervention in 10 of 12 studies. One study, however, reported significant decline in mean empathy scores for a particular group of undergraduate nursing students in a longitudinal study that was conducted over a year (Ward, Cody et al. 2012). Notably, this study could be regarded as a baseline study, as it tracked students' empathy before and after traditional teaching experiences (problem-based learning and clinical placements) without a specific empathy intervention. In two studies there was no change in empathy scores. Two studies used post-test only designs so no direct comparisons were available.

Only two of the single group studies showed a large effect size $r = >.5$. Fleming et al. (2015) reported improved ethno-cultural empathy after an 'Appreciating Diversity' workshop with first year nursing students; and Bas-Sarmiento et al. (2017) reported higher empathy scores following a 20-hour multi-method intervention which included simulation-based role plays for second-year students.

Despite the positive impact of the interventions in the single group studies, the effect sizes for change in overall empathy or for empathy subscales were generally low (see Table 2), ranging from $r = .09$ to a high of $r = .73$. When taken as an average effect from each study the overall mean was r

= .26. In five of the studies the empathy intervention effects were seen to be lower than .20. We conclude, therefore, that 6 of the 11 single group studies had a real-world impact on empathy.

INSERT Table 2 about here Single group designs

Empathy outcomes for case-control studies and RCTs

The eight studies that compared two groups of students included four studies that randomized participants into intervention and control groups and four case-control studies (see Table 3). Five of the eight studies reported a significant increase in empathy scores in the intervention group from pre-test to post-test ($p > 0.05$) and this outcome was also reflected in effect sizes which were greater in the intervention groups. Three of the studies had an effect size greater than $r = .2$. There was a mix of effects in both the case-control and RCT designs. Their average gain computed by extracting the highest effect reported in an instrument in the initial pre- and post-test in each study or an objective test (0.29; 0.59; 0.50), was $r = .45$, which is a medium effect size. All eight studies reported non-statistical change in pre- and post-test empathy scores for the control group. A range of different measurement instruments were used throughout the studies.

INSERT Table 3 about here: Change in empathy scores identified in case-control studies/RCTs

Effective studies

Six single-group studies reported an adequate level of the intervention effect in participants' empathy levels after exposure to an empathy intervention (Ozcan, Bilgin et al. 2011, Sheehan, Perrin et al. 2013, Fleming, Thomas et al. 2015, Bas-Sarmiento, Fernández-Gutiérrez et al. 2017, Levett-Jones 2017, Everson 2018) (see Table 2). Of the eight case-control studies, only three (Geçkil, Kaleci et al. 2017, Haley, Heo et al. 2017, Lee, Yu et al. 2018) showed a statistical improvement of ≥ 2 standard deviations over and above the change in a control group, thus demonstrating an effective empathy intervention. This suggests that participants in the intervention groups in the effective studies experienced a substantial increase in empathy levels when compared with the control group. In the other studies, less so. These results indicate that few of the designs used in the included empathy studies resulted in a substantial change in empathy scores.

Effective empathy education strategies

Some of the most effective empathy interventions were immersive and experiential simulation-based interventions. Six of the studies used this approach but with a range of different simulation modalities such as role plays (Bas-Sarmiento, Fernández-Gutiérrez et al. 2017, Geçkil, Kaleci et al. 2017), manikin-based scenarios (Haley, Heo et al. 2017), 3D e-simulations (Everson, Levett-Jones et al. 2015), and simulations where students wore hemiparesis suits (Levett-Jones 2017). The latter two were point-of-view simulations in which learners experienced the world ‘through the eyes’ of another person and learnt to ‘walk in the patient’s shoes’ (Levett-Jones et al, 2018). It should be noted that most of these empathy interventions included a guided reflection/debriefing component. Other effective empathy interventions included Fleming et al.’s. (2015) Appreciating Diversity workshop and Ozcan et al.’s (2011) arts and psychodrama program.

The foci of the empathy interventions typically included vulnerable patient groups such as those from culturally and linguistically backgrounds, frail older people, patients with a mental illness or who had an acquired brain injury. Additionally, while several of the effective empathy interventions consisted of a single simulation scenario, others were longitudinal in nature (eg. 21 hours over 6 weeks, or 20 hrs over a 12-week semester). However, there was no statistical correlation between program duration and the impact on empathy levels.

