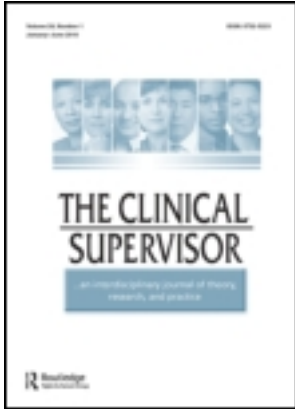


This article was downloaded by: [Monash University Library]

On: 07 September 2011, At: 20:42

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## The Clinical Supervisor

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/wcsu20>

### Evidence-Based Training for Clinical Supervisors: A Systematic Review of 11 Controlled Studies

Derek L. Milne<sup>a</sup>, Alia I. Sheikh<sup>a</sup>, Sue Pattison<sup>a</sup> & Andrew Wilkinson<sup>a</sup>

<sup>a</sup> Newcastle University, Newcastle upon Tyne, United Kingdom

Available online: 09 May 2011

To cite this article: Derek L. Milne, Alia I. Sheikh, Sue Pattison & Andrew Wilkinson (2011): Evidence-Based Training for Clinical Supervisors: A Systematic Review of 11 Controlled Studies, *The Clinical Supervisor*, 30:1, 53-71

To link to this article: <http://dx.doi.org/10.1080/07325223.2011.564955>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## **Evidence-Based Training for Clinical Supervisors: A Systematic Review of 11 Controlled Studies**

DEREK L. MILNE, ALIA I. SHEIKH, SUE PATTISON, and  
ANDREW WILKINSON

*Newcastle University, Newcastle upon Tyne, United Kingdom*

*Clinical supervision training is now considered essential, but has not been given serious attention. Therefore, in this study a systematic review of the research literature on clinical supervision training is undertaken. Eleven controlled studies were located and assessed to identify evidence-based practices in supervisor training. Addressing the two questions posed by Whitman, Ryan, and Rubinstein (2001), we found that these 11 studies provided empirical support for supervisor training (i.e., 15 specific elements of training had empirical support: primarily corrective feedback, educational role-play, and observational learning). Recommendations are outlined, based on these 11 studies, combined with narrative reviews, surveys, and expert consensus.*

*KEYWORDS* field instructor education, meta-analytic review, supervisor training, systematic review

### INTRODUCTION

Clinical supervision is now deemed a core professional competency within the mental health field (Falender et al., 2004), and Kilminster and Jolly (2000) identified effective clinical supervision as essential for patient and

---

The authors are grateful to Helen Aylott, Christopher Dunkerley, Helen Fitzpatrick, and Sarah Wharton for their initial work on this review (including literature searching); Kathryn Mark and Barbara Mellors for help in word processing this document; and to Tom Cliffe for help with searching for papers. This study was funded by the Higher Education Academy, Psychology Network.

Address correspondence to Derek L. Milne, Doctorate in Clinical Psychology, School of Psychology, Ridley Building, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. E-mail: D.L.Milne@ncl.ac.uk

practitioner safety. Yet the training of supervisors has long been problematic. For example, Spence, Wilson, Kavanagh, Strong, and Worrall (2001) examined clinical supervision in four mental health professions, noting that, despite the importance of the topic and the large amount of literature, “we have little information to guide us as to the most effective ways of training supervisors” (p. 135). Fleming, Oliver, and Bolton (1996) noted the “need for continued research into the specification, measurement, and training of supervisory skills” (p. 6), while Milne and James (2002) noted a paradox: the salience of supervision as the cornerstone in training therapists, but no concerted effort to train supervisors. Given that supervision may be as challenging a competence to develop as therapy (Watkins, 1995), Milne and James (2002) conjectured that much supervision is still being practiced incompetently, as has been suggested previously (Binder, 1993). In the counseling and psychotherapy profession, where clinical supervision is a requirement for ethical practice (British Association for Counselling and Psychotherapy [BACP], 2007; United Kingdom Council for Psychotherapy [UKCP], 2007), accreditation of therapist training courses takes place and a core curriculum is being developed (BACP, 2008). However, despite a number of supervision training courses being available, no apparent consensus exists on what constitutes effective supervisor training. According to Falender and Shafranske (2004), the majority of psychologists have not actually had any formal training in supervision. Worrying consequences of this “barren scape” are thought to include supervisors fleeing the role, supervisors who are overly influenced by the “tyranny of niceness” (Fleming, Gone, Diver, & Fowler, 2007), and supervisors who are “passive to the point of destruction, who are judgemental, authoritarian, and demanding to the point of sadism” (Hoffman, 1994, p. 24). This is of concern when the influence of supervision on the practitioner-client working alliance and client progress is documented (Lambert, 2006).

The problems of supervisor training have been noted for many years. Consistent and growing support for supervisor training has existed since the first workshops for clinical supervisors were offered to social workers in 1911 (Kadushin, 1985). In the 1980s, for example, a review by Loganbill and Hardy (1983) noted that increased attention was being accorded to supervisor training, and they outlined what they regarded as a comprehensive training program. But in a survey conducted in 1981 (Stanton et al., 1981, cited in Loganbill and Hardy, 1983), 19 of the 55 (35%) clinical programs that responded had no method of training supervisors. By 1998, a survey of 10 mental health professions in the United Kingdom indicated that all 10 endorsed the policy that supervisors should receive training (Milne, 1998). For example, discussing psychiatric nursing, Butterworth, Faugier, and Burnard (1998) noted that “training for...supervisors...is considered vital...” (p. 169). Since then, surveys of different professional groups have indicated that supervisor training has been increasing gradually. For

