Review

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

Tracy Levett-Jones[a,*], Robyn Cant[b], Samuel Lapkin[c]

[a] Nursing Education, University of Technology Sydney, Ultimo, NSW 2007, Australia
[b] Faculty of Health (Nursing), University of Technology Sydney, Ultimo, NSW 2007, Australia
[c] Faculty of Science Medicine & Health, School of Nursing, University of Wollongong, South Western Sydney Campus, Liverpool, NSW 2170, Australia

ARTICLE INFO

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.

1. Introduction

Empathy is a multidimensional construct and various definitions of the term are evident in the literature. Generally speaking, empathy involves the cognitive ability to comprehend what another person is feeling, an emotional resonance with those feelings, and the willingness to respond appropriately to the person's needs (Hatfield et al., 2011). In healthcare, empathy is considered a basic component of therapeutic relationships and a critical factor in patients' definitions of quality care (Hojat et al., 2013). It is suggested that over 200 studies have demonstrated the positive impact of empathic healthcare interactions on patient outcomes (Trzeciak et al., 2017). Empathetic encounters with healthcare professionals result in: decreased levels of depression, anxiety, distress; and increased levels of emotional wellbeing, satisfaction and adherence to treatment regimens (Hojat et al., 2013). Examples of physiological consequences of empathic healthcare interactions include: reduced rates of infection, improved wound healing, higher cancer survival rates, and a reduction in diabetes complications, and pain (Scott, 2011; Trzeciak et al., 2017). However, despite the increasing recognition of the impact of empathy on patient outcomes, there is compelling research indicating that contemporary healthcare is besieged by a generalised lack of empathy (Lown et al., 2011), which

---

* Corresponding author at: University of Technology Sydney, Ultimo, NSW 2007, Australia.
  E-mail addresses: Tracy.Levett-Jones@uts.edu.au (T. Levett-Jones), Robyn.Cant@uts.edu.au (R. Cant), slapkin@uow.edu.au (S. Lapkin).
  @Prof_TLJ (T. Levett-Jones), @DrLapkin (S. Lapkin).

can manifest itself as apathy, indifference, callousness, cruelty, and dehumanisation; and result in increased risk of harm to patients (West et al., 2006).

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

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

2. Review Methods

2.1. 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 and Green, 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 et al., 2009).

2.2. 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.

2.3. Search Strategy

A three-step search strategy was utilized 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 the conduct of searches in CINAHL Plus, SCOPUS, PubMed (Ovid Medline), and Education Resources Information Center (ERIC). We used a systematic strategy based on derivatives of 'students, nursing' (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.

2.4. 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).

2.5. 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 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 et al., 2007, Cook and Reed, 2015). In recent nursing research, interrater agreement using the MERSQI across 26 studies was high (ICC = 0.988) (Cant et al., 2018).

2.6. 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.

2.7. 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. Standardised effect sizes for the primary outcome (change in empathy level) were computed from study instrument

Box 1
Example of search strategy conducted in PubMed.

```
Search (((((((students nursing[MeSH Terms]) AND education[Title/Abstract]) AND empathy[Title/Abstract]) AND undergraduate[Title/
Abstract]) OR prelicensure[Title/Abstract]) AND effectiveness[Title/Abstract]) AND English[Language]) AND ("2000/01/01"[Date -
Publication] : "2018/05/30"[Date - Publication])
```

Table 1

Characteristics of included studies and the educational outcomes for empathy.

Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/ recommendations
Bas-Sarmiento et al. 2017 (Spain)	**Aim:** to measure impact of a set of educational interventions on empathy levels. **Design:** quasi-experimental single group study. **Data collection:** pre, post and at one month.	n = 47 second-year undergraduate nursing students from a population of N = 48	**Simulation** – role-plays, behaviour assays, reflective writing, and a flipped classroom approach. **Duration:** > 20 h of training as part of an interpersonal communication skills unit.	**Reynolds Empathy Scale** (RES) (Spanish version) used by students to self-evaluate their performance. **Consultation and Relational Empathy Measure** (CREM) used by patients to evaluate students' behaviour during simulation. **Carkhuff Scale** (similar to OSCE) used by three independent external observers to assess student's behaviour.	Jefferson Scale of Empathy-Student (JSE-S): V & R not given. **RES:** V = high concurrent validity 0.90; R: strong test-retest reliability alpha = 0.90. **CARE** (Spanish version): R: alpha 0.93.	Improved mean scores for all variables; significant for 8 of 9 measures including RES and CREM ($p \leq 0.001$); improved objective assessment (Carkhuff): effect size- 3 observers, pre-test to immediate post-test d = 3.29.	12.5	The multidimensional experiential intervention resulted in a significant improvement in both self-reported and objectively assessed empathy scores.
Bedoe & Murphy 2004 (USA)	**Aim:** to measure impact of mindfulness-based stress reduction (MBSR) course on students' stress and empathy levels. **Design:** quasi-experimental single group study. **Data collection:** pre-test post-test.	n = 16 baccalaureate nursing students from a population of N = 48	**MBSR** course consisting of relaxation, meditation, yoga and gentle exercise. **Duration:** 8 weeks – weekly 2-hour face-to-face sessions; and 5 × 30 min individual sessions per week using guided meditation audiotapes; completion of reflective journals.	**Derogatis Stress Profile** (DSP) **Interpersonal Reactivity Index** (IRI)	**DSP:** V = Not given. R: overall Cronbach's alpha = 0.90 **IRI:** V = Not given. R: Cronbach alpha > 0.73.	DSP significantly reduced (p > 0.05); IRI scores were moderately high pre-test, no significant change at post-test.	9.0	The MBSR course influenced anxiety levels but not empathy levels.
Cunico et al., 2012 (Italy)	**Aim:** to measure changes in empathy levels over a 3-year education program. **Design:** quasi-experimental two-group cohort study. **Data collection:** pre-test (year 1) post-test (year 3).	n = 103 first-year students (in IG), of 150 recruited, 62 (41%) completed third yr survey and of 72 in CG, 41 (58%) completed third year survey.	**Experiential learning activities** - IG - tutor-led seminars and small group laboratories to develop communicative and empathic abilities (teaching included films, guided examination of nurse-patient/ communication exercises, and role plays). **Duration:** 21 h over three years. CG – usual education.	Italian version of the **Balanced Emotional Empathy Scale** (BEES)	BEES: V: dimensions are described in the paper, scale data not given. R: not given.	Although both IG and CG improved a statistical increase in empathy was only seen in IG mean total scores (p = 0.047). Females showed greater improvement than males; empathy decreased in males between T_0 and T_3.	10.5	Empathy is a skill that may be developed over time. Men and women have different empathic traits and may show disparate empathic tendencies.
							9.5	