Nearly all of the studies in this review measured empathy in the short-term, immediately after an educational intervention. Those studies that measured empathy over the period of undergraduate enrolment (3 years) demonstrated that students’ empathy levels varied over time (Mete 2007, Cunico 2012). Few studies (Mete 2007, Henry, Ozier et al. 2011, Cunico 2012, Fleming, Thomas et al. 2015, Ward 2016, Haley, Heo et al. 2017) conducted longer-term empathy evaluations, for example a 4-week follow-up. None of the studies measured transfer to clinical practice, although some were longitudinal in nature and repeated empathy scores after a semester or after one year, which included the influence of clinical practice experience.

DISCUSSION

Few healthcare interventions have as much impact on a patient’s physical and emotional well-being as empathy (Trzeciak 2017). Thus, educators have a responsibility to ensure that empathy is an integral curricula component and that the outcomes of interventions designed to enhance students’ empathy levels are rigorously assessed.

In this study we reviewed 23 quantitative studies of empathy interventions for undergraduate nursing students and found that less than half (9/25) were effective in improving empathy levels in a significant and practical way.

Overall, single group studies without a control group (e.g. pre-test post-test single group designs) had a mean effect size of $r = 0.26$, with six of 11 showing a practical improvement of $r = \geq 0.2$ in empathy levels. Two of the RCTs and one case-control study reported significant practical effects with a mean effect size of $r = 0.45$. This difference in outcomes between different designs should not be interpreted as superiority among designs as there was much variability in the designs, sampling, teaching strategies, program duration and sample size. It is logical, however, that even in the studies that were not identified as effective in enhancing empathy, the interventions may well have resulted in attitudinal shifts and other beneficial affective and cognitive changes related to empathy that were not directly measured.

It should be noted that the validity of some empathy instruments has been called into question. For example, Yu and Kirk (2009) reviewed the properties of 12 empathy measures in nursing using objective criteria and found that overall quality scores were generally low. Although most of the instruments had undergone rigorous development and psychometric testing, according to Yu and Kirk, none were both psychometrically and conceptually satisfactory in examining all the domains of empathy. However, since 2009 most of the measures included in that paper have been revised or modified. For example, the Jefferson Scale of Physician Empathy (JSPE) was modified for healthcare professionals (JSE-HP) and nursing students (JSE-HP-S) and was found to be valid and reliable (Ward 2009, McMillan and Shannon 2011). Conversely, the Interpersonal Reactivity Index (IRI) used in three of the studies included in our review was not developed for healthcare professionals and showed low subscale correlations during development, although moderate test-retest reliability (Yu and Kirk 2009).

Measurement can also be fraught with difficulty because there are numerous definitions of empathy and debate about the elements/domains of empathy being measured, e.g. state versus trait empathy; affective versus cognitive empathy (Fields, Mahan et al. 2011). These issues suggest that researchers should consider the purpose, appropriateness and relevance of specific empathy instruments prior to use, and that psychometric testing should also be undertaken with the sampled populations.

While some studies indicate that nursing students' empathy levels can decline during the period of enrolment in an undergraduate degree (Ward, Cody et al. 2012), others suggest that nursing students may have higher empathetic tendencies than students from other disciplines, possibly diminishing a teaching effect (Gallagher, Moriarty et al. 2017). In this regard, McKenna et al. (2012) reported that nursing students from first to third year had overall consistently high empathy levels when measured using the JSPE- Health Professions. Several studies in the current review aligned with this notion, for example Everson et al. (2015) who examined the empathetic views of $n = 460$

second year nursing students found their empathy scores at pre-test averaged 86% (47.86/56) on the Modified Kiersma-Chen Empathy Scale. However, it should be noted that many of these studies used self-report surveys which can be influenced by social desirability bias. Overall, the current review does little to help unpack these underlying issues because of the use of numerous measurement instruments, small samples and the heterogeneous nature of the included studies which restricted a formal meta-analysis.