example, in the United States, Scott, Ingram, Vitanza, and Smith (2000) surveyed 123 respondents (each representing a different counseling or clinical psychology training program accredited by the American Psychological Association), who indicated that 84% of these programs offered some form of training in supervision. Similarly, multi-professional members of the British Association for Behavioural and Cognitive Psychotherapy (BABCP) were surveyed about supervisor training at around the same time, and 108 of the 170 respondents (64%) also reported having received some form of training (Townend, Iannetta, & Freeston, 2002). Workshops in supervision were the most frequent training method reported. Thankfully, these data question the view that “many, if not most, supervisors practice without the benefit of education, training or supervision” (Falender & Shafranske, 2004, p. 7), but problems clearly remain: “...to date, there are still no specific requirements for training for supervisors to ensure supervisory proficiency in APA accredited psychology doctoral and internship programs” (Whitman, Ryan, & Rubenstein, 2001, p.156), a position that they believe also applies to psychiatry and graduate social work programs. However, this gap is slowly being recognized in professional psychology, as illustrated by the formal incorporation of supervision competencies within a core competencies model (American Psychological Association [APA], 2007).

This situation is mirrored in the counseling and psychotherapy professions. However, in the United States, one exception to this is the training of counselors at the doctoral level, in which supervision training *is* required according to the Council for Accreditation of Counseling and Related Educational Programs (CACREP) accreditation standards (2001). This is not the case in the United Kingdom, where supervision training is a desired qualification but not an essential requirement. By contrast, clinical psychologists within the United Kingdom are explicitly required to receive appropriate training in supervision, training that moreover addresses “effective” methods of supervision (British Psychological Society, 2007). This current British stance is similar to a recent U.S. expert consensus statement about supervision competencies (Falender et al., 2004), which stated that supervisors should engage in course work (to develop relevant knowledge and skills) and receive supervision of their supervision (to include some form of direct observation and critical feedback). Finally, some professional groups (e.g., BACP) have added accreditation schemes for supervisors. In summary, policy directives from most but not all mental health professional bodies appear to require supervisor training, and recent surveys of supervisors suggest that workshops now reach between 64% and 84% of supervisors.

While this represents a great improvement in the availability of training since 1981, some still assert that “good supervision, like love... cannot be taught” (Hawkins & Shohet, 2000, p. 195). In keeping with these views, Kilminster and Jolly (2000) noted that training is rarely empirically or theoretically grounded, though in a guarded conclusion they also noted that “there

is some evidence that training can have a positive effect on supervisors” (p. 835). Less optimistically, Spence and colleagues (2001) concluded that there was only “tentative evidence to suggest that the training of supervisors could produce a change in supervisor practices” (p. 149), based on reviewing literature from occupational therapy, social work, speech pathology, and clinical psychology. At least, following years of recommendations (see Loganbill and Hardy, 1983), experts have latterly achieved a consensus on both the content and methods of supervisor training (Falender et al., 2004; Kaslow et al., 2004; Milne, Scaife, & Cliffe, 2009); however, survey data suggest wide variability in what is actually covered within workshops, at least in the United Kingdom (Fleming, 2004). In addition, the evidence base on which to design workshops is thought to be weak (Ellis & Ladany, 1997), the curriculum is not comprehensive (Borders et al., 1991), and agreement over which outcome measures could be used to assess competence in supervision is lacking (Kavanagh, Spence, Wilson, & Crow, 2002). Also, while some believe that clinical benefits are the acid test of successful supervision (and presumably in turn supervisor training), others argue that the primary goal of supervision is to foster supervisee development (e.g., Wampold & Holloway, 1997).

In summary, although there is a consensus on the broad content and methods of supervisor training, there is little clarity regarding precisely how supervisors should be trained and limited evaluation, yet a large number of unsystematic recommendations for supervisor training (Kavanagh et al., 2002). For example, there appears to be only one example of a manual to guide the training of supervisors (Neufeldt, 1994), itself based only on a simple user satisfaction evaluation. One perspective on this situation is that there exists a healthy diversity within supervisor training, possibly reflecting the different professional groups involved, the local circumstances, or the participants’ learning styles. However, our perspective is that we have entered an era in which evidence-based practices are increasingly required by employers and purchasers (e.g., science-informed supervision: Falender & Shafranske, 2004), making the present emphasis timely for a growing number of supervisor trainers.

In order to provide a constructive response to these supervisor training problems, the present review overcomes problems with the highly variable rigor of the research that does exist by adopting a selective form of systematic (meta-analytic) reviewing, called the “best evidence synthesis” (BES: Petticrew & Roberts, 2006). This strategy has been successful in prior attempts to answer pragmatic questions in the education and health fields (Petticrew & Roberts, 2006), and has helped to clarify an initial empirical basis for clinical supervision (Milne, 2007; Milne, Aylott, Fitzpatrick, & Ellis, 2008). One objective is therefore to utilize the BES approach to review controlled quantitative research on supervisor training, in order to determine the current answer to two questions posed by Whitman, Ryan, and Rubenstein

(2001): (1) What empirical support is there for supervisor training?, and (2) Which elements (content and methods) of supervisor development appear to be supported? In addition, we will then attempt to (3) outline emergent recommendations for the training of supervisors and (4) specify research directions related to the training of supervisors.

## METHOD

### Search Strategy

A systematic search of the primary research literature was conducted, in which three databases (Medline, PsycInfo, and ASSIA) were searched using the keywords “clinical supervision” and “clinical supervisor.” This yielded hundreds of articles. The search was then narrowed by adding the term “empirical.” The titles, abstracts, and, if necessary, the full documents were then examined in order to select papers for inclusion, based on specific inclusion criteria. This reduced the list to 11 studies. Further search methods were utilized to ensure a thorough procedure, such as hand-searching a number of relevant peer-reviewed journals and consulting experts for additional studies. These methods revealed no new studies.