Table 1 (*continued*)

Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/ recommendations
Everson et al. 2015 (Australia)	**Aim:** to measure the impact of a 3D e-simulation on nursing students' empathy towards patients from culturally and linguistically diverse backgrounds. **Design:** quasi-experimental one-group study. **Data collection:** pre-test post-test.	n = 460 second-year undergraduate nursing students from a population of N = 530 (RR 87%)	**Point-of-view simulation** - Students took on the role of a CALD patient while viewing a 10-minute 3D video of an unfolding scene in a hospital ward of a developing country. **Duration:** 60 min including briefing, e-simulation and debrief.	**Modified Kiersma-Chen Empathy Scale** (MKCES).	MKCES: V: verified during development of modified scale during testing - Cronbach's alpha: pre-test 0.73, post-test 0.79.	MKCES scores significantly improved (p < 0.001); mean scores for 'Perspective Taking' and 'Valuing Affective Empathy' subscales also increased significantly post simulation.		Point-of-view 3D e-simulations can have a positive impact on nursing students' empathy towards CALD patients.
Everson et al. 2018 (Australia)	**Aim:** to measure the impact of a 3D e-simulation on nursing students' empathic concern for patients from culturally and linguistically diverse backgrounds. **Design:** quasi-experimental one-group study. **Data collection:** pre-test post-test.	n = 460 second-year undergraduate nursing students from a population of N = 530 (RR 87%)	**Point-of-view simulation** - students took on the role of a CALD patient while viewing a 10-minute 3D video of an unfolding scene in a hospital ward of a developing country. **Duration:** 60 min including briefing, e-simulation and debrief.	**Empathic Concern subscale** (EC) of the Emotional Response Questionnaire (ERQ).	V: previously demonstrated. R: a one factor model demonstrated good internal consistency (α 0.87) on pre-test data and acceptable internal consistency (α 0.78) on post-test data.	EC scores significantly increased (p < 0.001).	10.5	Point-of-view 3D e-simulations can have a positive impact on nursing students' empathic concern for CALD patients.
Finch 2016 (PhD thesis) (USA)	**Aim:** to measure the impact of a final semester capstone immersive clinical experience on nursing students' emotional intelligence (EI) and empathy. **Design:** quasi-experimental cohort study. **Data collection:** pre-test post-test.	n = 44 from a population of 58 senior BSN students.	**Clinical placement** – undertaken in community health care facilities with students supported by preceptors. **Duration:** 150 h	**Mayer-Salovey-Caruso Emotional Intelligence Test V2.0 (MSCEIT)** which included an empathy subscale.	V: 0.93 for general, and 0.91 for expert scoring. R (inter-rater) r = 0.86. V & R data described for each construct.	EI scores significantly increased (p ≤ 0.001). No difference in empathy subscale scores (p = 0.721). Gender and prior health care experience had no impact of EI or empathy scores.	10.5	The 150-h clinical placement positively influenced EI but not empathy scores, a noteworthy finding.
Fleming et al. 2015 (USA)	**Aim:** to measure the impact of an Appreciating Diversity workshop on nursing (and dental) students' ethnocultural empathy. **Design:** quasi-experimental pilot study. **Data collection:** pre, post and at one month.	First year nursing (n = 40) and dental students (n = 42) (83% RR).	**Appreciating Diversity workshop** - guided pair and group activities focused on cross-relational skills. **Duration:** one-day workshop	**Scale of Ethnocultural Empathy** (SEE) (p < 0.05)	V = good internal consistency reliability. R: good test-retest reliability and convergent reliability. In current study Cronbach's alpha = 0.89.	SEE scores significantly increased (p < 0.05) and these gains were maintained at 1-month follow-up. (nursing students scores were not reported separately)	10.5	Results from this pilot study support the need for controlled trials designed to improve students' understanding of racial groups discordant from their own and determine if empathy and communication are improved.
Geckil 2017 (Turkey)	**Aim:** to measure the impact of experiential empathy activities on nursing students' attitudes	n = 116 nursing students (53 IG; 63 CG) from two universities.	**Experiential learning activities** - participants role played mobility and everyday tasks of a visually	**Attitudes Towards Disabled Persons** Scale (ATDP) Turkish version.	V = the scale has been found valid in Turkish studies. R = Cronbach's alpha 0.62 and 0.76.	ATDP scores of EG significantly higher than CG (p = 0.002)	10.5	Experiential learning activities designed to enhance empathy can improve students' attitudes *(continued on next page)*

Table 1 (*continued*)

Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/ recommendations
	towards people with a disability. **Design:** case-control study. **Data collection:** pre-test post-test at six months.		impaired person that aimed to enhance students' empathy followed by a reflective activity. **Duration:** part of first-year unit duration not given. CG – usual education.			and were maintained at six months (p = 0.025).		towards people with a disability.
Haley et al. 2017 (USA)	**Aim:** to measure the impact of a simulation on nursing students' empathy, active listening skills and self-awareness. **Design:** RCT. **Data collection:** pre-test post-test and at one month.	n = 53 from a population of 63 junior and senior pre-licensure baccalaureate and associate degree students (IG = 26; CG = 24).	**Simulation** - students completed the NLN-Advancing Care Excellence for Seniors Millie Larsen case study (consisting of audio narrative briefing, and manikin-based simulation). **Duration:** 20 min simulation followed by debrief.	**Active Empathic Listening Scale (AEL)** **Self-Consciousness Scaled-Revised (SCS-R)** **Kiersma-Chen Empathy Scale (KCES)** **Patient-Practitioner Orientation Scale (PPOS)** competed pre- and post, and after 1 month.	AEL: V: Described during development. R: Cronbach's alpha in present study = 0.89. SCS: V: construct validity was supported by factor analysis with loadings ≥ 0.45. R: Cronbach's alpha in present study = 0.91. KCES: V: construct and criterion validity supported during development by factor analysis. R: Cronbach's alpha during development = 0.89; in present study 0.87. PPOS: V: verified for physicians. R: Cronbach's alpha in present study = 0.92.	AEL and PPOS scores in the IG compared with the CG differed significantly at baseline, post intervention and at 1-month follow-up (p < 0.001). KCES and SCS-R scores were significantly different between groups the only at 1-month follow-up (p < 0.001).	12.5	NLN- Advancing Care Excellence for Seniors Millie Larsen case study was an effective education strategy for improving empathy (and the other three constructs).
Henry & Ozier 2011 (USA)	**Aim:** to measure the impact of a simulation-based ageing game on nursing students' knowledge, attitudes and empathy. **Design:** experimental two-group **Data collection:** pre-test post-test at beginning and end of semester.	N = 127 nursing (n = 73 of 76) and nutrition students (n = 54 of 54) randomly assigned to IG and CGs. n = 89 (70%) completed the study.	**Simulation** - role-playing the 'Ageing Game' (five stations). Students completed the NLN-Advancing Care Excellence for Seniors Millie Larsen case study (consisting of audio narrative briefing, and manikin-based simulation. **Duration:** 90 min including debrief. CG - instructor-led presentation and discussion.	**Anxiety about Aging Scale (AAS)** **Aging Semantic Differential (ASD)**, **Empathy:** four questions from the Maxwell and Sullivan Survey.	ASS: V & R not given. ASD: V: not given R: pre-test Cronbach alpha = 0.91. Empathetic response set: Cronbach alpha = 0.69 pre-test; = 0.70 post-test.	No significant change was seen in empathetic question response scores. A significant decrease in anxiety was seen in lower total AAS scores and nursing majors in particular.	10.5	While the game positively influenced anxiety levels, it had limited impact on empathy for either nursing or nutrition students.
Lee et al. 2018 (Taiwan)	**Aim:** to measure the impact of a simulation-based ageing activity on nursing students' empathy scores. **Design:** quasi-experimental two-group cohort study. **Data collection:** pre-test post-test.	n = 103 2nd-year nursing students from two BSN programs. (IG = 48; CG = 55).	**Simulation** - role plays, guided reflection and OSCE. **Duration:** over a 4-month period. CG - usual classroom lecture format.	**The Jefferson Scale of Empathy–Health Profession–Student (JSE-S)** (Chinese version **Objective Structured Clinical Examination (OSCE)** with empathy rubric.	JSE-HP-S: R & V previously established. Cronbach alpha for the Chinese version used in this study = 0.89. OSCE: checklist created by authors (CVI = 0.88). Cronbach alphas = 0.82 and 0.76 for teachers and standardised patients, respectively.	JSE-HP-S scores increased significantly in IG but no significant difference with CG. OSCE examiners and standardised patients gave significantly higher empathy scores to the IG than the CG (both improved:	12.5	This study attests to the value of objective observational methods of assessing empathy as well as the use of standardised patients.

Table 1 (*continued*)

Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/ recommendations
						$p \leq 0.001$). The JSE-HP-S were not correlated with OSCE scores.		
Levett-Jones et al. 2017 (Australia)	**Aim:** to measure the impact of a point-of-view simulation on nursing students' empathy towards people with an Acquired Brain Injury (ABI). **Design:** quasi-experimental two-group study. **Data collection:** pre-test post-test.	n = 390 second-year undergraduate nursing students from a population of N = 488 (RR 80%). Simulated patients n = 202, rehabilitation nurses n = 188.	**Point-of-view simulation** - students took on the role of either a simulated patient or a rehabilitation patient. The simulated patients wore a hemiparesis suit that replicated the experience of dysphasia, hemianopia and hemiparesis. **Duration:** 60 min including briefing, simulation and debrief.	**Comprehensive State Empathy Scale (CSES)**	CSES R: psychometric testing in current study showed good internal consistency, Cronbach alpha = 0.95. Confirmatory factor analysis identified 6 factors. Cronbach alpha for subscales were satisfactory: shared affect α 0.86; distress α 0.93; empathic concern α 0.87; helping motivation α 0.84; empathic imagination α 0.82; cognitive empathy α 0.93.	CSES scores significantly improved for both groups $p < 0.001$; the rehabilitation nurse group had significantly higher CSES scores than the simulated patients ($p < 0.05$).	12.0	The results from this study underscore the impact of point-of-view simulations to positively impact learners' empathy towards vulnerable patient groups.
Lobchuk 2016 (Canada)	**Aim:** to measure the impact of a video-based intervention on nursing students' perspective-taking and empathic accuracy. **Design:** pilot quasi-experimental study. **Data collection:** synchronous	n = 16, 3rd year undergraduate nursing students (8 pairs of nursing students) and carers.	**Perspective taking exercise** – students (working in pairs) participated in perspective-taking instructional and practice sessions, followed by a 10-minute video-recorded dialogue with carers followed by a video-tagging task. **Duration:** approximately 4 weeks allowing for practice: coaching 30 min, 10 min videoed communication plus video tagging.	Students and carers separately engaged in a video-tagging exercise on thoughts and feelings about their health risk behaviours when dialoguing with the student. Using Studiocode™ The computer recorded the time of each video-tagged instance, the self-reported 'thought' or 'feeling' as well as the specific content.	Both students and carers were trained in pre-briefing sessions. Performance measurements were obtained from the computer software 'Studiocode' with performances rated by faculty. Carers gave feedback.	Carers felt that students were 'non-judgmental', inquisitive, and helped them to 'gain new insights'. There was substantial Fleiss Kappa agreement among four raters across five dyads and 67 tagged instances.	11.0	This novel approach facilitated the development of perceptive taking and empathic accuracy; however, the small sample size limited the extent to which results can be extrapolated to larger groups.
Menenga 2016 (USA)	**Aim:** to measure the impact of a e-simulation nursing students' empathy levels and to identify differences in the impact of exposure to a single patient case, versus varied cases. **Design:** experimental two-group study. **Data collection:** pre-test and post-test.	n = 127 senior level nursing students from a population of N = 128, enrolled in a mental health unit (response 99%). Randomized into IG and CG.	**E-simulation** - IG completed an e-simulation (comprised of a picture of the patient with an audio narration) based on two scenarios of one patient; the CG's e-simulation included *two different* patients. **Duration:** two case studies and two e-simulations completed in class time-duration not stated.	**Jefferson Scale of Empathy for Health Professional Students (JSE-HPS).**	JSE-HPS: V: previously psychometrically tested/ validated R: reliability coefficient of 0.77 at pre-test and 0.80 at post-test in this study.	No significant difference in JSE-HPS ANCOVA scores between the IG compared with the CG at pre-test (p = 0.76) and post-test (p = 0.58).	12.0	The difference in the learning experiences was too subtle to evoke a difference in empathy scores between groups.
Mete, 2007 (Turkey)	**Aim:** to measure the impact of Problem-Based Learning (PBL) on nursing students' empathic tendencies and skills.	Stage 1. n = 20. Stage 2. n = 192	**Problem-Based Learning** (PBL) - as a routine teaching activity.	Stage 1. **Scale of Empathic Tendency** (ET) and **Scale of Empathic Skill** (ES) each year for the 4 years of enrolment. Stage 2. ET and ES cross-	ET: V: Cronbach alpha = 0.86, in the current study over 5 iterations. ES: V & R reported during	Stage 1: mean ET scores increased year 1–4 (p = 0.03). Stage 2: mean ES scores significantly increased	9.5	PBL has the potential to significantly impact nursing students' empathy levels, however, this is one of the first studies to measure