Irrespective of the issues outlined above, our review identified that educational interventions can be effective in improving nursing students' empathy levels. Immersive and experiential simulation-based interventions, particularly point-of-view simulations that allow learners to view the world 'through the eyes' of another person (Eymard, Crawford et al. 2010) were identified as effective, at least in the short term. This finding aligns with previous research indicating that simulations that allow learners to 'step into a patient's shoes' help them to gain new insights into the feelings, perspectives, experiences and needs of another person (Batt-Rawden 2013, Bearman 2015). Additionally, the findings from our review indicate that the factors identified as significantly associated with higher effect sizes include a focus on vulnerable patient groups and opportunities for guided reflection/debrief. Unlike previous research (Teding van Berkhout and Malouff 2016), the duration of the intervention and time between intervention and assessment were not statistically significantly associated with effect size.

Further research is necessary to fully understand the specific types of empathy that can be enhanced using simulation and the most effective educational strategies to use. Studies also require a higher level of evidence to address both the possibility of bias introduced through use of self-report questionnaires and other potential forms of bias and error (Polit 2012 p. 477). For example, single group studies that measure the impact of an educational intervention according to the difference between pre-test and post-test surveys have certain threats to internal validity in not allowing for comparison with other students in the same cohort who were not exposed to the intervention. Additionally, in studies where training is conducted over time (such as over a semester) bias can occur because it is not possible to verify whether any changes result from the educational intervention or from another learning activity, such as clinical exposure. Researchers should therefore consider the use of random samples with matched control groups and crossover designs where each group experiences the same intervention. Other critical aspects that have not been explored in the literature include transfer of empathy skills to practice and patients' perspectives of learners' empathy skills. These are important considerations and should be a feature of future studies.

Limitations

This review is limited by the exclusion of non-English language studies. Additionally, most of the included studies used self-report measures to identify outcomes, with little use of more objective measures. Regardless, in the current paper we have systematically applied research methods and appropriate guidelines to review the literature and thus, our results extend upon and update, what is known about empathy education for nursing students.

CONCLUSION

This paper reviewed 23 studies reporting on educational interventions designed to enhance nursing students' levels of empathy. The most effective empathy interventions were immersive and experiential simulations that focused on vulnerable patient groups and provided opportunities for guided reflection. The research designs of the included studies were, however, somewhat limited in terms of levels of evidence and use of subjective measures. A greater depth of exploration of the components of empathy and a focus on the behavioural domain is needed. Larger experimental studies are required to provide higher levels of evidence to identify unequivocal outcomes in terms of empathy research.

The results of this review, while informative and encouraging, indicate that future studies should consider transfer to practice and longer-term changes in empathy as study outcomes. Additionally, larger, well-planned and rigorous studies involving more than a single cohort, are required.

References

- Bas-Sarmiento, P., M. Fernández-Gutiérrez, M. Baena-Baños and J. M. Romero-Sánchez (2017). Efficacy of empathy training in nursing students: A quasi-experimental study. *Nurse Education Today* 59: 59-65.
- Batt-Rawden, S., Chisolm, M., Anton, B., and T. Flickinger (2013). Teaching empathy to medical students: An updated systematic review. *Academic Medicine* 8: 1171-1177.
- Bearman, M., Palermo, C., and B. Williams (2015). Learning empathy through simulation. *Simulation in Healthcare* 10: 308-319.
- Beddoe, A. E. and S. O. Murphy (2004). Does mindfulness decrease stress and foster empathy among nursing students? *Journal of Nurse Education* 43(7): 305-312.
- Cant, R., Levett-Jones T, and A. James (2018). Do simulation studies measure up? A simulation study quality review. *Clinical Simulation in Nursing* in press, 2018.
- Chen, A. M., Kiersma, M. E., Yehle, K .S., and K. S. Plake (2015). Impact of the Geriatric Medication Game^(R) on nursing students' empathy and attitudes toward older adults. *Nurse Education Today* 35(1): 38-43.
- Cook, D. A. and D. A. Reed. (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.
- Cunico, L., Sartori, R., Marognolli, O., and A. M. Meneghini (2012). Developing empathy in nursing students: a cohort longitudinal study. *Journal of Clinical Nursing* 21: 2016-2025.
- Everson, N., Levett-Jones, T., Lapkin, S., Pitt, V., van der Riet, P., Rossiter, R., Jones, D., Gilligan, C., and H. Courtney-Pratt (2015). Measuring the impact of a 3D simulation experience on nursing students' cultural empathy using a modified version of the Kiersma-Chen Empathy Scale. *Journal of Clinical Nursing* 24(19-20): 2849-2858.
- Everson, N., Levett-Jones, T., Pitt, V., Lapkin, S., Van Der Riet, P., Rossiter, R., .and H. C. Pratt (2018). Analysis of the Empathic Concern Subscale of the Emotional Response Questionnaire in a study evaluating the impact of a 3D Cultural Simulation. *International Journal of Nursing Education Scholarship* 15(1): DOI: 10.1515/ijnes-2017-0003.
- Eymard, A. S., Crawford, B. D., and T. M. Keller (2010). 'Take a Walk in My Shoes': Nursing students take a walk in older adults' shoes to increase knowledge and empathy. *Geriatric Nursing* 31(2): 137-141.