The inclusion and exclusion criteria were based on the NICE(R) Manual for Systematic Reviews of the Clinical Supervision Literature (Version 11; Milne & James, 2000). The eight inclusion criteria are intended to pinpoint the strongest available evidence base. According to these criteria, papers should be included if they

1. directly address clinical supervision (i.e., excluding related activities, such as training and consultancy) in the mental health field;
2. include an empirical intervention or experiment (good internal validity; excluding research designs that did not allow causal inferences to be drawn);
3. address the behaviors of the trainers and/or supervisors to provide an indication of the independent variable(s);
4. address the behaviors of the students and/or supervisees to provide an indication of the dependent variable(s);
5. present data from real clinical training practice (e.g., as compared to simulations or analog studies);
6. are relevant to routine clinical practice (good external validity);
7. demonstrate a clear benefit of training in clinical supervision (i.e., improvement of at least 20% to provide a sample of effective/successful manipulations); and
8. are reported in peer-reviewed scientific and professional journals over the past 20 years (1998–2008).

To illustrate, criterion 1 required that studies were concerned with clinical supervision, which was defined as “*The formal provision* (i.e., sanctioned by relevant organization/s); *by senior/qualified health practitioners* (or similarly experienced staff) *of an intensive education* (general problem-solving capacity; developing capability) *and/or training* (competence enhancement) *that is case-focused and which supports, directs, and guides* (including ‘restorative’ and/or ‘normative’ topics, addressed by means of professional methods, including objective monitoring, feedback, and evaluation) *the work of junior colleagues* (supervisees).” This incorporates the summative evaluation function (“gatekeeping”), and derives from the widely quoted definition by Bernard and Goodyear (2004), and is elaborated and justified in Milne (2007).

Of course, the highly focused application of such narrow inclusion criteria will inevitably exclude much of the potentially relevant literature, such as textbooks, as well as the kind of qualitative material that they contain. These other sources were knowingly excluded, as our aim was to ascertain what was known about supervisor training from the original and most relevant empirical sources, although below we do refer to prior literature reviews in discussing the empirical findings. Our rationale was to try to isolate a seam of the best available empirical evidence to complement the narrative style of review.

### Sample of Empirical Papers

Eleven studies were selected, based on this search methodology and the inclusion criteria listed above. These 11 studies represent the sample of studies on which the remainder of this review is based. The studies are: Bambling, King, Raue, Schweitzer, and Lambert (2006); Barrow and Domingo (1997); Busari, Scherpbier, Van der Vleuten, and Essed (2006); Ducharme, Williams, Cummings, Murray, and Spencer (2001); Fleming, Oliver, and Bolton (1996); Jensen, Parsons, and Reid (1998); McMahon and Simons (2004); Parsons, Reid, and Green (1993); Parsons and Reid (1995); Reid and colleagues (2003); and Richman, Riordan, Reiss, Pyles, and Bailey (1988).

### Procedure and Measure

All papers were next analyzed according to a study coding manual, to ensure that the systematic review was conducted in a reliable way. The manual is based on the United Kingdom’s National Institute for Clinical Excellence (NICE) approach to the systematic review (National Health Service’s Centre for Reviews and Dissemination, 2001). A manual was used to enhance the reliability of the analysis; in it we had to adapt the standard NICE manual, as it has a strongly medical emphasis, including some irrelevant material

(in relation to the present review). One feature of this coding system is the classification of the methodological quality of studies. Based on detailed prior ratings (e.g., the use of psychometrically sound instruments), the reviewers grade studies as fulfilling few of these criteria (“-”); fulfilling some of them (“+”); or fulfilling all or most of the criteria, so that the main threats are adequately addressed, and the study is judged to be fundamentally sound (coded as “++”). The NICE method was revised (R), as this primarily medical orientation to reviewing was clearly inappropriate in some respects for summarizing research on supervision. A copy of the NICE(R) manual is available from the first author.

### Reliability of Coding

All 11 studies reviewed were coded by one of the authors, by reference to the manual. In addition, in order to ensure that this coding was conducted reliably, all authors independently coded the same two randomly selected studies, one at the start (Busari et al., 2006), and a second at the end of the coding process (Sundin, Ogren, & Boethius, 2008). It should be noted that both were originally included within the review sample; it was only after some final checking of the sample that Sundin and colleagues’ (2008) study was excluded, as it was not a controlled study. In other respects, it was thought to be sufficiently similar to the review sample to be a representative test of our inter-rater agreement. The agreement calculations were based on the number of agreements between the first author (as the most experienced coder, with a history of demonstrated reliability: e.g., Milne et al., 2008) and each other author in turn, using all 48 criteria within the coding manual. Agreement at the outset ranged from 71% to 89% across the four raters (mean agreement: 81%), while at the end of the coding process this ranged from 80% to 83% (mean agreement: 82%). This indicates that the coding of the 11 papers was conducted in a reliable way, throughout the review period.

## RESULTS

Only 2 of the 11 studies offered explicit definitions of clinical supervision, but 9 of them did both specify and measure their supervision variables, an encouraging degree of precision. All 11 studies were judged to have hypotheses or study objectives that were logically consistent with the reasoning within their Introduction sections, so these studies appeared to have hypothesis validity (Wampold, David, & Good, 1990).

