

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/338118465>

What do clinical supervision research reviews tell us? Surveying the last 25 years

Article in *Counselling and Psychotherapy Research* · December 2019

DOI: 10.1002/capr.12287

CITATIONS

0

READS

260

1 author:



[Clifton Edward Watkins, Jr.](#)

University of North Texas

130 PUBLICATIONS 1,690 CITATIONS

SEE PROFILE

REVIEW ARTICLE

What do clinical supervision research reviews tell us? Surveying the last 25 years

C. Edward Watkins Jr. 

Department of Psychology, University of North Texas, Denton, Texas

Correspondence

C. Edward Watkins, Jr., Department of Psychology, University of North Texas, 76203, Denton, TX, USA.
Email: watkinsc@unt.edu

Abstract

What do clinical supervision research reviews across the last 25 years tell us? That question is subsequently examined. Based on database and literature searches, 20 reviews appearing from 1995 through 2019 were identified for survey examination; consistencies, inconsistencies and other defining features were determined across reviews; and the survey findings and their implications are considered. Primary findings are as follows: (a) 'proof' for supervision appears to be more 'proof by association' than otherwise, being primarily a product of ex post facto, cross-sectional, correlational study; (b) evidence supporting supervision impact of any type is weak at best, especially so for worker and client outcomes; (c) supervision models generally lack empirical foundation; (d) evidence-based supervision appears to be more a hope and dream than supervision-based reality at present; and (e) the primary methodological problems that plagued supervision research in the 1990s are still all too frequent in modern research. Some questions to entertain about supervision going forward, and some remedies for improving its research, are proposed.

KEYWORDS

clinical, counselling, psychotherapy, research, review, supervision

1 | INTRODUCTION

Supervision research has been around for over 60 years now, and its importance for the field's advancement has been increasingly recognised (Inman et al., 2014). As supervision studies have continued to accumulate and supervision has acquired greater empirical mass, supervision research reviews have become ever more common fare. Reviews of supervision studies have long been valued for providing a critical perspective on the available research, identifying (a) what can be learned to guide practice, (b) what obstacles emerge as limitations across investigations and (c) what areas are in need of attention going forward (cf. Mulrow, 1994; Munn, Peters, et al., 2018). Such reviews serve as barometers of progress, ideally reflecting increasing methodological sophistication and data base solidification. But is that the case for clinical supervision? Do supervision research reviews reflect such 'progress'? Those questions are subsequently examined.

2 | RESEARCH IN CLINICAL SUPERVISION: A SHORT STATUS REPORT

Five broad areas of supervision research study are as follows: (a) supervision effects on client outcomes; (b) supervision effects on the supervisor-supervisee interaction; (c) supervision's direct effects on supervisee competence; (d) factors that mediate and moderate supervisor impact on supervisee competence; and (e) supervisor and supervisee characteristics (Bernard & Goodyear, 2014). Most research has focused on mediating and moderating factors (e.g. relationship variables, developmental considerations; Bernard & Goodyear, 2014), what Ladany and Malouf (2010) refer to as 'inside supervision' matters. Data support the supervisor-supervisee relationship, perhaps the most substantial 'inside' matter, as being integral to fostering supervisee change (Carifio & Hess, 1987; Ellis, 2010; Park, Ha, Lee, Lee, & Lee, 2019). Data further suggest that supervision positively impacts supervisees, resulting in such gains as enhanced self-awareness, enhanced self-efficacy

and enhanced skill acquisition (Goodyear & Guzzardo, 2000; Inman et al., 2014; Wheeler & Richards, 2007). But research supporting supervision's impact on skill transfer remains limited, as does research addressing supervision's impact on client outcomes.

The current state of supervision research can be compared to psychotherapy research in the 1950s or 1960s: measurement and effectiveness issues loom large as major concerns (Milne et al., 2012). It remains the case that '...although our knowledge and understanding of supervision has bourgeoned..., that which we do not understand or understand well continues to be vast' (Inman et al., 2014, p. 86). Hampering that understanding, supervision research has often been, and continues to be, criticised for the following reasons: small sample sizes, over-reliance on self-report measures, limited number of valid supervision measures, ex post facto designs, limited attention to client outcomes and lack of longitudinal data (Ellis, D'Iuso, & Ladany, 2008; Ellis & Ladany, 1997; Hill & Knox, 2013; Russell, Crimmings, & Lent, 1984). As Milne et al. (2012) have stated, '... we are currently about "half-way there", working on the "search for scientific rigour"...' (p. 144).

3 | SURVEYING SUPERVISION RESEARCH REVIEWS: RATIONALE, FOCUS AND APPROACH

3.1 | Rationale

In continuing that search, I wondered this: Might a survey of supervision research reviews add to our current understanding about supervision's impact, how previously identified supervision research needs are being addressed and how supervision study has changed over time? Furthermore, if '...the evidence of supervisor impact has been well established' (Bernard & Goodyear, 2014, p. 301), how might that evidence be on display across reviews?

Survey is defined here as a 'detailed critical inspection and comprehensive considering' (see WordWeb; <https://wordweb.info/>). To my knowledge, no survey of supervision research reviews has been conducted. It could be informative to assemble review findings, so as to potentially reinforce and bolster current perspective, identify new, promising findings that might not have been properly recognised, or indicate areas that continue to show lack of progress and sorely need attention. Moreover, because a spike in supervision research reviews has occurred across these last five years alone (e.g. Alfonsson, Parling, Spännargård, Andersson, & Lundgren, 2018; Barrett, Gonsalvez, & Shires, 2019; Forshaw, Sabin-Farrell, & Schröder, 2019; Kühne, Maas, Wiesenthal, & Weck, 2019; Newman, Simon, & Swerdlik, 2019), the time for such a survey would seem especially apropos.

3.2 | Focus

The last 25 years, 1995 through 2019, was selected as the review period. That period of time was chosen for three reasons: (a) it covers the most recent generation of supervision scholarship;

(b) research is reported to have increased in both quantity and quality during that time period (Inman et al., 2014); and (c) that particular period captures a time of significant transformation in supervision's history, where the shift to a 'competency-based' and 'evidence-based' perspective has been increasingly embraced and become defining (e.g. Falender & Shafranske, 2017; O'Donoghue, Ju, & Tsui, 2018).

Focus was given to what might broadly be thought of as social (as opposed to medical) services, the specific provision of counselling, psychotherapy or emotional support being an important part of professional role functioning. Social work, counselling, psychology, psychiatric nursing and psychiatry were all included for study. However, reviews were excluded that combined both social and medical services (Cutcliffe, Sloan, & Bashaw, 2018; Dawson, Phillips, & Leggat, 2013; Ducat & Kumar, 2015; Farnan et al., 2012), focused exclusively on medical services (Snowdon, Leggat, & Taylor, 2017), relied primarily on a noncounselling/psychotherapy database (e.g. learning disabilities; Milne & James, 2000) or were primarily scoping (e.g. Goodyear, Bunch, & Claiborn, 2006; Sewell, 2018) or summary (MacDonald & Ellis, 2012; Weerasekera, 2013) in nature. Focus was given to systematic research reviews, defined here as 'a type of research synthesis...to identify and retrieve international evidence that is relevant to a particular question or questions and to appraise and synthesize the results of this search to inform practice, policy and in some cases, further research...' (Munn, Stern, Aromataris, Lockwood, & Jordan, 2018, p. 144). Such reviews are conducted for a host of reasons: to confirm current practices, identify new practices, identify conflicting results, identify areas for future research, uncover international evidence and produce statements to guide decision-making (Higgins & Thomas, 2019; Munn, Peters, et al., 2018). Journal articles, which met that definition and comported with those reasons, were included. Both in-print (already in bound journal form) and online (early view) review articles were allowed. Book chapters, with but two exceptions, were excluded. The decision was made to include the two chapters by Ellis and colleagues (Ellis et al., 2008; Ellis & Ladany, 1997) because they are (a) well-done, highly rigorous systematic research reviews of supervision study methodology and (b) unique in the thoroughgoing attention and scrutiny given to methodological matters.

3.3 | Selection/Analysis approach

Four steps were taken to identify articles for examination: (a) database searches (e.g. Google Scholar, PsycINFO), using such search words as 'supervision', 'clinical', 'research' and 'review', were conducted for the 1995–2019 period; (b) reference sections of identified articles were examined to identify reviews that might have been missed; (c) supervision journals or journals that publish some supervision material were examined for recent reviews; and (d) recent supervision texts (e.g. Bernard & Goodyear, 2019; Scaife, 2019) were also examined to further find possible missed work.