Table 1 (*continued*)

Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/ recommendations
				sectional survey completed at end of either year 1, 2, 3, or 4.	development and by test-retest 0.91.	year 1–4 (p ≤ 0.05) but changes in mean ET were not significant (p > 0.05).		empathy as an outcome of this well used pedagogical approach.
Nosek 2014 (USA)	**Design:** quasi-experimental single group mixed methods. **Data collection:** pre-test post-test. **Aim:** to measure the impact of a Nonviolent Communication intervention on nursing students' empathy levels.	n = 55 first year baccalaureate nursing students undertaking a psychiatric unit (attrition 24%)	**Nonviolent Communication** (NVC) teaching intervention – including compassionate communication using role-plays. **Duration:** 2 × 1.75-hour sessions.	**Interpersonal Reactivity Index** (IRI) to measure empathy.	IRI: V &R: well documented during development and Cronbach alpha in present study > 0.70.	IRI scores significantly increased (p = 0.037) post training. Qualitative findings revealed positive impacts of training.	10.5	Incorporating NVC into curricula is an effective strategy for enhancing communication skills and empathy levels in students.
Ozcan et al. 2011 (Turkey)	**Aim:** to measure the impact of expressive arts on nursing students' empathy skills. **Design:** quasi-experimental single group study. **Data collection:** pre-test, post-test.	n = 48, 3rd year and senior students in five groups.	**Expressive art/ psychodrama** - focusing on knowing oneself, developing self-awareness; and relationships with 'others'. **Duration:** 2 h × 12 weeks.	**The Scale of Empathic Skill** (SES) was administered before and after the group studies. (6 scenarios are presented; each is related to people with a different problem. For each problem, there are 12 statements to be told to the person having the problem and responses are recorded).	SES: R: test-retest reliability 0.83.	SES scores increased significantly post intervention (p = 0.001).	9.0	The expressive art/ psychodrama approach enhanced participants' gaining awareness of effective interpersonal dynamics with patients.
Ozcan et al. 2012 (Turkey)	**Aim:** to measure the impact of a lecture program on structured lecture course on nursing students' empathy and communication skills. **Design:** quasi-experimental two group study. **Data collection:** pre-test, post-test.	n = 257 first year students) from a medical (n = 143) and a nursing school (n = 83) in one Turkish university were recruited (88% completed the study).	**Program of didactic lectures** – focused on empathy and communication skills. **Duration:** 2 h × 5 weeks.	**Empathic Communication Skills Scale** (ECSS) **Empathic Tendency Scale** (ETS).	ECSS: V & R are reported- the retest correlation coefficient was r = 0.91 and Cronbach alpha 0.83. ETS: V & R are reported- Cronbach alpha 0.88.	ECSS and ETS scores significantly improved for both groups (p < 0.05). (Nursing students scores were not reported separately).	11.5	Didactic and targeted lectures have the potential to improve empathy levels.
Sheehan et al. 2013 (USA)	**Aim:** to measure the impact of an elective course about the dimensions of human suffering on nursing students' empathy levels. **Design:** quasi-experimental single group study. **Data collection:** pre-test, post-test.	n = 99 senior baccalaureate junior and senior nursing students from a population of N = 125 (response rate 79%).	**Elective human suffering course** - included topics related to palliative care and utilizing both traditional and non-traditional teaching strategies (details not given). **Duration:** course unit over 4 months.	**Jefferson Scale of Empathy, Nursing Student Version** (JSE-Nurs)	V & R: large studies have demonstrated acceptable internal consistency reliability coefficient (0.77–0.89).	JSE-Nurs scores significantly increased (p ≤ 0.001) with positive findings replicated in different student groups over 5 year period. Collectively, students scored seven points higher at course completion (p < 0.001).	10.5	The paper provides sample exercises from the Human Suffering course that will be of benefit to educators and researchers.
Streklova 2017 (USA)	**Aim:** to measure the impact of a virtual patient simulation on nursing	n = 343 undergraduate nursing students attended a	**Virtual patient simulation** using message design logistics - health	A **codebook** was developed to assess which message design logic was used in the	V & R: purposely developed tool. Intercoder reliability ranged from 0.70 and 0.96.	Students were largely conventional in their empathic statements	9.0	The results suggest the feasibility and utility of novel approaches such as

Table 1 (continued)

Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/recommendations
	students' expressive empathy skills. **Design:** cross-sectional single group analysis. **Data collection:** pre-test, post-test.	standardised health assessment course at nursing schools in eight US states.	history information in a pre-programmed virtual patient. Students typed questions to obtain health history information. The interface enabled questioning or provision of empathetic statements by choosing one of the options: 'ask', 'emphasize', or 'educate'. **Duration:** one home assignment in second week.	statements that nursing students self-identified as empathic. Statements were coded as expressive, conventional, or rhetorical, assigning values of 1, 2, or 3 respectively.		(mean score 2.00 of 3.00). Students encountered up to 9 information disclosures that warranted the expression of empathy, but they recognised only 33.5%. of them. Empathic statements in response to pain complaints had relatively high scores (mean 2.23; SD 0.61) while other scores were lower.		using virtual patients to enhance empathy skills.
Ward et al. 2010 (USA)	**Aim:** to measure changes in nursing students' empathy levels over one academic year. **Design:** quasi-experimental single group study. **Data collection:** pre-test, post-test.	n = 214 senior undergraduate nursing students from a population of N = 333 (sample 64%).	**Traditional learning;** students completed the Jefferson Scale of Empathy at the beginning and and at end of 2006–2007 academic year.	**Jefferson Scale of Empathy, Nursing Student Version** (JSE-Nurs)	JSE-Nurs: V & R: reported during instrument development.	JSE-Nurs scores did not significantly change overall, however, there was a significant change in students exposed to more patient encounters during study period (p = 0.01).	9.5	Further research is necessary to identify why and to what extent empathy scores decline in relation to increasing patient encounters.
Ward et al. 2016 (USA)	**Aim:** to measure the impact of a standardised patient simulation on nursing students' empathy levels. **Design:** quasi-experimental two-phase mixed methods. **Data collection:** pre-test post-test and at one year.	n = 146 senior nursing students (n = 98 BSN, n = 72 fast track 12-month program (FACT)).	**Simulation** - using a standardised patients and family members. **Duration:** 15-min simulation followed by debrief.	**Jefferson Scale of Empathy-Health Providers** (JSE-HPs) standardised patients and family members scored students' empathy skills using a **checklist.**	JSE-HPs: V & R: demonstrated (Cronbach alpha = 0.78). Checklist: V & R not reported.	JSE-HPs scores did not change overall but FACT students' scores increased: JSE-HPs scores (p < 0.04). Self-reported empathy was associated with SPs' assessment of student empathy (group 1 (p < 0.01) and group 2 p < 0.02).	11.0	Standardised patients offer and important opportunity to improve empathy scores however, appropriate scenarios and debriefing approaches are required.
Webster, 2010 (USA)	**Aim:** to measure the impact of a community psychosocial program and a creative writing reflective assignment on nursing students' empathy levels. **Design:** quasi-experimental two-group study. **Data collection:** pre-test post-test.	n = 73 baccalaureate nursing students enrolled in a psychiatric clinical course. (IG = 29; CG = 44).	**Community psychosocial program and reflection** - IG group completed a 4-week community psychosocial program and a creative reflective assignment to represent the life experience of a client with mental illness; CG completed a 14-week clinical placement.	**The Interpersonal Reactivity Index** (IRI); Data were also collected from interviews (n = 14), field observation, and students' creative reflective assignments.	IRI: V & R: internal reliability 0.71 to 0.77; test-retest reliability 0.62 to 0.71.	IRI scores did not differ significantly between groups but positive and negative trends were noted in sub-scale factors (Compassion Satisfaction; Compassion Fatigue.	10.0	Vastly different approaches and duration of intervention between IG and CG in this study may have undermined the study outcomes.
Wikstrom 2003 (Sweden)	**Aim:** to measure the impact of an art-based intervention on nursing students' engagement in learning about empathy.	N = 144 first-year nursing students (IG = 72; CG = 72).	**Art-based intervention** - IG); students engaged in small group discussions over 1 h using a reproduction of Edvard	Engagement in learning was measured with the semi-structured **Wheel Questionnaire** which assesses parameters of motivation,	V & R: no data given.	IG: scores significantly improved compared with CG. (IG Motivation 9.6; 12.0 respectively; Structure 0.2, 0.4;	10.0	Findings suggest Edvard Munch's painting could be effectively used as a stimulus for teaching