The most frequently taught workshop topic was feedback, which was cited in all 11 studies. This was followed by agenda-setting (including goal-setting, which was cited in six studies: 55%). Six further supervision activities were all ranked third: educational needs assessment; evaluation;



modeling; observation; prompting (e.g., suggestions); and time management (all cited in three studies: 27%). Next came teaching (didactic instruction) and educational role-play, both cited twice. A further 10 activities were specified in only 1 of our 11 studies. For convenience, these can be bracketed in terms of the three broad functions of supervision (Kadushin, 1985): “formative” (including promoting insight; behavioral rehearsal; discussion; question-and-answer sessions), “restorative” (including satisfaction and comfort with therapy work; support), and “normative” (including troubleshooting, providing a rationale, and promoting generalization). Thus, a total of 20 different supervision topics were cited within our sample of 11 controlled studies. The total frequency with which these activities were utilized was 49, which indicates that, on average, supervision entailed the use of 4.5 methods per study. This suggests that it was treated as a complex process, requiring a blend of activities to succeed.

Five of the research designs that were employed within these 11 studies were control group evaluations (e.g., randomized controlled trials, crossover designs) and the remaining 6 were multiple baseline  $N=1$  designs. Overall, nine of the studies were judged to have achieved very good internal validity (defined as a study in which all major threats were thought to have been controlled), the remaining two being graded as good (the inference that supervisor training had produced the obtained effect was judged to be valid). This is an impressive degree of rigor, given that these studies were all conducted within naturalistic service settings, which improves external validity (including application to real-life settings and acceptability to practitioners). The studies were largely based on supervisors who were mental-health- and/or human-service-related professionals. While all studies described characteristics of the 145 supervisors and 251 supervisees involved, 3 of the 11 studies also included a description of the clients ( $N=150$ ), although only 2 studies actually assessed the clinical effects. The recruitment of the samples included volunteers (directly from student and patient groups and through advertisements; and convenience samples), but in some studies the method of recruitment was unclear. The age of participants was reported as either a range or mean, or was unclear. For clients, the age range was reported as 8 to 23 years in one study (Ducharme et al., 2001; this is the study that did not evaluate clinical outcomes); in a second study the mean was 39.1 (Bambling et al., 2006), while in a third no age or gender information was provided (Parsons et al., 1993). Supervisees' ages ranged from 20 to 61 years. Supervisors' ages are reported as means of 30 and 49.9. In terms of gender, all but four of the clients were male; supervisees included 50 males and 107 females; supervisors were predominantly female ( $N=81$ ) with 32 males. In some studies the participants' gender was not reported clearly. Where applicable, cases and controls were mostly taken from comparable populations, with the same inclusion and exclusion criteria being used for both cases and controls. In summary, although the 11 studies contained some helpful descriptive data,

some data were missing and some marked biases (e.g., nearly all clients were male) existed, which makes this a rather unsatisfactory account of the demographics and an unrepresentative sample. Future work should report demographic details and aim for representative samples.

The methods that were used to train the 145 supervisors in these 11 studies mirrored those used within supervision. Table 1 summarizes the 15 distinct methods reported, indicating that feedback was again employed in all studies, followed by educational role-play and modeling (both cited in 73% of studies). Comparison with the methods used in these same studies to specify supervision indicates that 12 of the 15 methods in Table 1 were also cited as part of supervision (the exceptions were the use of guided reading, homework, and quizzes in supervisor training). With a total frequency of 56 supervisor training methods used across all 11 studies, there is also a marked parallel in the deploying of a blend of methods—on average 5.1 per study. Overall, the methods tended to develop competencies in four separate areas: (1) skill acquisition and behavioral change (e.g., ability to adjust behavior to the needs of supervisees); (2) knowledge (e.g., increase in theoretical and conceptual knowledge); (3) generalization of skills (e.g., from simulated to clinical settings); and (4) supervisees' satisfaction (e.g., increase in confidence, sense of independence).

The impact of this training was assessed primarily in terms of the participating supervisors' development (all 11 studies) and supervisee development (assessed in 10 studies). Other popular measures were assessments of transfer (mostly across settings, from the workshop to the workplace), followed by assessments of the supervisors' reactions to their training (6 studies). Only two of the studies examined clinical effects with patients (Bambling et al., 2006; Parsons et al., 1993).

The impact of training on these outcomes was expressed via a standardized 7-point rating of the overall effectiveness of the training, following the procedure reported in Milne and colleagues (2008). The rating could range from  $-3$  to  $+3$ , defined in terms of a spectrum of possible outcomes, from very negative effects to very positive effects across all assessed outcome variables. This rating was anchored in the proportion of study participants who were affected. Thus, a study that reported that at least 75% of participants achieved highly significant improvements in their score on a quiz would be rated as a  $+3$  impact (assuming no accompanying negative results).

A moderately large positive impact was reported in the majority of instances across these participant groups (e.g., five of the studies reported such an impact on the supervisors). The mean impacts were rated as 2.0 for the supervisors and the supervisees, and 2.5 for the patients. Expressed in percentage terms, this represents 67% impact for the supervisors and supervisees and 83% for the patients. These results indicate that the blend of training and supervisory methods, as outlined above, were effective in facilitating supervisor and supervisee (therapist) development, which in turn

was associated with patient benefits. However, it should be recalled that only two studies measured such benefits directly (Bambling et al., 2006; Parsons et al., 1993), so this clinical outcome estimate should be treated with great caution.

These impacts subsumed follow-up assessments in 4 of the studies (range: 1–40 months; mean interval: 7.8 months) and analyses of generalization across settings in 6 studies, and across people in 3 studies. There appeared to be no analyses of response generalization.