Potential articles/chapters were identified and examined to determine their being a systematic research review, a set of reviews was then identified for further examination, and each review was then read and studied so as to map its main features. Mapped features across the set of reviews included review focus, inclusion criteria, appraisal procedure, findings and limitations/issues. Other important features (e.g. attention to/omission of multicultural factors) were also examined and noted. Next, all reviews were examined for consistencies, inconsistencies and defining features, and all such facets were identified and recorded, the hope being that the resulting supervision information might confirm current practices, identify new practices, identify conflicting results, identify areas for future research and uncover international evidence (cf. Munn, Peters, et al., 2018).

Examination of reviews was approached in two ways. First, a broad view was taken, with consistencies, inconsistencies and defining features being determined across the entire body of identified reviews. Second, the body of reviews was divided into two groupings of 12 years each: (a) the 1996–2007 years; and (b) the 2008–2019 years. No supervision reviews were identified as having appeared in 1995, so that year was excluded in order to make for two groupings equivalent in number of years covered. Those two periods were examined for any defining differences.

4 | RESULTS

A total of 20 clinical supervision research reviews were identified for study. Table 1 provides the mapped features of each review.

4.1 | What do these reviews tell us? The broad view

4.1.1 | Focus of reviews

The reviews were varied in focus and can be roughly grouped as follows: general reviews, where ‘the whole’ of supervision was examined (Bogo & McKnight, 2006; Borders, 2006; Buus & Gonge, 2009; Newman et al., 2019; O’Donoghue & Tsui, 2015; Tsui, 1997); reviews primarily focused on supervision effects on supervisees and/or clients (Alfonsson et al., 2018; Carpenter et al., 2013; Freitas, 2002; Kühne et al., 2019; Reiser & Milne, 2014; Watkins, 2011; Wheeler & Richards, 2007); single reviews about evidence-based practice (Barrett et al., 2019), supervision models (Simpson-Southward, Waller, & Hardy, 2017), the supervisee’s experience of receiving supervision (Wilson, Davies, & Weatherhead, 2016) and the supervisor’s experience of providing supervision (Forshaw et al., 2019); and reviews focused on methodology and rigour (Ellis et al., 2008; Ellis & Ladany, 1997; Ellis et al., 1996).

Quantitative, qualitative, single-case or other

Virtually all reviews focused exclusively on quantitative studies. However, two reviews were exclusively qualitative in nature

(Forshaw et al., 2019; Wilson et al., 2016), and one was a content analysis (Simpson-Southward et al., 2017).

The reviewers reviewing

The reviews cut across several mental health disciplines, being provided by representatives from social work, counsellor education, nursing and psychology. Countries from which the reviews originated were Australia, Canada, China, Denmark, Germany, New Zealand, Sweden, the UK and the United States.

Journals in which reviews appeared (excluding two book chapters)

Reviews appeared in a host of varied journals: *BMC Psychology*, the *British Journal of Social Work*, *Children and Youth Services Review*, *Clinical Psychologist*, *Clinical Psychology and Psychotherapy*, *Clinical Supervisor*, *Cognitive Behaviour Therapy*, *Counselling and Psychotherapy Research*, *International Journal of Mental Health Nursing*, *Journal of Social Service Research*, *Mental Health Review Journal*, *Psychology in the Schools*, *Psychotherapy* and *Training and Education in Professional Psychology*.

4.1.2 | Inclusion criteria

Two inclusion criteria were consistently applied across most reviews: that the selected articles have been peer-reviewed and be in English. Where not explicitly stated, those criteria still seemed to be in play. Other criteria typically were reflective of any given review’s particular focus (e.g. that a measure of outcome be included where outcome was the primary concern).

4.1.3 | Appraisal procedure/criteria

Thirteen of 20 reviews involved the use of some type of objective appraisal procedure, with some type of rating of study quality typically being made. Different rating tools mentioned included the Critical Appraisal Skills Programme, Weight of Evidence approach and researcher-created evaluation checklists. PRISMA reporting guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) were applied in at least five reviews, and in three cases, a research protocol was produced and registered (with PROSPERO; International Prospective Register of Systematic Reviews) before the review’s initiation. With PRISMA (Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009) and PROSPERO (Page, Shamseer, & Trico, 2018) both being products of the last approximate decade, their use across supervision reviews is understandably a most recent phenomenon.

4.1.4 | Features of reviewed publications

The number of reviewed articles across reviews ranged from a low of four (Barrett et al., 2019) to a high of 131 (Ellis et al., 1996), with about

TABLE 1 Supervision reviews across the last 25 years

Authors	Focus of review	Inclusion criteria	Appraisal procedure/criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Alfonsson et al. (2018)	Effects of supervision on supervisees and patients	<ul style="list-style-type: none"> a. That supervision focused on ongoing cases b. That treatment conducted within broad cognitive-behavioural framework c. That experimental, quasi-experimental or single-case design used d. That quantifiable outcome measure of supervisees and/or patient effects used e. In English language 	<p>Jadad scoring system used and CONSORT guidelines followed</p> <p>Review process followed guidelines in <i>Cochrane Handbook of Interventions</i> and by PRISMA</p> <p>Review protocol registered prospectively with PROSPERO and published</p>	<p>5 articles reviewed (out of 4,103 initial hits)</p> <p>Years covered: 2006–2016</p> <p>4 studies used randomised controlled design, and one used nonrandomised controlled design</p> <p>2 studies conducted in the United States, 1 in Australia, 1 in Germany and 1 in the UK; UK study involved supervisees from Russia and Ukraine</p>	<ul style="list-style-type: none"> a. Most frequent methodological shortcomings: failure to report power analyses, inadequate reporting of randomisation procedures and inadequate reporting of post hoc analyses b. Supervision may impact supervisee competence c. Any supervision impact on patients is weak at best d. No support provided for any particular supervision model e. Need to explore supervision process, a neglected variable 	<ul style="list-style-type: none"> a. Number of supervisors and patients involved unreported in two studies b. Duration of supervision unspecified in one study c. Frequency of supervision unspecified in one study d. Supervision intervention in one study consisted of a single 30-min supervision session per month for three months in one study
Barrett et al. (2019)	To review evidence for scientist-practitioner and evidence-based practice (EBP) in clinical supervision	<ul style="list-style-type: none"> a. That supervision be provided by accredited supervisors b. Evidence-based practice and/or scientist-practitioner competence as an outcome c. Publication in peer-reviewed journal 	<p>Appraisal tool for Cross-Sectional Studies</p> <p>Review process followed PRISMA guidelines</p> <p>Review protocol developed in accordance with <i>Cochrane Handbook of systematic reviews</i>, registered prospectively with PROSPERO and published</p>	<p>4 articles reviewed (out of 1,287 initial hits)</p> <p>Years covered: 2015–2017</p> <p>2 studies conducted in Australia and 2 others from the United States</p>	<ul style="list-style-type: none"> a. Virtual absence of EBP/supervision research b. Supervisors encouraged to use evidence-based supervisory activities (e.g. video feedback, modelling) c. Supervisors encouraged to use competency evaluation rating forms in monitoring supervisee progress 	<p>Narrow inclusion criteria</p> <p>Focus on clinical psychology trainees only</p>
Bogo & McKnight (2006)	To examine social work supervision research	<ul style="list-style-type: none"> a. That publication be peer-reviewed b. Appear during designated ten-year review period 	<p>None specified</p>	<p>13 articles reviewed</p> <p>Years covered: 1994–2004</p> <p>All studies from the United States</p>	<ul style="list-style-type: none"> a. Dearth of empirical studies on social work supervision across the past decade b. Most studies used small, convenience samples and were cross-sectional, single projects c. Race/ethnicity studied in only two projects d. No evidence that supervision affects worker or client outcomes e. Supervision models remain untested f. Supervision's evidence base embryonic 	<p>All studies from the United States</p> <p>Criteria for article selection unspecified</p> <p>No objective appraisal procedure used</p> <p>Process of arriving at conclusions unspecified</p>