Table 1 (*continued*)

| Study	Design/aim	Student/sample	Intervention	Evaluation/measures	Validity (V) and reliability (R) of measures	Results	Study quality (MERSQI score/18)	Outcomes/ recommendations
	Design: randomized study. **Data collection**: post-test.		Munch's painting, 'The Sick Girl' (representing empathy). CG: discussed the same issues without the use of a picture.	structure, and emotional investment. The instrument was used as a teaching tool and students individually wrote their responses/scores based on their feelings during the session.		Emotional Investment 1.2, 1.7). IG students were also more engaged in learning about empathy parameters of structure, motivation, and emotional investment.		empathy in nursing programs.
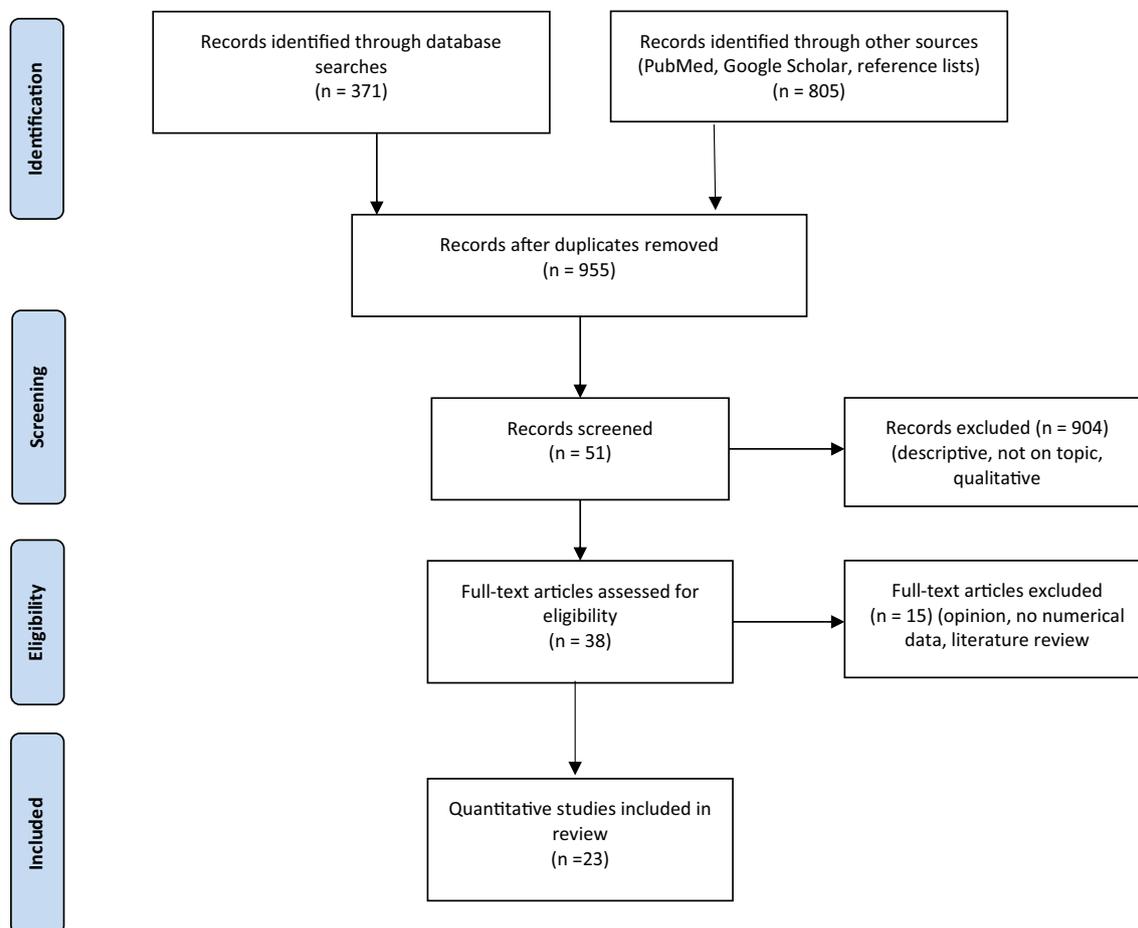
Fig. 1. PRISMA flow chart.

scores and standard deviations using an online calculator (https://www.uccs.edu/lbecker/). They were reported using Cohen's *d* statistic presented as the effect size *r* (a Pearson correlation co-efficient). The *d* is the standardised mean difference between two groups and *r* describes the proportion of the variance that is explained (with possible values between zero and 1). A small effect for the *r* statistic is generally interpreted as 0.10, a medium effect 0.30, and a large effect 0.50 or greater. In this study the change in mean empathy scores was considered to be negligible when *r* was < 0.2, as this indicates that the means of two groups differ by < 0.2 standard deviations and the difference, even when statistically significant, is trivial (Polit and Tatano-Beck, 2012, p. 478). An effect size of *r* = 0.20 or more was regarded as a valid measure of change in empathy that was both practical and likely to be clinically meaningful.