## DISCUSSION

We first set out to clarify the answers to two questions posed by Whitman and colleagues (2001): (1) What empirical support was there for supervisor training? (2) Which elements of such training were supported by evidence? In relation to the question about empirical support, we located 11 controlled studies that were judged by the present authors to have hypothesis validity (Wampold et al., 1990) and good internal and external validity. Together they provided a detailed and informative account of clinical supervision, suggesting that it consisted of up to 20 interwoven variables, primarily based on an educational needs assessment that leads to agenda-setting, various developmental activities designed to achieve these objectives (e.g., modeling, prompting, and direct observation), and concluding with the provision of feedback. Studies utilized a blend of such methods—on average 4.5 per study. The impact of this training was estimated at 67% for both the supervisors and their supervisees, with an impact on their patients (in the two relevant studies) of 83%. This suggests that clear empirical support exists for supervisor training, at least in a preliminary sense, given that this conclusion is based on a small sample of only 11 controlled studies. However, as the majority of the training methods (i.e., 12; see Table 1) were utilized in less than half of the studies, there is a suggestion that only feedback, educational role-play, and modeling (live/video demonstration) should be considered as provisionally supported methods.

Turning to Whitman and colleagues' (2001) second question (Which elements of training had empirical support?), these supervisors were trained by means of 15 different educational methods, primarily feedback, role-play, and modeling (see Table 1 for a summary). Mirroring the supervision methods, the training also consisted of a blend of relevant methods—on average 5.1 per study. Because these methods were associated with the good impacts noted above, we infer that they are indeed supported by the available empirical evidence.

In relation to the above answers to both of Whitman and colleagues' (2001) questions, it is probably worth clarifying that they draw heavily on the cognitive-behavioral (CBT) model, itself based on the scientific study

**TABLE 1** Summary of the 15 Methods Used to Train the 145 Supervisors in a Carefully Selected Sample of 11 Controlled Studies

Rank	Training method	Frequency	Percentage of studies
1	Feedback	11	100
2.5	Educational role-play	8	73
2.5	Modeling (live/video demonstration)	8	73
4	Teaching (verbal instruction)	5	45
5	Written assignments	4	36
7	Behavioral rehearsal	3	27
7	Providing a rationale	3	27
7	Guided reading	3	27
9.5	Discussion	2	18
9.5	Educational needs assessment	2	18
9.5	Direct observation	2	18
9.5	Questions and answers	2	18
13	Agenda-setting	1	9
14	Homework assignments	1	9
15	Quiz	1	9

of learning (Mayer, 2008). This probably follows from our search for the best available empirical evidence, which tends to be associated with CBT because it is an avowedly evidence-based approach. In any case, the CBT model represents a distinctive theoretical approach, in relation to such alternatives as constructivist and problem-based approaches (Kirschner, Sweller, & Clark, 2006). It is therefore particularly relevant to researchers who seek to develop supervision by empirical means.

Our third objective was to offer emergent recommendations for supervisor training. In this respect it is convenient that our findings and conclusions are consistent with prior reviews and expert consensus statements. For instance, compare these evidence-based methods of supervisor training with those put forward more than a quarter of a century ago by Loganbill and Hardy (1983). According to them, three necessary components of training exist: theoretical content, simulated experience, and in vivo practice with supervisees. Based on the recommendations of Loganbill and Hardy (1983) and other subsequent suggestions, Russell and Petrie (1994) similarly suggested a combination of didactic and experiential components. The didactic aspect was to include theoretical models, relevant research, plus ethical and professional issues. The experiential training aspect required opportunities to rehearse supervision with therapists in training, something that they regarded as a "critical element in an effective supervisory training programme" (p. 39). A systematic review of the closely related field of faculty development also endorsed the need for experiential learning, feedback, effective peer and colleague relationships, and well-designed interventions that followed the principles of teaching and learning (Steinert et al., 2006). The recent expert consensus statements endorse this emphasis on didactic and experiential learning (Falender et al., 2004; Kaslow et al., 2004). Thankfully, useful

materials exist to support training that attempts to achieve this balance (Baltimore & Crutchfield, 2003; Fall & Sutton, 2004; Milne, 2009). In summary, the present findings are consistent with prior narrative reviews and with expert consensus, but bolster these opinions with an emphasis on the empirical evidence base. Taken together, these sources suggest that there is now preliminary empirical support for supervisor training, and for the experiential methods on which it is based.

An interesting finding was the marked overlap in the range of methods that were used to promote both supervisor and supervisee development. For example, 11 of the 15 methods used to train supervisors in our sample of studies were also used in supervision, in order to develop supervisees (therapists). This finding is perhaps not that surprising when one views the development of supervisors and supervisees as fundamentally similar, both entailing experiential learning that is based on such variables as an assessment of development needs, graded experiences over time, and consistent feedback (Bernard & Goodyear, 2004). These are also established elements within a general, evidence-based “science of human learning” (Mayer, 2008). This also recognizes the need to prime the learner to cognitively process the instructional material (in order to organize and integrate it with prior knowledge), and to manage that process to avoid difficulties such as overloading. An expert consensus has also underscored the appropriateness of this fundamental overlap (Milne et al., 2009).