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Borders (2006)	To review conceptual and empirical supervision literature from the fields of counselling and counselling education	<p>a. That publication appear in American Counseling Association journal or one of three international counselling journals (<i>International Journal for the Advancement of Counselling</i>, <i>British Journal of Guidance and Counselling</i> and <i>Canadian Journal of Counselling</i>)</p> <p>b. Appear during designated five-year review period</p>	None specified	84 articles reviewed Years covered: late 1999 to early 2005	<p>a. The supervisory relationship emerged as highly important, that being especially so for multicultural supervision</p> <p>b. Multicultural supervision received increased attention, feedback was given minimal research attention, use of technology began to be investigated, and school counselling received the most attention across counselling specialities</p> <p>c. Few studies employed an experimental design, and qualitative studies were comparatively more frequent, with even a few mixed-methods studies also being conducted</p> <p>d. White females were preponderant as research study participants</p> <p>e. Studies often involved small samples, were based in but a single programme and were ex post facto in design</p>	<p>Author sole evaluator</p> <p>No objective appraisal procedure used</p>
Buus & Gonge (2009)	To summarise and critically evaluate all empirical studies of clinical supervision in psychiatric nursing	<p>a. That publication be empirical study of clinical supervision in psychiatric nursing</p> <p>b. In English language</p>	Articles systematically examined using specialised checklists from CONSORT, STROBE and COREQ	34 articles reviewed Years covered: 1990-2007 Diverse representation of geographical contribution; Sweden primary contributor; other contributors included Australia, Finland, the UK and the United States	<p>a. Clinical supervision in psychiatric nursing commonly viewed as a good practice, but empirical evidence supporting this view is limited</p> <p>b. Studies were often small-scale, did not control for confounding factors and had reliability/validity issues</p> <p>c. Unable to confidently say supervision in psychiatric nursing settings has desired outcomes on nurses or their patients</p> <p>d. Identified research obstacles include the following: general lack of consensus among researchers about what research instruments to use; and general lack of consensus about which definitions and models should guide supervision research. Furthermore, such variables as shifting workloads, stressful events, interpersonal and interprofessional conflicts, and patients' degenerating or volatile conditions render randomisation of participants and controlling for confounding factors particularly problematic</p>	<p>Conceptualisations of nursing practices show some variation around the world, making data comparisons difficult</p>

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/ criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Carpenter, Webb, & Bostock (2013)	To determine what is known about supervision effectiveness; and appraise evidence base for supervision in child welfare services	<ul style="list-style-type: none"> a. That publication be peer-reviewed b. That study be quantitative or qualitative and include social workers or other child welfare practitioners c. In English language 	Weight of Evidence approach employed (ratings provided for each study on trustworthiness, appropriateness of study design and topic relevance) Review process followed PRISMA guidelines and methods developed by Social Care Institute for Excellence	22 articles reviewed (out of 1,590 initial hits) Years covered: 2000–2012 All studies from the United States	<ul style="list-style-type: none"> a. Primary study problems included the following: no randomised controlled trials or quasi-experimental investigations; small sample sizes; and inadequate descriptions b. The vast majority of studies were cross-sectional and correlational, with only one study being intervention (pre/post) in design c. Supervision found to be positively associated with job satisfaction, job retention and ability to manage workload d. Supervision works best when attending to the supervisor–supervisee relationship, providing social and emotional support and providing task assistance e. Impact of supervision models on outcomes not studied f. Very limited, weak evidence for supervision impact on worker outcomes g. Impact of supervision on consumer outcomes an unstudied issue 	<ul style="list-style-type: none"> All studies from the United States Inclusion criteria possibly too broad Weak research designs perhaps admitted Restricted time frame
Ellis et al. (2008)	To provide update of earlier review (Ellis & Ladany, 1997), giving specific focus to supervision measures and evaluating 37 validity threats	<ul style="list-style-type: none"> a. That article be about clinical supervision b. That article be data-based c. That article focus on supervision measurement 	Three raters trained for five hours in rating 37 threats 90% rater agreement reached before beginning study	6 articles reviewed Years covered: 1995–2007 All studies from the United States	<ul style="list-style-type: none"> a. New supervision measures continue to be developed; some evidence that more stringent approach is being taken to measure development b. Most research still not scientifically rigorous, with numerous threats to validity being on display c. More sound construction of supervision-specific measures necessary for field to advance 	The same as Ellis, Ladany, Krengel, and Schult (1996)/ Ellis and Ladany (1997)

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/ criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Ellis and Ladany (1977)	Replication and extension of Ellis et al. (1996), conducting a more circumscribed methodological review specifically targeting supervisors and clients and evaluating 37 validity threats	The same as Ellis et al. (1996)	The same as Ellis et al. (1996)	96 publications reviewed (one empirically based book and 95 research articles) Years covered: 1981–1995 Of asterisked studies in chapter References, indicating their inclusion in review, most from the United States	<p>a. Quantitative findings closely parallel earlier review (see Ellis et al., 1996, entry below)</p> <p>b. Quality of supervision research across 15-year period judged substandard</p> <p>c. About 75% of studies ex post facto in nature (no random assignment, no independent variable manipulation)</p> <p>d. Supervision models remain minimally studied</p> <p>e. Supervision replication studies are scarce</p> <p>f. Continued lack of sound supervision measures</p> <p>g. Clients generally not included in supervision study; quality of supervision/client outcome research poor</p>	The same as Ellis et al. (1996) Possible threats to hypothesis validity Review subject to Type II error Some review-based conclusions possibly spurious
Ellis et al. (1996)	Critique supervision research with regard to scientific rigour and methodological quality; evaluate 49 threats to validity across studies; threats included statistical conclusion validity, internal validity, construct validity, external validity, hypothesis validity and methodological threats	<p>a. That article be consistent with provided definition of clinical supervision</p> <p>b. That article be data-based</p> <p>c. That article involve supervision of individual counselling/therapy</p>	<p>Four raters trained for 10 hr in rating 49 threats</p> <p>90% rater agreement reached before beginning the study</p>	<p>131 publications (one empirically based book and 130 research articles)</p> <p>Years covered: 1981–1993</p> <p>Of asterisked studies in article References, indicating their inclusion in review, virtually all from the United States</p>	<p>a. About 75% of studies ex post facto in nature</p> <p>b. On average, investigators conducted their research with a 100% probability of committing one or more Type II errors</p> <p>c. At least 80% of the studies were judged to have inflated Type I or Type II error rates or unreliable measures</p> <p>d. Selection bias and ambiguity of causal direction, respectively, were threats to 77% and 69% of the studies</p> <p>e. Mono-method bias, construct confounding and inadequate preoperational explication were three consistent threats to construct validity</p> <p>f. Inconsequential hypotheses, ambiguous hypotheses and diffuse statistical hypotheses frequently emerged as threats to hypothesis validity</p> <p>g. Most studies had inadequate sample sizes (78%) and used only self-report data (66%)</p> <p>h. Most supervision studies were simultaneously likely to find spurious significant results and unlikely to detect true effects</p>	<p>Criteria definitions could have precluded detecting design variations in supervision research</p> <p>49 threats not all inclusive</p> <p>No direct comparison group</p> <p>Selection bias possible</p>

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Forshaw et al. (2019)	To examine supervisors' experience of providing supervision: 'What are the supervisor's experiences of providing clinical supervision to qualified therapists?' (p. 53) To perform qualitative meta-ethnographic synthesis of reviewed studies	a. Supervisor's experience of doing supervision being the focus of the study b. Qualitative methodology employed c. Supervisors being either psychologists, psychotherapists or counsellors d. In English language	Critical Appraisal Skills Programme Review process followed PRISMA guidelines	6 articles reviewed (out of initial set of possible 236) Years covered: 1999–2016 4 studies conducted in the UK, 1 in Australia and 1 in the United States	a. Primary findings revolved around four themes: experiencing difficulties in supervision, responsibility, similarities to treatment and feeling capable as a supervisor b. Supervision, because of its multifaceted nature, inherent power imbalance and competing demands, can be highly challenging c. Doing supervision can have significant personal and professional impacts, positive and negative, on the supervisor d. Doing supervision often leads to perceived growth as a supervisor e. Supervisor training can render supervisor unpreparedness and bewilderment avoidable	Three studies failed to specify participant gender and race/ethnicity, two failed to indicate supervisors' number of years providing supervision Two studies unpublished Article quality ratings variable
Freitas (2002)	To review in detail research on supervision outcome, re-examining several studies from Ellis and Ladany (1997) and also adding in three other studies (appearing from 1993 to 1997) that they did not cover	That supervision outcome be addressed	None specified	10 articles reviewed Years covered: 1981–2001 8 articles from the United States, 1 from Canada and 1 from Sweden	a. Being informed by and drawing from review of Ellis and Ladany (1997), Freitas' findings highly consistent with theirs b. Two most common problems to emerge across studies were as follows: failure to control for Type I and Type II error; and failure to provide psychometric data on measures employed	Author sole evaluator No objective appraisal procedure used

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Kühne et al. (2019)	To review current status of supervision interventions and methodological quality of empirical literature	<ul style="list-style-type: none"> a. Supervision of psychotherapy the focus of each study b. Adult patients only c. Publication peer-reviewed d. Outcome measure included 	<p>Reviewer-developed comprehensive tool for evaluating methodological quality; ratings provided of how well key methodological issues addressed (e.g. control of confounds, sources of bias)</p> <p>PRISMA guidelines followed and review protocol registered and published with PROSPERO</p>	<p>19 articles reviewed (out of 8,786 initial hits)</p> <p>Years covered: 2001–2017</p> <p>5 studies used randomised controlled design, and 1 used cluster randomised controlled design; follow-up data collected in only 3 studies</p> <p>Majority of reviewed studies from the United States; other reviewed studies from Australia, Germany, Hong Kong, Sweden and the UK</p>	<ul style="list-style-type: none"> a. Supervision generally seen as helpful by supervisees, as benefitting their therapeutic competence b. A variety of supervision interventions were employed, with case discussion and providing feedback being most evident c. A supervision manual or supervisor training not routinely a part of research studies d. Most studies were uncontrolled or used small samples, with about half relying exclusively on self-report questionnaires e. Virtually all studies were at high risk with regard to three methodological quality issues: participant selection, control of confounds and other sources of bias f. Supervision research lags well behind psychotherapy research g. These questions still remain without definitive answers: What are supervision's active ingredients? Does supervision effect client change? 	<p>Evaluation tool reviewer-developed</p> <p>Stricter operationalisation of inclusion criteria may have resulted in fewer reviewed studies being selected, accordingly limiting drawables conclusions</p>