3. Results

3.1. Study Selection

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

3.2. 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 et al.

(2011), Ward et al. (2012), Sheehan et al. (2013), Nosek et al. (2014), Everson et al. (2015), Fleming et al. (2015), Ward (2016), Bas-Sarmiento et al. (2017), and Levett-Jones et al. (2017). Two further studies used post-test only designs: Lobchuk et al. (2016) and Strekalova et al. (2017).

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 et al. (2011), Cunico et al. (2012), Mennenga et al. (2016), Geçkil et al. (2017), Haley et al. (2017), and Lee et al. (2018). Four of these studies randomized students into an intervention or control group: Henry et al. (2011), Mennenga et al. (2016), Geçkil et al. (2017), and Haley et al. (2017). Two longitudinal studies re-assessed empathy levels at four weeks (Haley et al., 2017) or six months (Geçkil et al., 2017) after the intervention.

3.3. 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.

3.4. Types of Outcome Measures

In all, the included studies reported the use of 21 different empathy instruments, including translations of several English-language

Table 2
Change in empathy scores identified in single group studies.

Study first author	Sample n	Instrument/s	Intervention group		Statistically significant change in empathy	Outcome
			Pre-test Mean (SD)	Post-test Mean (SD)	Yes/no	Effect size (r)
Bas-Sarmiento 2017	47	Reynolds Empathy Scale (RES) (Spanish)	66.81 (14.27)	82.57 (13.39)	Yes	0.49
		Consultation and Relational Empathy Measure (CREM)	22.56 (2.20)	28.50 (6.60)		0.52
Bedoe 2004	16	Derogatis Stress Profile (DSP)		DSP increased significantly.		N/A
		Interpersonal Reactivity Index (IRI).			No	No data
Everson 2015	460	Modified KCES:			Yes	
		- Perspective taking	47.86 (4.64)	49.24 (5.18)		0.14
		- Valuing	2853 (375)	2959 (414)		0.13
		- Affective empathy	1934 (167)	1965 (176)		0.09
Everson 2018	460	Empathic Concern subscale of the IRI	5.57 (1.04)	6.10 (0.95)	Yes	0.26
Finch 2016	44	Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) (includes empathy subscale)	98.71 (17.10)	97.29 (15.24)	No	0.04
Fleming 2015	82	Scale of Ethnocultural Empathy (SEE)			Yes- no data reported.	
		- Empathetic feeling & expression	4.54 (0.07)	4.66 (0.09)		T_0-$T_1$0.60
		- Empathetic perspective taking	3.49 (0.10)	3.84 (0.10)		0.87
		- Acceptance of cultural differences	4.75 (0.08)	4.88 (0.10)		0.58
		- Empathetic awareness	4.44 (0.09)	4.74 (0.09)		0.86
						[M] r = 0.73
Levett-Jones 2017	390	Comprehensive State Empathy Scale (CSES)	3.38 (0.61)	3.75 (0.66)	Yes	0.28
Lobchuk 2016	16	Video assessment		Student empathetic accuracy 49%, range 31–71%.	N/A-no data reported	N/A
Mete 2007	20	Empathetic Tendency/	69.80	75.40	Yes	N/A-no data
		Empathetic Skill	147.80	186.45		N/A-no data
Nosek 2014	55	Interpersonal Reactivity Index (IRI)	69.1 (12.3)	71.4 (12.3)	Yes	0.09
Ozcan 2011	48	Scale of Empathetic Skill	127.97 (21.26)	138.87 (20.40)	Yes	0.25
Sheehan 2013	99	Jefferson Scale of Empathy, NS Vs R (JSE-Nurs)	116.95 (9.8)	123.97 (7.78)	Yes	0.37
Streklova 2017	343	Quantified empathy assessment statements	N/A	2.01 (0.61) (of possible 3.00)	Yes	N/A-no data available
Ward 2010	214	Jefferson Scale of Empathy for Nursing Students (JSE-N-S)	114.6 (11.8)	112.7 (12.1)	Yes	0.16
Ward 2016	146	Jefferson Scale of Empathy (JSE-HPs)	112.3 (13.0)	114.3 (11.9)	No	0.08

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) also reporting reliability outcomes from their current study. Details of the empathy instruments used are provided in 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 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 et al. (2018), in addition to participant surveys, and in the study by Bas-Sarmiento et al. (2017). Ward (2016) developed a checklist so that standardised patients and simulated family members could rate students' empathy.

3.5. Methodological Quality Components

The studies were assessed for methodological quality. Following MERSQI assessment, one study (Chen et al., 2015) was removed because of a low score of < 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 standardised patient, or computer-based ratings of performance). Only four 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 criterion limited quality scores.

3.6. 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 studies identifying a significant change in overall or 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,

Table 3
Change in empathy scores identified in case-control studies and RCTs.