408 Field, A. P., and R. Gillett (2010). How to do a meta-analysis. *British Journal of Mathematical and*
409 *Statistical Psychology* 63(3): 665-694.

410 Fields, S. K., Mahan, P., Tillman, P., Harris, J., Maxwell, K., and M. Hojat (2011). Measuring empathy
411 in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider--
412 student version. *Journal of Interprofessional Care* 25(4): 287-293.

413 Fleming, B. D., Thomas, S. E., Burnham, W. S., Charles. L. T., and D. Shaw (2015). Improving
414 Ethnocultural empathy in healthcare students through a targeted intervention. *Journal of Cultural*
415 *Diversity* 22(2): 59-63.

416 Gallagher, P., Moriarty, H., Huthwaite, M., and B. Lim (2017). Challenging some assumptions about
417 empathy. *Clinical Teacher* 14(6): 437-440.

418 Geçkil, E., Kaleci, E., Cingil, D., and F. Hisar. (2017). The effect of disability empathy activity on the
419 attitude of nursing students towards disabled people: a pilot study. *Contemporary Nurse*: 53(1): 82-
420 93.

421 Haley, B., Heo, S., Wright, P., Barone, C., Rettigantid, M. R., and M. Anders (2017). Effects of using an
422 Advancing Care Excellence For Seniors simulation scenario on nursing student empathy: a
423 randomized controlled trial. *Clinical Simulation in Nursing* 13(10): 511-519.

424 Hatfield, E., Rapson, R. L., and Le, Y. C., (2011). Emotional contagion and empathy. *The Social*
425 *Neuroscience of Empathy* January(19).

426 Henry, B. W., Ozier, A. D., and A. Johnson (2011). Empathetic responses and attitudes about older
427 adults: How experience with the aging game measures up. *Educational Gerontology* 37(10): 924-941.

428 Higgins, J., and S.Green (2011). *Cochrane Handbook for Systematic Reviews of Interventions* version
429 5.1.0., Cochrane Collaboration.

430 Hojat, M., Louis, D.Z., Maio, V., and J. S. Gonnella (2013). Empathy and health care quality. *American*
431 *Journal of Medical Quality* 28: 6-7.

432 Lee, K. C., Yu, C. C., Hsieh, P. L., Li, C. C., and Y. C. Chao (2018). Situated teaching improves empathy
433 learning of the students in a BSN program: A quasi-experimental study. *Nurse Education Today* 64:
434 138-143.

435 Levett-Jones, T., Lapkin, S., Govind, N., ...and N. Everson (2017). Measuring the impact of a 'point of
436 view' disability simulation on nursing students' empathy using the Comprehensive State Empathy
437 Scale. *Nurse Education Today* 59: DOI10.1016/j.nedt.2017.1009.1007.

438 Lobchuk, M., Halas, G., West, C., Harder, N., Tursunova, Z., and C. Ramraj (2016). Development of a
 439 novel empathy-related video-feedback intervention to improve empathic accuracy of nursing
 440 students: A pilot study. *Nurse Education Today* 46: 86-93.

441 Lown, B., Rosen, J., and J. Marttila (2011). An agenda for improving compassionate care: a survey
 442 shows about half of patients say such care is missing. *Health Affairs* 30: 1772–1778.

443 McKenna L, Brown, B. M., Williams, B., Molloy, A., Lewis, B., and L. Molloy (2012). Levels of empathy
 444 in undergraduate nursing students. *International Journal of Nursing Practice*. 18:246-251.