Newman et al. (2019)	To review current evidence about supervision in school psychology	<ul style="list-style-type: none"> Studies required to focus conceptually or empirically on school psychological supervision processes and practices 	<p>Systematic mapping; review descriptive as opposed to interpretive</p>	<p>37 articles reviewed (out of initial sample of 70 articles), with 21 being empirical and 16 being conceptual; of 21 empirical publications, 13 were surveys, four were qualitative, two were mixed-methods, and two were quantitative</p> <p>Years covered: 2000–2017</p> <p>Most studies focusing on school psychology supervision in the United States; other contributions coming from Australia, Hong Kong, New Zealand and the UK</p>	<ul style="list-style-type: none"> a. School psychology supervision literature quite limited, its research lagging well behind supervision research in related fields b. Topics receiving minimal attention include the following: supervision processes; outcome; models, methods and formats; technology; supervisor training; certain areas of diversity; and gatekeeping c. A methodologically pluralistic approach to doing school psychology supervision research recommended 	<p>Empirical rigour of reviewed studies not assessed</p> <p>Methods and methodology descriptions taken at face value</p>
----------------------	---	---	--	---	--	--

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/ criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
O'Donoghue & Tsui (2015)	To provide comprehensive review of research on practising social workers	<ol style="list-style-type: none"> That articles be peer-reviewed That articles include firsthand empirical information That articles be about practising social workers, not students 	None specified	<p>86 articles reviewed</p> <p>Years covered: 1970–2010</p> <p>About 60% of studies from the United States; other notable contributors were Australia, Hong Kong, Israel and New Zealand</p>	<ol style="list-style-type: none"> Social work supervision research increased in number and geographical spread Research designs have increasingly diversified and show increasing sophistication of data analyses Supervisor emotional support within trusting relationship mitigates work stress impact and positively related to job satisfaction Cultural differences examined in a few studies Clients minimally included in studies, influence of supervision on client outcomes largely unexamined Empirically supported supervision practice model lacking 	<p>No appraisal procedure used</p> <p>Process of arriving at conclusions unspecified</p>
Reiser & Milne (2014)	Applying fidelity framework to supervision outcome studies	<ol style="list-style-type: none"> Articles from peer-reviewed journals over the last 30 years Direct measure of client outcome included Supervision provided had to be of psychotherapy or counselling services Supervision provided consistent with Milne's (2007) empirical definition of clinical supervision 	Fidelity Framework Checklist	<p>12 articles reviewed (out of initial search sample of 48)</p> <p>Years covered: 1981–2010</p> <p>7 studies from the United States; 2 from Australia, 2 from the UK and 1 from Canada</p>	<ol style="list-style-type: none"> Areas of infidelity were identified across most reviewed studies, with the most troubling issues being receipt, training and delivery Fidelity framework considered useful for assessing adherence in supervision studies Argument made for also giving 'acid test' weight to other outcomes beyond client outcomes alone 	<p>Possible rater bias</p> <p>Exclusionary criteria may have been overly restrictive</p>

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/ criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Simpson-Southward et al. (2017)	Examination of consistencies across supervision models	<ul style="list-style-type: none"> a. That the supervision model describes what happens in supervision b. That model be about one-to-one supervision c. In English language 	<p>Content analysis conducted</p> <p>Articles evaluated with five questions in mind:</p> <ul style="list-style-type: none"> a. Was supervisee learning and/or development addressed? b. Were managerial and/or ethical responsibilities considered? c. Were emotional effects of the work considered? d. Was the model based on empirical evidence? e. Does the model take into account supervisor, supervisee and client? 	<p>52 articles reviewed</p> <p>Years covered: 1964–2015</p>	<ul style="list-style-type: none"> a. Most supervision models, while focusing on supervisee learning and/or development, lack consistency and are not empirically based b. Supervision models are largely assumption-based in nature c. The client remains the neglected factor in supervision models 	<p>The authors conclude that 'There is little evidence for the effectiveness of supervision' (p. 1,228). While research about the particular models in question was considered, supervision outcome research was not specifically evaluated as part of conceptual review. Perhaps the most precise statement is this: 'There is little evidence for the effectiveness of supervision models'.</p> <p>Various aspects of the content analysis can be questioned, some examples being as follows: Although at least two models were labelled and described as developmental, they were simultaneously identified as not addressing supervisee development (see Hess and Ward/House entries) under the appraisal question, 'was supervisee learning and/or development addressed?';</p> <p>Some supervision models are identified as not addressing a particular supervision dimension that they do indeed seem to address (see entries for the Frawley-O'Dea/Samat relational model; Hawkins/Shohet seven-eyed model; Mueller/Keil coping with conflict model; and Scaife model)</p> <p>One model, a model about supervisor development, is faulted for not focusing on the supervisee</p>

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Tsui (1997)	To provide comprehensive review of research on practising social workers	<ul style="list-style-type: none"> a. That articles be peer-reviewed b. That articles include firsthand empirical information c. That articles be about practising social workers, not students 	None specified	30 articles reviewed Years covered: 1970–1995 Mention made of supervision research coming from Australia, Canada, Hong Kong, Israel and the United States	<ul style="list-style-type: none"> a. Social work supervision research embryonic b. Most studies were one-shot, cross-sectional surveys, relied on self-report questionnaires, and had problems with internal, external, and/or construct validity c. Lack of programmatic investigation d. No study addressed culture e. Client outcomes minimally addressed 	<ul style="list-style-type: none"> Author sole evaluator No objective appraisal procedure used Process of arriving at conclusions unspecified
Watkins (2011)	To examine the effects of supervision on client outcomes	<ul style="list-style-type: none"> a. Used identified outcome studies from four previous reviews b. Identified additional articles for inclusion appearing since four reviews, using 'supervision' and 'outcome' as search words 	None specified	18 articles reviewed Years covered: 1981–2010 11 studies from the United States; other contributions coming from the UK (3), Australia (2), Canada (1) and Sweden (1)	<ul style="list-style-type: none"> a. Misidentification, data duplication and inadequate measurement identified as problems with set of studies b. Not able to confidently say that supervision positively affects client outcome c. Three studies identified as pointing the way forward for future supervision outcome investigations 	<ul style="list-style-type: none"> Author sole evaluator No objective appraisal procedure used
Wheeler & Richards (2007)	Review conducted so as to answer this question: What impact does supervision have on counsellors/therapists, their practice and their clients?	<ul style="list-style-type: none"> a. Studies either published or unpublished, quantitative or qualitative (but no unpublished studies included in final sample) b. Valid, reliable measure or rigorous qualitative analysis used to show supervision's impact on supervisee c. Supervisees had to be practising counsellors or therapists d. In English language 	<p>Review conducted using EPPI-Reviewer software</p> <p>Inclusion/exclusion criteria programmed into software</p> <p>Data extraction procedure also programmed into software, so that various matters of methodology and design could be recorded</p> <p>Quality ratings made as to each study's methodological soundness and about how well each study fit with the inclusion criteria</p>	<ul style="list-style-type: none"> a. Supervision shown to have some impact on supervisee self-awareness, skill development and self-efficacy b. Minimal research attention given to the client: '...no studies ...offer substantial evidence to support improvement in client outcomes' (p. 63) 	<ul style="list-style-type: none"> Based on quality ratings, only two of 18 studies rated 'very good' Specifics about empirical rigour of reviewed studies (or lack thereof) not provided Studies identified as being supportive of any given area (e.g. supervision and self-awareness) quite limited Self-report measures predominant 14 studies were from the United States 	

(Continues)

TABLE 1 (Continued)

Authors	Focus of review	Inclusion criteria	Appraisal procedure/criteria	Features of reviewed publications	Findings/conclusions	Study limitations/issues
Wilson et al. (2016)	To examine trainee therapists' experiences of receiving supervision To perform qualitative meta-synthesis of reviewed studies	a. That qualitative design be employed, where interviews or focus groups were used b. That focus be on trainees' previous supervision experiences c. In English language	Critical Appraisal Skills Programme	15 articles reviewed (out of 1,019 initial hits) Years covered: 1996–2012 9 studies from the United States, 3 from the UK, 1 from Australia, 1 from Canada and 1 from Norway	a. Four crucial themes identified: supervision as a learning opportunity; the supervisory relationship; power in supervision; and supervision impact. Positive and negative aspects of each theme considered b. Supervision can contribute to supervisee personal and professional development, but also has the potential to cause supervisee stress, anxiety and self-doubt c. Supervisors should remain mindful of the power differential and its impact on supervision	Majority of participants female

Note: CONSORT, Consolidated Standards of Reporting Trials; EPPI, Evidence for Policy and Practice Information and Co-ordinating Centre; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PROSPERO, International Prospective Register of Systematic Reviews; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology. EPPI-Reviewer software is a Web-based software program for managing and analysing data in literature reviews.

half of the reviews involving anywhere from 10 to 40 articles. The total span of years covered ranged from 1964 (Simpson-Southward et al., 2017) up to 2017 (Barrett et al., 2019), with most reviews accounting for a span of 10–30 years. The vast majority of reviewed studies came from the United States. Other primary contributors were the UK and Australia, followed (in alphabetical order) by Canada, Finland, Germany, Hong Kong, Israel, New Zealand, Norway and Sweden.