Study first author	Sample n	Instrument/s	Intervention group		Control group		Statistically significant change in empathy	Group outcome	Net effect: gain over control group
			Pre-test Mean (SD)	Post-test Mean (SD)	Pre-test Mean (SD)	Post-test Mean (SD)	Yes/no	Effect size (r)	(r)
Cunico 2012	104	BEE (Spanish version)	30.19 (17.2)	37.87 (16.5) (T_0-T_3: p = 0.047).	31.56 (20.04)	35.07 (17.4) (T_0-T_3: p ≥ 0.05) NS.	IG: yes CG: no	IG T_0-T_1 0.20 CG T_0-T_1 0.10	0.10
Geckil[a] 2017	116: IG 53; CG 63	Attitudes Towards Disabled Persons Scale (ATDP) Time 0, 1 & 6 months	55.17 (11.76)	66.81 (14.27) T_2: 63.58 (13.46)	55.68 (10.94)	59.02 (11.71) T_2: 58.43 (11.03)	IG: yes	IG: T_0-T_1: 0.41 T_0-T_2: 0.31 CG: T_0-T_1: 0.14 T_0-T_2: 0.12	M = 0.27 F/U: M = 0.19
Haley[a] 2017	53: IG = 26; CG = 24.	Kiersma-Chen Empathy Scale	73.4 (4.50)	79.7 (3.0) & 83.3 (2.0) [T_0-T_1 & T_0-T_3 p = 0.001]	72.8 (5.9)	72.3 (5.6) T_2: 73.0 (5.5) NS change	IG: yes CG: no	IG: T_0-T_1: 0.63 T_0-T_2: 0.82 CG: T_0-T_1: 0.04 T_0-T_2: 0.02	M = 0.59 F/U: M = 0.80
Henry[a] 2011	89	Empathy questions	12.29 (1.78)	12.64 (1.78) NS change	11.75 (2.10)	12.15 (1.90) NS change	IG: no CG: no	IG: r = −0.10 CG: r = −0.10	NS: 0.0
Lee 2018	103: IG 48; CG 53	Jefferson Scale of Empathy for Health Professional Students (JSE-HPS) Objective Structured Clinical Examination (OSCE) rated by examiners	106.67 (8.14)	111.83 (7.54) 20.67 (7.98)	108.16 (10.36)	108.89 (12.82) 11.61 (7.52) (t = −5.97, p ≤ 0.0001)	Yes IG to CG: yes	IG: T_0-T_1: 0.27 CG: T_0-T_1: 0.03 OSCE: 0.50	0.20 0.50
Menenga[a] 2016	128	Jefferson Scale of Empathy for Health Professional Students (JSE-HPS)	111.8 (9.09)	114.77 (8.95)	112.3 (9.04)	116.1 (12.4)	IG: no CG: no	IG: r = −0.16, CG: r = −0.17	−0.01
Webster 2010	73	Interpersonal Reactivity Index (IRI) Perspective-Taking Empathetic Concern	19.76 (3.98) 21.79 (4.30)	20.31 (3.68) 21.31 (3.82)	19.40 (3.76) 21.82 (2.82)	19.98 (3.83) 22.16 (3.77)	IG: no CG: no IG: no CG: no	Overall not described: NS NS difference IG and CG p ≥ 0.05	
Wikstrom 2003	144	Wheel Questionnaire: Emotional Investment Motivation subscale Structure subscale	1.4 9.6 0.2	1.7 12.0 0.4	1.3 9.2 0.2	1.2 10.3 0.3	IG to CG: Yes NS NS	Overall not reported p = 0.01 p ≥ 0.05 p ≥ 0.05	

Notes: Time 0 (pre-test), time 1 (post-test), time 2 (delayed post-test); IG = intervention, CG = controls.
[a] Denotes RCT.

reported significant decline in mean empathy scores for a group of undergraduate nursing students in a longitudinal study that was conducted over a year (Ward 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 $r ≥ 0.5$. 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 indicated by inferential statistics, the effect sizes for change in overall empathy or for empathy subscales were generally low (see Table 2), ranging from $r = 0.09$ to a high of $r = 0.73$. When taken as an average effect from each study the overall mean was $r = 0.26$. In five of

the studies the empathy intervention effects were seen to be lower than 0.20. We conclude, therefore, that 6 of the 11 single group studies had a real-world impact on empathy.

3.7. 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 $r = 0.2$. 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 $r = 0.45$, 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.

3.8. 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 et al., 2011, Sheehan et al., 2013, Fleming et al., 2015, Bas-Sarmiento et al., 2017, Levett-Jones et al., 2017, Everson et al., 2018) (see Table 2). Of the eight case-control studies, only three (Geçkil et al., 2017; Haley et al., 2017; Lee et al., 2018) showed a statistical improvement of ≥ 0.2 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.

3.9. 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 et al., 2017; Geçkil et al., 2017), manikin-based scenarios (Haley et al., 2017), 3D e-simulations (Everson et al., 2015), and simulations where students wore hemiparesis suits (Levett-Jones et al., 2017). The latter two were point-of-view simulations in which learners experienced the world 'through the eyes' of another person and learned to 'walk in the patient's shoes' (Levett-Jones, 2017). 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 (e.g. 21 h over 6 weeks, or 20 h 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 et al., 2012). Few studies (Mete, 2007; Henry et al., 2011; Cunico et al., 2012; Fleming et al., 2015; Ward, 2016; Haley 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.

4. Discussion

Few healthcare interventions have as much impact on a patient's physical and emotional well-being as empathy (Trzeciak et al., 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 (9/23) 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 et al., 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 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 et al., 2012), others suggest that nursing students may have higher empathetic tendencies than students from other disciplines, possibly diminishing a teaching effect (Gallagher 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% (47.8/56) 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 constrained 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 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 et al., 2013; Bearman et al., 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 and Tatano-Beck, 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.

4.1. 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 few using 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.

5. 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 because they used 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.

Declarations of Interest

None.

Funding

References

Bas-Sarmiento, P., Fernández-Gutiérrez, M., Baena-Baños, M., Romero-Sánchez, J.M., 2017. Efficacy of empathy training in nursing students: a quasi-experimental study. Nurse Educ. Today 59, 59–65.

Batt-Rawden, S., Chisolm, M., Anton, B., Flickinger, T., 2013. Teaching empathy to medical students: an updated systematic review. Acad. Med. 8, 1171–1177.

Bearman, M., Palermo, C., Williams, B., 2015. Learning empathy through simulation. Simul. Healthc. 10, 308–319.

Beddoe, A.E., Murphy, S.O., 2004. Does mindfulness decrease stress and foster empathy among nursing students? J. Nurse Educ. 43 (7), 305–312.

Cant, R., Levett-Jones, T., James, A., 2018. Do simulation studies measure up? A simulation study quality review. Clin. Simul. Nurs. 21, 23–39. https://doi.org/10.1016/j.ecns.2018.06.002.

Chen, A.M., Kiersma, M.E., Yehle, K.S., Plake, K.S., 2015. Impact of the Geriatric Medication Game(R) on nursing students' empathy and attitudes toward older adults. Nurse Educ. Today 35 (1), 38–43.

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. Acad. Med. 90 (8), 1067–1076.

Cunico, L., Sartori, R., Marognolli, O., Meneghini, A.M., 2012. Developing empathy in nursing students: a cohort longitudinal study. J. Clin. Nurs. 21, 2016–2025.

Everson, N., Levett-Jones, T., Lapkin, S., Pitt, V., van der Riet, P., Rossiter, R., Jones, D., Gilligan, C., Courtney-Pratt, H., 2015. Measuring the impact of a 3D simulation experience on nursing students' cultural empathy using a modified version of the Kiersma-Chen Empathy Scale. J. Clin. Nurs. 24 (19–20), 2849–2858.

Everson, N., Levett-Jones, T., Pitt, V., Lapkin, S., Van Der Riet, P., Rossiter, R., Pratt, H.C., 2018. Analysis of the Empathic Concern Subscale of the Emotional Response Questionnaire in a study evaluating the impact of a 3D Cultural Simulation. Int. J. Nurs. Educ. Scholarsh. 15 (1). https://doi.org/10.1515/ijnes-2017-0003.