Our fourth and final objective was to note some of the implied research directions. Although these 11 controlled studies represent a promising empirical basis for supervisor training, much remains to be done. Even our brief summary indicates that there is considerable variety in the content of training programs (as also reported in a survey by Fleming, 2004), which implies equally diverse learning objectives, both representing sources for confounding the impact of training with the training methods that are used. Latterly, work has begun to build expert consensus on the objectives of such training (Roth & Pilling, 2008), which, together with survey data (Lyon, Heppler, Leavitt, & Fisher, 2008), may in turn contribute to some clarity about the content of training. Linked to this concern, more research is needed to clarify the conceptual framework underpinning training, which is usually implicitly some form of experiential learning (Kolb, 1984). But accounts such as Kolb’s have been criticized precisely because of a lack of precision (see Milne, 2009, for a summary). This makes the clarity of the conceptualization that guides one’s training efforts another research task (e.g., the role of multimedia: Mayer, 2008). In turn, we need to develop training technologies (e.g., manuals and guidelines), and then determine afresh the clinical impact of supervisor training (similar conclusions were reached in the reviews by Kilminster and Jolly, 2000, and by Spence and colleagues, 2001). In conducting this work, it is important to realize that traditional thinking about effectiveness evaluation may be fundamentally flawed. In particular, the

much-cited Kirkpatrick (1967) evaluation framework does not withstand empirical scrutiny (Colquitt, LePine, & Noe, 2000), indicating the need to concentrate more on variables such as knowledge gain and skill acquisition (as opposed to reactions, such as learner satisfaction with training). Carefully designed evaluations of effectiveness have much to contribute to the training of supervisors (Belfield, Thomas, Bullock, Eynon, & Wall, 2001). There is also a case for making more use of qualitative analyses of the process of supervisor development (Steinert et al., 2006), as this can illuminate complex interactions of the kind we can expect within supervision. For example, it may help us to recognize the process by which the different instructional methods best operate, potentially highlighting the salience of critical events such as “responsive” training (Stiles & Shapiro, 1994), and the “co-construction” of training (Milne, 2009), as opposed to a simplistic listing of the methods used.

### STUDY LIMITATIONS

It is recognized that the BES strategy necessarily limits our findings to only one form of evidence, as it does not, for example, include qualitative data. Our rationale is that BES affords an appropriate quantitative answer to a pragmatic question, one that has to date been addressed primarily in a qualitative way (i.e., through expert reviews, case studies, and consensus statements: see, for example, Watkins, 1995; Butterworth et al., 1998; Falender et al., 2004). We therefore believe that the form of evidence that BES provides has a valid, valuable, and complementary place within research and practice. In this context, it is also recognized that the term *best* implies the superiority of controlled, empirical data over all other forms of research, but this is certainly not assumed in the present case. Rather, within this positivistic perspective, the best evidence of this quantitative kind is assembled (i.e., relative to other quantitative research: controlled as opposed to uncontrolled, etc.), in order to answer a specific pragmatic question, as viewed in the context of a particular organizational culture that prizes a quantitative form of evidence-based practice (e.g., is compatible with some commissioner’s or employer’s demands for a quantitative brand of evidence-based practice). It follows that other forms of evidence will be the ‘best’ in other contexts, particularly for addressing other questions (e.g., qualitative, phenomenological case studies for clarifying participants’ perceptions about the acceptability of training). Of course, in practice most researchers adopt a pluralistic perspective, drawing on and combining the different approaches to collate complementary forms of evidence, within a multi-strand research program (Glaziou, Vandenbroucke, & Chalmers, 2004).

Within the BES approach, one problem with the present review is that our inclusion criteria were so strict that we only included those studies that had research designs that allowed valid inferences to be drawn. This meant

that only 11 adequately controlled evaluations of supervisor training were included in the review. However, a great many more evaluations of supervisor training have been conducted, using less rigorous designs. A popular example is the pretest/posttest design, which, unless it is supplemented by a network of additional evidence or rich contextual information, does not allow one to be confident about causal relationships (Elliott, 2002). It is possible, however, that these less rigorous studies provide a corroborating or complementary account of effective supervisor training, and this may merit a separate review.

A major disadvantage of this BES systematic review methodology is that a broad overview of the extant literature is sacrificed for a highly detailed micro-analysis of a handful of carefully selected studies. For instance, our sample all utilized individual (1:1) supervision, so it may be that supervisors who work within a group approach would be best trained by other methods. Similarly, our focus has been on workshops as the method of supervisor training, whereas these are sometimes complemented by formal course work, supervision-of-supervision, and other methods. While we believe firmly, along with others (Petticrew & Roberts, 2006), that this intensive approach has value (primarily in terms of ensuring adequate internal validity), we also recognize the need to integrate this kind of approach with wider-ranging review methodologies, both narrative and systematic (e.g., Spence et al., 2001; Ellis and Ladany, 1997, respectively). For instance, the BACP commissioned a systematic review on the impact of clinical supervision on counselors, therapists, their practice, and clients (Wheeler & Richards, 2007). Although the purpose of this BACP review was not to evaluate supervision training, there may be some utility in the methodology based on the Evidence for Policy Practice Information (EPPI; 2009) review that was used to evaluate and synthesize a large number of qualitative studies. Ultimately, we believe, all such attempts at conceptualizing how supervisors might best be trained needs to be put to the empirical test, and it is for this reason that we have also set out how we believe this research agenda might best progress.

## CONCLUSIONS

We began by noting the inconsistency between policy and practice in relation to supervisor training, a situation that has been likened to a “dirty little secret” within the professions (Hoffman, 1994, p. 40). By contrast, based on an impressive track record (Bambling et al., 2006; Gilbody, Bower, Fletcher, Richards, & Sutton, 2006; Heaven, Clegg, & Maguire, 2006), clinical supervision is now regarded as an “essential” ingredient in modern mental health care (Department of Health, 2008, p. 29) and a “key activity” in relation to ensuring that therapies are implemented with fidelity (Turpin,

2008, p. 102). Yet paradoxically, supervisors do not tend to receive suitably systematic training: “something does not compute” (Watkins, 1997, p. 604). Furthermore, it has been said that “research on the effectiveness of (supervision) training remains virtually non-existent” (Bernard & Goodyear, 2004, p. 291), making a systematic approach problematic.