4.1.5 | Findings/conclusions

Findings/conclusions can be placed together into four rough groupings: supervision impact, supervision models, methodological matters and other (miscellaneous concerns or conclusions).

Supervision impact

Supervision, found to be positively associated with job satisfaction, job retention and ability to manage workload (Carpenter et al., 2013; O'Donoghue & Tsui, 2015), appears to be seen as helpful by supervisees (Kühne et al., 2019) and may even benefit their therapeutic competence (e.g. enhanced self-awareness, enhanced sense of self-efficacy; Alfonsso et al., 2018; Kühne et al., 2019; Wheeler & Richards, 2007). But supervision's favourable impact on worker outcomes is weak at best, yet to be proven (Bogo & McKnight, 2006; Carpenter et al., 2013). Furthermore, the client has been, and continues to be, summarily neglected in supervision research: supervision's impact on client outcome has yet to be proven (Alfonsso et al., 2018; Bogo & McKnight, 2006; Buus & Gonge, 2009; Carpenter et al., 2013; Ellis & Ladany, 1997; Ellis et al., 1996; O'Donoghue & Tsui, 2015; Simpson-Southward et al., 2017; Tsui, 1997; Watkins, 2011; Wheeler & Richards, 2007).

Study data about giving and receiving supervision also accentuated this: supervision, because of its multifaceted nature, inherent power imbalance and competing demands, can be highly challenging and have significant personal and professional impacts, positive and negative, on both supervisor and supervisee, and both parties are advised to remain mindful of those realities (Forshaw et al., 2019; Wilson et al., 2016).

Supervision models

Supervision models largely went untested, with no clear evidence supporting any given model (Alfonsso et al., 2018; Bogo & McKnight, 2006; Buus & Gonge, 2009; Carpenter et al., 2013; Ellis & Ladany, 1997; Ellis et al., 1996; Simpson-Southward et al., 2017). Supervision lacks evidence-based practice research and an evidence-based practice model (Barrett et al., 2019; O'Donoghue & Tsui, 2015; Simpson-Southward et al., 2017).

Methodological matters

Methodological issues and study shortcomings were consistently identified and can be placed into three (not necessarily mutually exclusive) broad groupings: (a) failure to control for compromising variables; (b) type of study conducted; and (c) recurring issues about critical study components. Frequently identified compromising variables included unaddressed threats to validity, sources of bias and confounds, as well

as inadequate reporting and description of study specifics (Alfonsson et al., 2018; Buus & Gonge, 2009; Carpenter et al., 2013; Ellis et al., 2008; Ellis & Ladany, 1997; Ellis et al., 1996; Freitas, 2002; Kühne et al., 2019; Tsui, 1997). The vast majority of studies were described as being *ex post facto*, cross-sectional and correlational (i.e. no randomisation, no independent variable manipulation), with very few being experimental investigations (Bogo & McKnight, 2006; Borders, 2006; Carpenter et al., 2013; Ellis & Ladany, 1997; Ellis et al., 1996; Tsui, 1997). Recurring problematic issues included small sample sizes, an over-reliance on self-report measures and retrospective accounts, and a lack of sound supervision measures from which to draw (Alfonsson et al., 2018; Bogo & McKnight, 2006; Borders, 2006; Buus & Gonge, 2009; Ellis et al., 2008; Ellis & Ladany, 1997; Ellis et al., 1996; Kühne et al., 2019; O'Donoghue & Tsui, 2015; Tsui, 1997).

4.1.6 | Other study limitations/issues

Other study limitations/issues included inclusion criteria being overly restrictive or perhaps not restrictive enough (e.g. Barrett et al., 2019; Carpenter et al., 2013), some studies selected for review having incomplete information (Alfonsson et al., 2018) and being of variable quality (Wheeler & Richards, 2007), and about one-third of the reviews involving no objective appraisal procedure (e.g. Bogo & McKnight, 2006; Watkins, 2011). Additional issues also mentioned or identified were as follows: supervision process is a neglected variable and could benefit from study (Alfonsson et al., 2018); and neither a supervision manual nor supervisor training is routinely included in supervision research (Kühne et al., 2019).

4.1.7 | What do these reviews tell us? The two 12-year periods compared

Perhaps the clearest comparative results to emerge would be the following: (a) the primary problematic features of and limitations that attended clinical supervision research from 1996 to 2007 (e.g. *ex post facto* designs, small samples) have loudly lingered into the 2008–2019 period and remain as current problematic features and limitations; (b) the use of some type of study appraisal procedure was standard fare for most reviews appearing from 2008 to 2019, but the reverse was the case for reviews appearing during the earlier time period; and (c) systematic supervision reviews appearing in the last several years have been more apt to be conducted in accordance with established systematic review guidelines (e.g. PRISMA and PRISMA-P [PRISMA for Protocols; Moher, Shamseer, et al., 2015; Shamseer et al., 2015]).

5 | DISCUSSION

These survey findings can be viewed in two ways, as the 'glass being half empty' and the 'glass being half full'. It seems important to hold both views in mind.

5.1 | The 'Glass Half Full'

These reviews appear to reflect (a) a growing and lively research interest in supervision that increasingly stretches around the globe, that interest being most clearly on display across these last 15 years (cf. Pelling, Abbott, & Lack, 2017); (b) some diversification in study designs and sophistication of statistical analyses, particularly the welcome embrace of qualitative studies (Borders, 2006; O'Donoghue & Tsui, 2015); and (c) continuing efforts being made across mental health disciplines to closely scrutinise supervision research so as to advance the field. Those observations are also reflected in other recent assessments about supervision and its research and seemingly are givens going forward (Bernard & Goodyear, 2019; Inman et al., 2014; Sewell, 2018). Supervision researchers are clearly 'working on the "search for scientific rigour"...' (Milne et al., 2012, p. 144), and all indications are that they will continue to vigorously do so.

5.2 | The 'Glass Half Empty'

Four issues are most concerning: supervision impact, supervision models, methodological matters and evidence-based supervision.

5.2.1 | Supervision impact

Scholarly opinion—based upon the supposed weight of empirical evidence—is that supervision contributes to supervisee competence development and skill transfer, but any impact on client outcome has yet to be proven (e.g. Bernard & Goodyear, 2019). But what is most surprising from this survey of reviews is the limited evidence that affirms any type of supervision impact at all: although attention to clients and client outcome was identified frequently and repeatedly as being sorely needed (e.g. Alfonsson et al., 2018; Kühne et al., 2019), any evidence that supervision benefits supervisees was highly tentative and quite thin at best. Supervision has indeed been found to be associated with some positive outcomes (e.g. increased sense of supervisee self-efficacy; Carpenter et al., 2013; O'Donoghue & Tsui, 2015; Wheeler & Richards, 2007), but beyond saying that, these reviews offer little else that can be definitively concluded about supervision's impact. If '...the evidence of supervisor impact has been well established' (Bernard & Goodyear, 2014, p. 301), where is that evidence?

The question 'Does supervision work?' still seems to be a wide open question, not answered affirmatively and conclusively on any front across these reviews. Could it be that claims about supervision's impact and effectiveness have been overstated, maybe even oversold (Alfonsson et al., 2018; Carpenter et al., 2013; White, 2017)?