445 McMillan, L. R. and D. M. Shannon (2011). Psychometric analysis of the JSPE nursing student version
 446 R: comparison of senior BSN students and medical students attitudes toward empathy in patient
 447 care. *ISRN Nursing*: 1-7.

448 Mennenga, H. A., Bassett, S., and L. Pasquariello (2016). Empathy development through case study
 449 and simulation. *Nurse Educator* 41: 139-142.

450 Mete, S. (2007). The empathetic tendencies and skills of nursing students. *Social Behavior and*
 451 *Personality* 35: 1181-1188.

452 Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and The PRISMA Group (2009). Preferred
 453 Reporting Items for Systematic reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 6:
 454 e1000097. doi:10.1371/journal.pmed1000097 6(6): e1000097.

455 Nosek, M., Gifford, E., and B. Kober (2014). Nonviolent Communication (NVC) training increases
 456 empathy in baccalaureate nursing students: A mixed method study. *Journal of Nursing Education*
 457 *and Practice* 4 (10): 1.

458 Ozcan, N. K., Bilgin, H., and N. Eracar (2011). The use of expressive methods for developing empathic
 459 skills. *Issues in Mental Health Nursing* 32: 131-136.

460 Polit, D., and C. Tatano-Beck (2012). *Nursing research: generating and assessing evidence for*
 461 *nursing practice*. Sydney: Wolters Kluwer.

462 Reed D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., and S. M. Wright (2007). Association
 463 between funding and quality of published medical education research. *JAMA* 298: 1002-1009.

464 Reed, D. A., Beckman, T. J., and S. M. Wright (2009). An assessment of the methodologic quality of
 465 medical education research studies published in *The American Journal of Surgery*. *American Journal*
 466 *of Surgery* 198: 442-444.

467 Scott, H. (2011). *Empathy in healthcare settings*. Goldsmiths, University of London, Goldsmiths
 468 *Research Online*.

469 Sheehan, C. A., Perrin, K. O., Potter, M. L., Kazanowski, M. K., and L. A. Bennett (2013). Engendering
 470 empathy in baccalaureate nursing students. *International Journal of Caring Sciences* 6: 456-464.

471 Strekalova, Y. A., Krieger, J. L., Kleinheksel, A. J., and A. Kotranza (2017). Empathic communication in
 472 virtual education for nursing students. *Nurse Educator* 42: 18-22.

473 Teding van Berkhout, E., and J. M. Malouff (2016). The efficacy of empathy training: A meta-analysis
 474 of randomized controlled trials. *Journal of Counseling Psychology* 63: 32.

475 Trzeciak, S., Roberts, B.W., and A. J. Mazzairelli (2017). Compassionomics: Hypothesis and
 476 experimental approach. *Medical Hypotheses* 107: 92-97.

477 Ward, J. (2016). The Empathy Enigma: does it still exist? comparison of empathy using students and
 478 standardized actors. *Nurse Educator* 41: 134-138.

479 Ward, J., Cody, J., Schaal, M., and M. Hojat (2012). The Empathy Enigma: an empirical study of
 480 decline in empathy among undergraduate nursing students. *Journal of Professional Nursing* 28: 34-
 481 40.

482 Ward, J., Schaal, M., Sullivan, J., and M. E. Bowen (2009). Reliability and validity of the Jefferson
 483 Scale of Empathy in undergraduate nursing students. *Journal of Nursing Management* 17: 3.

484 Webster, D. (2010). Promoting empathy through a creative reflective teaching strategy: a mixed-
 485 method study. *Journal of Nursing Education* 49: 87-94.

486 West, C. P., Huschka, M. M., Novotny, P. J., Sloan, J. A., Kolars, J. C., Habermann, T. M., and T. D.
 487 Shanafelt (2006). Association of perceived medical errors with resident distress and empathy: a
 488 prospective longitudinal study. *JAMA* 296: 1071-1078.

489 Wikstrom, B. M. (2003). A picture of a work of art as an empathy teaching strategy in nurse
 490 education complementary to theoretical knowledge. *Journal of Professional Nursing* 19: 49-54.

491 Yu, J. and M. Kirk (2009). Evaluation of empathy measurement tools in nursing: systematic review.
 492 *Journal of Advanced Nursing* 65: 1790-1806.