In this damning context, we believe that our integration of the results of a small but rigorous sample of empirical supervisor training evaluations with survey data, expert consensus, and narrative reviews provides a much-needed boost, one that could improve the consistency between policy and practice. It also adds an evidence-based way to make supervisor training “compute,” preparing mental health professionals for the recognized core competence of supervision after some 25 years of theorizing about supervisor development. But we also recognize the complementary benefits that can arise from alternative forms of evidence, such as “practice-based evidence” and “empirically based” findings, as these can help to bridge the gulf that is created by the relative absence of research on which to base training recommendations (Calhoun, Moras, Pilkonis, & Rehm, 1998). In short, we hope that, through these varied scientific approaches, supervisor training can now enter the era of evidence-based practice (Milne, 2009).

## REFERENCES

- American Psychological Association (APA). (2007). *Assessment of competency benchmarks work group: A developmental model for defining and measuring competence in professional psychology*. Washington, DC: Author.
- Baltimore, M. L., & Crutchfield, L. B. (2003). *Clinical supervisor training: An interactive CD-ROM training programme for the helping professions*. London: Allyn & Bacon.
- 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.
- Barrow, M., & Domingo, R. A. (1997). The effectiveness of training clinical supervisors in conducting the supervisory conference. *The Clinical Supervisor*, *16*, 55–78.
- Belfield, C., Thomas, H., Bullock, A., Eynon, R., & Wall, D. (2001). Measuring effectiveness for best evidence medical education: A discussion. *Medical Teacher*, *23*, 164–170.
- Bernard, J. M., & Goodyear, R. K. (2004). *Fundamentals of clinical supervision* (3rd ed.). London: Pearson.
- Binder, J. L., (1993). Is it time to improve psychotherapy training? *Clinical Psychology Review*, *13*, 301–318.
- Borders, L. D., Bernard, J. M., Dye, H. A., Fong, M. L., Henderson, P., & Nance, D. W. (1991). Curriculum guide for training counseling supervisors: Rationale,



- development and implementation. *Counselor Education and Supervision*, 31, 58–80.
- British Association for Counselling and Psychotherapy (BACP). (2007). *Ethical framework for good practice in counselling and psychotherapy*. Lutterworth: Author.
- British Association for Counselling and Psychotherapy (BACP). (2008). *Skills for health/HPC update (core curriculum)*. Retrieved from <http://ccyp.co.uk/regulation/index/php?newsId=957&count=6&start=3&filter=&cat=9&year=2008>
- British Psychological Society. (2007). *Criteria for the accreditation of postgraduate training in clinical psychology*. Leicester: Author.
- Busari, J. O., Scherpbier, A. J. J. A., Van der Vleuten, C. P. M., & Essed, G. G. M. (2006). A two-day teacher-training programme for medical residents: Investigating the impact on teaching ability. *Advances in Health Sciences Education*, 11, 133–144.
- Butterworth, T., Faugier, J., & Burnard, P. (1998). *Clinical supervision and mentorship in nursing*. Cheltenham: Chapman & Hall.
- Calhoun, K. S., Moras, K., Pilkonis, P. A., & Rehm, L. P. (1998). Empirically-supported treatments: Implications for training. *Journal of Counselling & Clinical Psychology*, 66, 151–162.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of twenty years of research. *Journal of Applied Psychology*, 85, 678–707.
- Department of Health. (2008). *Improving access to psychological therapies (IAPT): Commissioning toolkit*. London: Author.
- Ducharme, J. M., Williams, L., Cummings, A., Murray, P., & Spencer, T. (2001). General case quasi-pyramidal staff training to promote generalisation of teaching skills in supervisory and direct-care staff. *Behavior Modification*, 25, 233–254.
- Elliott, R. (2002). Hermeneutic single-case efficacy designs. *Psychotherapy Research*, 12, 1–21.
- 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: Wiley.
- Evidence for Policy and Practice Information (EPPI). (2009). *Social science and public policy, systematic reviews*. Retrieved from <http://epi.ioe.ac.uk/cms>
- Falender, C., Cornish, J. A. E., Goodyear, R., Hatcher, R., Kaslow, N. J., Leventhal, G., et al. (2004). Defining competencies in psychology supervision: A consensus statement. *Journal of Clinical Psychology*, 60, 771–785.
- Falender, C. A., & Shafranske, E. (2004). *Clinical supervision: A competency based approach*. Washington, DC: APA.
- Fall, M., & Sutton, J. M. (2004). *Clinical supervision: A handbook for practitioners*. London: Pearson Education.
- Fleming, I. (2004). Training clinical psychologists as supervisors. In I. Fleming, & L. Steen (Eds.), *Supervision and clinical psychology: Theory, practice, and perspectives* (pp. 72–92). New York: Brunner-Routledge.
- Fleming, I., Gone, R., Diver, A., & Fowler, B. (2007). Risk supervision in Rochdale. *Clinical Psychology Forum*, 176, 22–25.
- Fleming, R. K., Oliver, J. R., & Bolton, D. M. (1996). Training supervisors to train staff: A case study in a human service organization. *Journal of Organizational Behavior Management*, 16, 3–25.