5.2.2 | Supervision models

Supervision models generally went unexamined across these reviews, any such study appearing to be more often a product of the

past than present. The idea of evidence-supported supervision models appears to be far more hope and dream than reality at present. Even when the one model-specific review about cognitive-behavioural supervision (Alfonsson et al., 2018) is closely examined, one could question whether it is truly a model-specific review. For example, one of the five reviewed 'so designated' CBT supervision studies—the often cited research investigation by Bambling, King, Raue, Schweitzer, and Lambert (2006)—mentions CBT minimally, instead the focus being on alliance and problem-solving conditions, and involved the use of an unsound, nonspecific supervision manual (Milne, 2016). Three of the other four studies, while occurring within the context of supervised CBT, were examinations about the impact of different modes of supervision delivery (e.g. bug-in-the-eye vs. video feedback, cotherapy treatment/supervision vs. single-provider treatment/supervision). The fifth study's supervision condition consisted of but one 30-min Skype contact per month for a total of three months. Again, that these purportedly 'carefully selected studies' are actual tests of CBT supervision can be questioned.

Pushing this line of thinking even further, how much do supervision models really impact daily supervision practice anyway? Could it be that 'most supervisors eventually develop their own unique integrationist perspectives..., [that that outcome] probably is inevitable' (Bernard & Goodyear, 2004, p. 100)? If that is the case, what specific purpose does the multitude of models serve (Simpson-Southward et al., 2017)? Perhaps such models provide supervisors with the bits and pieces from which they construct those unique integrationist perspectives, with some of those larger bits and pieces (e.g. supervisory alliance; Park et al., 2019) ideally enjoying some level of empirical support. But beyond possible piecemeal support, empirical backing for the vast majority of models (with the discrimination and developmental models excepted; Bernard & Goodyear, 2019; Rønnestad, Orlinsky, Schröder, Skovholt, & Willutzki, 2019) appears lacking. Based on this survey of reviews, it does not seem beyond reasonable to wonder, 'Is research on supervision models dead?'

5.2.3 | Methodological matters

The primary methodological problems that plagued supervision research in the 1990s are still the problems that plague research today: 'The conclusions from previous systematic reviews of clinical supervision, that there is a dire need for more empirical evaluations and that many studies in this area lack adequate methodological rigor, unfortunately still hold' (Alfonsson et al., 2018, p. 219). Supervision's long lingering methodological issues—predominance of *ex post facto*, cross-sectional, correlational designs, small sample sizes, over-reliance on self-report measures, lack of psychometrically sound supervision measures, lack of experimental designs and longitudinal study—remain in force, temper any resulting research findings, compromise the field's advancement and leave us with this reality: we still know relatively little about the effects of supervision on supervisees, supervisors or clients (Hill & Knox, 2013). Numerous supervision scholars agree that substantive research benefits will only accrue

when supervision studies do a better job of the following: employing a pluralistic research approach that, in conjunction with *ex post facto*, cross-sectional study, also includes experimental investigation and randomisation as well as data-driven case studies (Borders, in Goodyear et al., 2016; Pelling et al., 2017); conducting multisite studies so as to increase sample size (Hill & Knox, 2013; Russell et al., 1984); developing psychometrically sound supervision measures and subjecting existing substandard measures to proper validation procedures (Ellis et al., 2008; Ellis & Ladany, 1997); employing a multithread, multitrait measurement approach, where self-report is but one type of data gathered (Ladany & Malouf, 2010; Lambert & Ogles, 1997); attending seriously to particular methodological issues (e.g. inadequate statistical power, threats to validity) during study development and execution (Ellis, 1991; Wampold & Holloway, 1997; cf. Schutt, 2012); examining supervision longitudinally (Holloway, 1992; Hill & Knox, 2013; cf. Potts, 2018); and studying supervision as a full-fledged triad, including the client and client outcomes as important parts of the supervisory picture (Ladany & Inman, 2012; Lambert & Ogles, 1997). Although those remedies have been identified for decades now, their implementation still lags far behind, negatively affects supervision research product and awaits realisation.

5.2.4 | Evidence-based supervision

Evidence-based supervision—built on a foundation of research evidence, expert consensus and practitioner expertise—can be understood practically as involving two fundamental features: (a) supervisor utilisation of research-supported supervisory methods and techniques during supervision; and (b) supervisor encouragement of supervisees to use empirically validated psychological interventions during therapy (Barrett et al., 2019; Milne, 2018). Evidence-based supervision (along with that which is competency-based) has indeed emerged as supervision's *Zeitgeist*, is generally touted as being a widely beneficial scientific stance to adopt (Bernard & Goodyear, 2019) and is considered crucial for good supervision practice and supervision's advancement (Reiser & Milne, 2012; O'Donoghue et al., 2018). But one contrary reality remains: any evidence-based supervision studies seem to be a most scant commodity (Barrett et al., 2019), highly desirable but largely unidentifiable in the literature. If supervision is an evidence-based or evidence-informed affair, or is at least moving increasingly in that evidence-based direction, where is the actual evidence that points to that being so (cf. Barrett et al., 2019; O'Donoghue & Tsui, 2015)?

5.3 | An emerged trend? An emerging trend? And a caution

5.3.1 | An emerged trend?

That effort has been increasingly made to bring some type of more objective appraisal procedure to the systematic review process

seems good. Whereas 13 of these 20 reviews involved the use of an appraisal procedure or criteria, almost all of those 13 reviews appeared during the last approximate decade alone. This would seem to reflect increasingly ongoing efforts to render the review process ever tighter and more rigorous. Those efforts seem destined and designed to bring benefit to any resulting supervision product.

5.3.2 | An emerging trend?

Efforts have also been increasingly made across these last 15 years to put more 'systematic' in the systematic review process and its product. Examples of that include the (2012) establishment of the journal, *Systematic Reviews* (Moher, Stewart, & Shekelle, 2015), release of the second edition of the *Cochrane Handbook of Systematic Reviews of Interventions* (Higgins & Thomas, 2019), introduction and widening embrace of the PRISMA (systematic review) guidelines (Moher et al., 2009), and the widening embrace of protocols that document the systematic review plan and procedures (e.g. PRISMA-P; Moher, Shamseer, et al., 2015; Shamseer et al., 2015) and the subsequent registration of those protocols (e.g. with PROSPERO, the international register launched in 2011; Page et al., 2018). All indications are that these systematic review changes are highly valued and here to stay because they (a) provide a purpose-driven organising structure for reviews to follow, (b) increase the rigour of the review process, (c) make prespecification explicit and (d) promote consistency, accountability, research integrity and transparency (Moher et al., 2009; Moher, Shamseer, et al., 2015). That three of the most recent supervision reviews surveyed here (Alfonsson et al., 2018; Barrett et al., 2019; Kühne et al., 2019) followed these systematic review and registration guidelines also seems good, reflecting efforts by supervision researchers to maintain the highest research standards, be highly rigorous in the process and produce the promised product for study and scrutiny. Future supervision systematic reviews would do well to follow the lead of these three efforts (Alfonsson et al., 2018; Barrett et al., 2019; Kühne et al., 2019).

5.3.3 | A caution

But just as there is much value to be found in their review approach, two of these reviews (Alfonsson et al., 2018; Barrett et al., 2019) also raise this question for consideration: At what point does restriction of allowed studies become so restrictive that finding meaningful results becomes a virtual impossibility? Four studies were reviewed in Barrett et al. (2019) and five in Alfonsson et al. (2018). It would seem that whereas more studies reviewed do not necessarily lead to quality findings, studies that are both more variable in content and fewer in number will be increasingly apt to lead to highly limited, less robust review findings. For example, Alfonsson et al. (2018), in commenting on their selected studies,

stated: 'The five studies varied greatly regarding background, scope and quality, making it difficult to compare the studies and summarize the findings...' (p. 219). The combination of study number and variability would seem a most important consideration to bear in mind when thinking about supervision study selection for a systematic review.

5.4 | Survey limitations/Cautions

Three particular limitations or cautions attend this survey of reviews and merit mention. First, being a single-author effort, this survey does not benefit from having another set or multiple sets of eyes and perspectives brought to bear on what was done. Although I have tried to be careful in my thinking about this survey and be meticulous in its conduct, readers may reasonably take issue with some aspects of what was done here and how it was done (e.g. my decision about what reviews to include and exclude). Second, review findings are a product of the reviewed studies; these 20 reviews, while providing generally good coverage of the supervision literature, did not cover the entire expanse of supervision investigations. It is possible that some of my conclusions, while consistent with what is in these 20 reviews, are not in sync with the broader body of supervision literature. Third, I have taken a highly critical view here; some might even say too much so. If nothing else, perhaps this highly critical perspective might serve as a stimulus for counterpoint discussion.

6 | CONCLUSION

Practising supervisors and supervisees tend to believe in, and have conviction about, the benefits, power and potential of supervision (Rast, Herman, Rousmaniere, Whipple, & Swift, 2017). But belief and conviction do not necessarily translate into empirical reality. It could be reasonably argued that supervision is still too much a product of 'proof by association', that supervision's methodological morass long lingers (e.g. compare Hill & Knox [2013] with Holloway & Hosford [1983]), and that if supervision is to ever substantially and fruitfully advance, research has to move beyond a preponderance of 'proof by association', more studiously address methodological shortcomings in research planning and execution and seriously abide by long-available, well-informed, admittedly challenging yet doable recommendations (e.g. Ellis, 1991; Hill & Knox, 2013; Russell et al., 1984). If supervision is to ever be evidence-based, then there is sore need for more, better and broader evidence.