Eymard, A.S., Crawford, B.D., Keller, T.M., 2010. 'Take a walk in my shoes': nursing students take a walk in older adults' shoes to increase knowledge and empathy. Geriatr. Nurs. 31 (2), 137–141.

Field, A.P., Gillett, R., 2010. How to do a meta-analysis. Br. J. Math. Stat. Psychol. 63 (3), 665–694.

Fields, S.K., Mahan, P., Tillman, P., Harris, J., Maxwell, K., Hojat, M., 2011. Measuring empathy in healthcare profession students using the Jefferson Scale of Physician Empathy: health provider–student version. J. Interprof. Care 25 (4), 287–293.

Fleming, B.D., Thomas, S.E., Burnham, W.S., Charles, L.T., Shaw, D., 2015. Improving ethnocultural empathy in healthcare students through a targeted intervention. J. Cult. Divers. 22 (2), 59–63.

Gallagher, P., Moriarty, H., Huthwaite, M., Lim, B., 2017. Challenging some assumptions about empathy. Clin. Teach. 14 (6), 437–440.

Geçkil, E., Kaleci, E., Cingil, D., Hisar, F., 2017. The effect of disability empathy activity on the attitude of nursing students towards disabled people: a pilot study. Contemp. Nurse 53 (1), 82–93.

Haley, B., Heo, S., Wright, P., Barone, C., Rettigantid, M.R., Anders, M., 2017. Effects of using an Advancing Care Excellence For Seniors simulation on nursing student empathy: a randomized controlled trial. Clin. Simul. Nurs. 13 (10), 511–519.

Hatfield, E., Rapson, R.L., Le, Y.C., 2011. Emotional contagion and empathy. Soc. Neurosci. Empathy 19 January.

Henry, B.W., Ozier, A.D., Johnson, A., 2011. Empathetic responses and attitudes about older adults: how experience with the aging game measures up. Educ. Gerontol. 37 (10), 924–941.

Higgins, J., Green, S., 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Cochrane Collaboration.

Hojat, M., Louis, D.Z., Maio, V., Gonnella, J.S., 2013. Empathy and health care quality. Am. J. Med. Qual. 28, 6–7.

Lee, K.C., Yu, C.C., Hsieh, P.L., Li, C.C., Chao, Y.C., 2018. Situated teaching improves empathy learning of the students in a BSN program: a quasi-experimental study. Nurse Educ. Today 64, 138–143.

Levett-Jones, T., Lapkin, S., Govind, N., ... Everson, N., 2017. Measuring the impact of a 'point of view' disability simulation on nursing students' empathy using the Comprehensive State Empathy Scale. Nurse Educ. Today 59. https://doi.org/10.1016/j.nedt.2017.1009.1007.

Lobchuk, M., Halas, G., West, C., Harder, N., Tursunova, Z., Ramraj, C., 2016. Development of a novel empathy-related video-feedback intervention to improve empathic accuracy of nursing students: a pilot study. Nurse Educ. Today 46, 86–93.

Lown, B., Rosen, J., Marttila, J., 2011. An agenda for improving compassionate care: a survey shows about half of patients say such care is missing. Health Aff. 30, 1772–1778.

McKenna, L., Brown, B.M., Williams, B., Molloy, A., Lewis, B., Molloy, L., 2012. Levels of empathy in undergraduate nursing students. Int. J. Nurs. Pract. 18, 246–251.

McMillan, L.R., Shannon, D.M., 2011. Psychometric analysis of the JSPE nursing student version R: comparison of senior BSN students and medical students attitudes toward empathy in patient care. In: ISRN Nursing, pp. 1–7.

Mennenga, H.A., Bassett, S., Pasquariello, L., 2016. Empathy development through case study and simulation. Nurse Educ. 41, 139–142.

Mete, S., 2007. The empathetic tendencies and skills of nursing students. Soc. Behav. Pers. 35, 1181–1188.

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 Med. 6 (6), e1000097. https://doi.org/10.1371/journal.pmed1000097.

Nosek, M., Gifford, E., Kober, B., 2014. Nonviolent communication (NVC) training increases empathy in baccalaureate nursing students: a mixed method study. J. Nurs.

Educ. Pract. 4 (10), 1.

Ozcan, N.K., Bilgin, H., Eracar, N., 2011. The use of expressive methods for developing empathic skills. Issues Ment. Health Nurs. 32, 131–136.

Polit, D., Tatano-Beck, C., 2012. Nursing Research: Generating and Assessing Evidence for Nursing Practice. Wolters Klewer, Sydney.

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

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. Am. J. Surg. 198, 442–444.

Scott, H., 2011. Empathy in healthcare settings. In: Goldsmiths Research Online. Goldsmiths, University of London.

Sheehan, C.A., Perrin, K.O., Potter, M.L., Kazanowski, M.K., Bennett, L.A., 2013. Engendering empathy in baccalaureate nursing students. Int. J. Caring Sci. 6, 456–464.

Strekalova, Y.A., Krieger, J.L., Kleinheksel, A.J., Kotranza, A., 2017. Empathic communication in virtual education for nursing students. Nurse Educ. 42, 18–22.

Teding van Berkhout, E., Malouff, J.M., 2016. The efficacy of empathy training: a meta-analysis of randomized controlled trials. J. Couns. Psychol. 63, 32.

Trzeciak, S., Roberts, B.W., Mazzarelli, A.J., 2017. Compassionomics: hypothesis and experimental approach. Med. Hypotheses 107, 92–97.

Ward, J., 2016. The empathy enigma: does it still exist? Comparison of empathy using students and standardized actors. Nurse Educ. 41, 134–138.

Ward, J., Schaal, M., Sullivan, J., Bowen, M.E., 2009. Reliability and validity of the Jefferson Scale of Empathy in undergraduate nursing students. J. Nurs. Manag. 17, 3.

Ward, J., Cody, J., Schaal, M., Hojat, M., 2012. The empathy enigma: an empirical study of decline in empathy among undergraduate nursing students. J. Prof. Nurs. 28, 34–40.

Webster, D., 2010. Promoting empathy through a creative reflective teaching strategy: a mixed-method study. J. Nurs. Educ. 49, 87–94.

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

Wikstrom, B.M., 2003. A picture of a work of art as an empathy teaching strategy in nurse education complementary to theoretical knowledge. J. Prof. Nurs. 19, 49–54.

Yu, J., Kirk, M., 2009. Evaluation of empathy measurement tools in nursing: systematic review. J. Adv. Nurs. 65, 1790–1806.