ORCID

C. Edward Watkins  <https://orcid.org/0000-0001-9137-5526>

REFERENCES

Alfonsson, S., Parling, T., Spännargård, Å., Andersson, G., & Lundgren, T. (2018). The effects of clinical supervision on supervisees and patients in cognitive behavioral therapy: A systematic review. *Cognitive*

- Behaviour Therapy*, 47, 206–228. <https://doi.org/10.1080/16506073.2017.1369559>
- Bambling, M., King, R., Raue, P., Schweitzer, R., & Lambert, W. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research*, 16, 317–331. <https://doi.org/10.1080/10503300500268524>
- Barrett, J., Gonsalvez, C. J., & Shires, A. (2019). Evidence-based practice within supervision during psychology practitioner training: A systematic review. *Clinical Psychologist*, <https://doi.org/10.1111/cp.12196> (early view, available online)
- Bernard, J. M., & Goodyear, R. K. (2004). *Fundamentals of clinical supervision*, 3rd ed. Upper Saddle River, NJ: Pearson.
- Bernard, J. M., & Goodyear, R. K. (2014). *Fundamentals of clinical supervision*, 5th ed. Upper Saddle River, NJ: Pearson.
- Bernard, J. M., & Goodyear, R. K. (2019). *Fundamentals of clinical supervision*, 6th ed. Upper Saddle River, NJ: Pearson.
- Bogo, M., & McKnight, K. (2006). Clinical supervision in social work: A review of the research literature. *The Clinical Supervisor*, 24(1–2), 49–67. https://doi.org/10.1300/J001v24n01_04
- Borders, L. D. (2006). Snapshot of clinical supervision in counseling and counselor education: A five-year review. *The Clinical Supervisor*, 24(1–2), 69–113. https://doi.org/10.1300/J001v24n01_05
- Buus, N., & Gonge, H. (2009). Empirical studies of clinical supervision in psychiatric nursing: A systematic literature review and methodological critique. *International Journal of Mental Health Nursing*, 18(4), 250–264. <https://doi.org/10.1111/j.1447-0349.2009.00612.x>
- Carifio, M. S., & Hess, A. K. (1987). Who is the ideal supervisor? *Professional Psychology: Research and Practice*, 18, 244–250. <https://doi.org/10.1037/0735-7028.18.3.244>
- Carpenter, J., Webb, C. M., & Bostock, L. (2013). The surprisingly weak evidence base for supervision: Findings from a systematic review of research in child welfare practice (2000–2012). *Children and Youth Services Review*, 35(11), 1843–1853. <https://doi.org/10.1016/j.childyouth.2013.08.014>
- Cutcliffe, J. R., Sloan, G., & Bashaw, M. (2018). A systematic review of clinical supervision evaluation studies in nursing. *International Journal of Mental Health Nursing*, 27, 1344–1363. <https://doi.org/10.1111/inm.12443>
- Dawson, M., Phillips, B., & Leggat, S. (2013). Clinical supervision for allied health professionals: A systematic review. *Journal of Allied Health*, 42(2), 65–73.
- Ducat, W. H., & Kumar, S. (2015). A systematic review of professional supervision experiences and effects for allied health practitioners working in non-metropolitan health care settings. *Journal of Multidisciplinary Healthcare*, 8, 397–407. <https://doi.org/10.2147/JMDH.S84557>
- Ellis, M. V. (1991). Research in clinical supervision: Revitalizing a scientific agenda. *Counselor Education and Supervision*, 30(3), 238–251. <https://doi.org/10.1002/j.1556-6978.1991.tb01204.x>
- Ellis, M. V. (2010). Bridging the science and practice of clinical supervision: Some discoveries, some misconceptions. *The Clinical Supervisor*, 29(1), 95–116. <https://doi.org/10.1080/07325221003741910>
- Ellis, M. V., D'lusio, N., & Ladany, N. (2008). State of the art in assessment, measurement, and evaluation of clinical supervision. In A. K. Hess, K. D. Hess, & T. A. Hess (Eds.), *Psychotherapy supervision: Theory, research, and practice*, 2nd ed. (pp. 473–499). Hoboken, NJ: Wiley.
- Ellis, M. V., & Ladany, N. (1997). Inferences concerning supervisees and clients in clinical supervision: An integrative review. In C. E. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 447–507). New York, NY: Wiley.
- Ellis, M. V., Ladany, N., Kregel, M., & Schult, D. (1996). Clinical supervision research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology*, 43(1), 35–50. <https://doi.org/10.1037/0022-0167.43.1.35>
- Falender, C. A., & Shafranske, E. P. (2017). *Supervision essentials for the practice of competency based supervision*. Washington, DC: American Psychological Association.
- Farnan, J. M., Petty, L. A., Georgitis, E., Martin, S., Chiu, E., Prochaska, M., & Arora, V. M. (2012). A systematic review: The effect of clinical supervision on patient and residency education outcomes. *Academic Medicine*, 87, 428–442. <https://doi.org/10.1097/ACM.0b013e31824822cc>
- Forshaw, G., Sabin-Farrell, R., & Schröder, T. (2019). Supervisors' experience of delivering individual clinical supervision to qualified therapists: A meta-ethnographic synthesis. *Mental Health Review Journal*, 24(1), 51–68. <https://doi.org/10.1108/MHRJ-09-2018-0028>
- Freitas, G. J. (2002). The impact of psychotherapy supervision on client outcome: A critical examination of 2 decades of research. *Psychotherapy*, 39, 354–367. <https://doi.org/10.1037/0033-3204.39.4.354>
- Goodyear, R. K., Borders, L. D. A., Chang, C. Y., Guiffrida, D. A., Hutman, H., Kemer, G., ... White, E. (2016). Prioritizing questions and methods for an international and interdisciplinary supervision research agenda: Suggestions by eight scholars. *The Clinical Supervisor*, 35, 117–154. <https://doi.org/10.1080/07325223.2016.1153991>
- Goodyear, R. K., Bunch, K., & Claiborn, C. D. (2006). Current supervision scholarship in psychology: A five year review. *The Clinical Supervisor*, 24(1–2), 137–147. https://doi.org/10.1300/J001v24n01_07
- Goodyear, R. K., & Guzzardo, C. R. (2000). Psychotherapy supervision and training. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology*, 3rd ed. (pp. 83–108). New York, NY: Wiley.
- Higgins, J. P., & Thomas, J. (Eds.) (2019). *Cochrane handbook for systematic reviews of interventions*. West Sussex, UK: Wiley.
- Hill, C. E., & Knox, S. (2013). Training and supervision in psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*, 6th ed. (pp. 775–811). Hoboken, NJ: Wiley.
- Holloway, E. L. (1992). Supervision: A way of teaching and learning. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology*, 2nd ed. (pp. 177–214). New York, NY: Wiley.
- Holloway, E. L., & Hosford, R. E. (1983). Towards developing a prescriptive technology of counselor supervision. *The Counseling Psychologist*, 11(1), 73–77. <https://doi.org/10.1177/0011000083111012>
- Inman, A. G., Hutman, H., Pendse, A., Devdas, L., Luu, L., & Ellis, M. V. (2014). Current trends concerning supervisors, supervisees, and clients in clinical supervision. In C. E. Watkins, & D. Milne (Eds.), *Wiley international handbook of clinical supervision* (pp. 61–102). Oxford, UK: Wiley.
- Kühne, F., Maas, J., Wiesenthal, S., & Weck, F. (2019). Empirical research in clinical supervision: A systematic review and suggestions for future studies. *BMC Psychology*, 7(1), 1–11. <https://doi.org/10.1186/s40359-019-0327-7>
- Ladany, N., & Inman, A. G. (2012). Training and supervision. In E. M. Altmaier, & J. Hansen (Eds.), *Oxford handbook of counseling psychology* (pp. 179–207). Oxford, NY: Oxford University Press.
- Ladany, N., & Malouf, M. A. (2010). Understanding and conducting supervision research. In N. Ladany, & L. J. Bradley (Eds.), *Counselor supervision* (pp. 353–388). New York, NY: Routledge.
- Lambert, M. J., & Ogles, B. M. (1997). The effectiveness of psychotherapy supervision. In C. E. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 421–446). New York, NY: Wiley.
- MacDonald, J., & Ellis, P. M. (2012). Supervision in psychiatry: Terra incognita? *Current Opinion in Psychiatry*, 25, 322–326. <https://doi.org/10.1097/YCO.0b013e328354ecc>
- Milne, D. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology*, 46(4), 437–447. <https://doi.org/10.1348/014466507X197415>
- Milne, D. (2016). Guiding CBT supervision: How well do manuals and guidelines fulfil their promise? *The Cognitive Behaviour Therapist*, 9, <https://doi.org/10.1017/S1754470X15000720>

- Milne, D. L. (2018). *Evidence-based CBT supervision: Principles and practice*, 2nd ed. Chichester, UK: Wiley.
- Milne, D. L., & James, I. (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology*, 39(2), 111–127. <https://doi.org/10.1348/014466500163149>
- Milne, D. L., Leck, C., James, I., Wilson, M., Procter, R., Ramm, L., & Weetman, J. (2012). High fidelity in clinical supervision research. In I. Fleming, & L. Steen (Eds.), *Supervision and clinical psychology: Theory, practice and perspectives*, 2nd ed. (pp. 142–158). London, UK: Routledge.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., and the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135> (PRISMA guidelines currently being updated; www.prisma-statement.org/)
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1. <https://doi.org/10.1136/bmj.g7647>
- Moher, D., Stewart, L., & Shekelle, P. (2015). All in the family: Systematic reviews, rapid reviews, scoping reviews, realist reviews, and more. *Systematic Reviews*, 4(1), 1. <https://doi.org/10.1186/s13643-015-0163-7>
- Mulrow, C. D. (1994). Systematic reviews: Rationale for systematic reviews. *BMJ*, 309(6954), 597–599. <https://doi.org/10.1136/bmj.309.6954.597>
- Munn, Z., Peters, M. D., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Medical Research Methodology*, 18(1), 5. <https://doi.org/10.1186/s12874-017-0468-4>
- Newman, D. S., Simon, D. J., & Swerdluk, M. E. (2019). What we know and do not know about supervision in school psychology: A systematic mapping and review of the literature between 2000 and 2017. *Psychology in the Schools*, 56, 306–334. <https://doi.org/10.1002/pits.22182>
- O'Donoghue, K., Ju, P. W. Y., & Tsui, M. (2018). Constructing an evidence-informed social work supervision model. *European Journal of Social Work*, 21, 348–358. <https://doi.org/10.1080/13691457.2017.1341387>
- O'Donoghue, K., & Tsui, M. (2015). Social work supervision research (1970–2010): The way we were and the way ahead. *British Journal of Social Work*, 45, 616–633. <https://doi.org/10.1093/bjsw/bct115>
- Page, M. J., Shamseer, L., & Tricco, A. C. (2018). Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Systematic Reviews*, 7(1), 32. <https://doi.org/10.1186/s13643-018-0699-4>
- Park, E. H., Ha, G., Lee, S., Lee, Y. Y., & Lee, S. M. (2019). Relationship between the supervisory working alliance and outcomes: A meta-analysis. *Journal of Counseling and Development*, 97(4), 437–446. <https://doi.org/10.1002/jcad.12292>
- Pelling, N., Abbott, D., & Lack, C. W. (2017). The state of supervision scholarship in the 21st century. In N. Pelling, & P. Armsrtong (Eds.), *The practice of counselling and clinical supervision*, 2nd ed. (pp. 109–116). Samford Valley, Qld: Australian Academic Press.
- Pott, T. (2018). *How highly effective psychotherapy supervisors supervise: A longitudinal study of supervisees' experiences (unpublished doctoral dissertation)*. Edmonton, CA: University of Alberta.
- Rast, K. A., Herman, D. J., Rousmaniere, T. G., Whipple, J. L., & Swift, J. K. (2017). Perceived impact on client outcomes: The perspectives of practicing supervisors and supervisees. *SAGE Open*, 7(1), <https://doi.org/10.1177/2158244017698729>
- Reiser, R. P., & Milne, D. (2012). Supervising cognitive-behavioral psychotherapy: Pressing needs, impressing possibilities. *Journal of Contemporary Psychotherapy*, 42(3), 161–171. <https://doi.org/10.1007/s10879-011-9200-6>
- Reiser, R. P., & Milne, D. L. (2014). A systematic review and reformulation of outcome evaluation in clinical supervision: Applying the fidelity framework. *Training and Education in Professional Psychology*, 8(3), 149–157. <https://doi.org/10.1037/tep0000031>
- Rønnestad, M. H., Orlinsky, D. E., Schröder, T. A., Skovholt, T. M., & Willutzki, U. (2019). The professional development of counsellors and psychotherapists: Implications of empirical studies for supervision, training and practice. *Counselling and Psychotherapy Research*, 19, 214–230. <https://doi.org/10.1002/capr.12198>
- Russell, R. K., Crimmings, A. M., & Lent, R. W. (1984). Counselor training and supervision: Theory and research. In S. D. Brown, & R. W. Lent (Eds.), *Handbook of counseling psychology* (pp. 625–681). New York, NY: Wiley.
- Scaife, J. (2019). *Supervision in clinical practice: A practitioner's guide*, 3rd ed. London, UK: Routledge.
- Schutt, M. A. (2012). *Replication and extension of Ellis, Ladany, Krenzel, and Shult (1996); Clinical Supervision and Research from 1981 to 1993: A Methodological Critique (unpublished doctoral dissertation)*. Bethlehem, PA: Lehigh University.
- Sewell, K. M. (2018). Social work supervision of staff: A primer and scoping review (2013–2017). *Clinical Social Work Journal*, 46, 252–265. [doi.10.1007/s10615-018-0679-0](https://doi.org/10.1007/s10615-018-0679-0)
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., ... Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, 349, g7647. <https://doi.org/10.1136/bmj.g7647>
- Simpson-Southward, C., Waller, G., & Hardy, G. E. (2017). How do we know what makes for “best practice” in clinical supervision for psychological therapists? A content analysis of supervisory models and approaches. *Clinical Psychology & Psychotherapy*, 24, 1228–1245. <https://doi.org/10.1002/cpp.2084>
- Snowdon, D. A., Leggat, S. G., & Taylor, N. F. (2017). Does clinical supervision of healthcare professionals improve effectiveness of care and patient experience? A systematic review. *BMC Health Services Research*, 17(1), 786. <https://doi.org/10.1186/s12913-017-2739-5>
- Tsui, M. S. (1997). Empirical research on social work supervision: The state of the art (1970–1995). *Journal of Social Service Research*, 23(2), 39–54. https://doi.org/10.1300/J079v23n02_03
- Wampold, B. E., & Holloway, E. L. (1997). Methodology, design, and evaluation in psychotherapy supervision research. In C. E. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 11–27). New York, NY: Wiley.
- Watkins, C. E. Jr (2011). Does psychotherapy supervision contribute to patient outcomes? Considering thirty years of research. *The Clinical Supervisor*, 30(2), 235–256. <https://doi.org/10.1080/07325223.2011.619417>
- Weerasekera, P. (2013). The state of psychotherapy supervision: Recommendations for future training. *International Review of Psychiatry*, 25(3), 255–264. <https://doi.org/10.3109/09540261.2013.769431>
- Wheeler, S., & Richards, K. (2007). The impact of clinical supervision on counselors and therapists, their practice and their clients: A systematic review of the literature. *Counselling & Psychotherapy Research*, 7, 54–65. <https://doi.org/10.1080/14733140601185274>
- White, E. (2017). Claims to the benefits of clinical supervision: A critique of the policy development process and outcomes in New South

Wales, Australia. *International Journal of Mental Health Nursing*, 26(1), 65–76. <https://doi.org/10.1111/inm.12292>

Wilson, H. M. N., Davies, J. S., & Weatherhead, S. (2016). Trainee therapists' experiences of supervision during training: A meta-synthesis. *Clinical Psychology & Psychotherapy*, 23(4), 340–351. <https://doi.org/10.1002/cpp.1957>

How to cite this article: Watkins CE. What do clinical supervision research reviews tell us? Surveying the last 25 years. *Couns Psychother Res*. 2019;00:1–19. <https://doi.org/10.1002/capr.12287>

AUTHOR BIOGRAPHY

C. Edward Watkins, Jr., Ph.D., is Professor of Psychology at the University of North Texas. His primary professional interests focus on psychotherapy supervision and integrative and psychoanalytic theories, practice, and research. He is editor of the *Handbook of Psychotherapy Supervision* (1997), co-editor (with Derek Milne) of the *Wiley International Handbook of Clinical Supervision* (2014), and volume coordinator (with Loredana-Ileana Vișcu) of *Metode si Tehnici de Interventie in Consilierea si Psihoterapia Copilului si Adolescentului* (2019). He currently serves as Associate Editor (the Editor being Dr. Vișcu) of the *International Journal of Supervision in Psychotherapy*. He is a Fellow of Divisions 29 (Psychotherapy) and 17 (Counseling Psychology) of the American Psychological Association.