- Gilbody, S., Bower, P., Fletcher, J., Richards, D., & Sutton, A. J. (2006). Collaborative care for depression: A meta analysis and review of longer-term outcomes. *Archives of Internal Medicine*, *166*, 2314–2320.
- Glaziou, P., Vandenbroucke, J., & Chalmers, I. (2004). Assessing the quality of research. *British Medical Journal*, *328*, 39–41.
- Hawkins, P., & Shohet, R. (2000). *Supervision in the helping professions: An individual, group and organizational approach*. Milton Keynes, Buckinghamshire, UK: Open University Press.
- Heaven, C., Clegg, J., & Maguire, P. (2006). Transfer of communication skills training from workshop to workplace: The impact of clinical supervision. *Patient Education and Counseling*, *60*, 313–325.
- Hoffman, L. W. (1994). The training of psychotherapy supervisors: A barren scape. *Psychotherapy in Private Practice*, *13*, 23–42.
- Jensen, J. M., Parsons, M. B., & Reid, D. H. (1998). Supervisory training for teachers: Multiple, long-term effects in an education program for adults with severe disabilities. *Research in Developmental Disabilities*, *19*, 449–463.
- Kadushin, A. (1985). *Supervision in social work*. New York: Columbia University Press.
- Kaslow, N. J., Borden, K. A., Collins, F. L., Forrest, L., Ilfelder-Kaye, J., Nelson, P. D., et al. (2004). Competencies conference: Future directions in education and credentialing in professional psychology. *Journal of Clinical Psychology*, *60*, 699–712.
- Kavanagh, D. J., Spence, S. H., Wilson, J., & Crow, N. (2002). Achieving effective supervision. *Drug and Alcohol Review*, *21*, 247–252.
- Kilminster, S. M., & Jolly, B. C. (2000). Effective supervision in clinical practice settings: A literature review. *Medical Education*, *34*, 827–840.
- Kirkpatrick, D. L. (1967). Evaluation of training. In R. L. Craig, & L. R. Bittel (Eds.), *Training and development handbook* (pp. 87–112). New York: McGraw-Hill.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86.
- Kolb, D. A. (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice Hall.
- Lambert, M. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction on the brief treatment of major depression. *Psychotherapy Research*, *16*, 317–331.
- Loganbill, C., & Hardy, E. (1983). Developing training programs for clinical supervisors. *The Clinical Supervisor*, *1*, 15–21.
- Lyon, R. C., Heppler, A., Leavitt, L., & Fisher, L. (2008). Supervisory training experiences and overall supervisory development in predoctoral interns. *The Clinical Supervisor*, *27*, 268–284.
- Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*, 760–769.
- McMahon, N., & Simons, R. (2004). Supervision training for professional counselors: An exploratory study. *Counselor Education and Supervision*, *43*, 301–309.
- Milne, D. L. (1998). Clinical supervision: Time to reconstruct or to retrench? *Clinical Psychology and Psychotherapy*, *5*, 199–203.

- Milne, D. L. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology, 46*, 437–447.
- Milne, D. L. (2009). *Evidence-based clinical supervision: Principles and practice*. Chichester: BPS Blackwell.
- Milne, D. L., Aylott, H., Fitzpatrick, H., & Ellis, M. V. (2008). How does clinical supervision work? Using a Best Evidence Synthesis approach to construct a basic model of supervision. *The Clinical Supervisor, 27*, 170–190.
- Milne, D., & James, I. (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology, 39*, 111–127.
- Milne, D. L., & James, I. A. (2002). The observed impact of training on competence in clinical supervision. *British Journal of Clinical Psychology, 41*, 55–72.
- Milne, D. L., Scaife, J., & Cliffe, T. (2009). A British consensus on facilitating clinical supervisors' experiential learning. *Clinical Psychology Forum, 203*, 7–12.
- National Health Service, Centre for Reviews and Dissemination. (2001). *Undertaking systematic reviews of research on effectiveness*. Retrieved from <http://www.york.ac.uk/inst/crd/report4.htm>
- Neufeldt, S. A. (1994). Use of a manual to train supervisors. *Counselor Education and Supervision, 33*, 327–336.
- Parsons, M. B., & Reid, D. H. (1995). Training residential supervisors to provide feedback for maintaining staff teaching skills with people who have severe disabilities. *Journal of Applied Behavior Analysis, 28*, 317–322.
- Parsons, M. B., Reid, D. H., & Green, C. W. (1993). Preparing direct service staff to teach people with severe disabilities: A comprehensive evaluation of an effective and acceptable training program. *Behavioral Residential Treatment, 8*, 163–185.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell.
- Reid, D. H., Rotholz, D. A., Parsons, M. B., Morris, L., Braswell, B. A., Green, C. W., & Schell, R. M. (2003). Training human service supervisors in aspects of PBS: Evaluation of a statewide, performance-based program. *Journal of Positive Behavior Interventions, 5*, 35–46.
- Richman, G. S., Riordan, M. R., Reiss, M. L., Pyles, D. A. M., & Bailey, J. S. (1988). The effects of self-monitoring and supervisor feedback on staff performance in a residential setting. *Journal of Applied Behavior Analysis, 21*, 401–409.
- Roth, A., & Pilling, S. (2008). *A competence framework for the supervision of psychological therapies*. Retrieved from [http://www.ucl.ac.uk/clinical-psychology/CORE/supervision\\_framework.htm](http://www.ucl.ac.uk/clinical-psychology/CORE/supervision_framework.htm)
- Russell, R. K., & Petrie, T. (1994). Issues in training effective supervisors. *Applied and Preventive Psychology, 3*, 27–42
- Scott, K. J., Ingram, K. M., Vitanza, S. A., & Smith, N. G. (2000). Training in supervision: A survey of current practices. *The Counseling Psychologist, 28*, 403–422.
- Spence, S. H., Wilson, J., Kavanagh, D., Strong, J., & Worrall, L. (2001). Clinical supervision in four mental health professions: A review of the evidence. *Behaviour Change, 18*, 135–155.
- Stanton, A., Sanchez, V. C., & Kesges, R. C. (1981). Family therapy and systems supervision with the 'Pick-a-Dali-Circus' model. In F. W. Kaslow *Supervision*

- and training: *Models, dilemmas and challenges*. Binghamton, NY: The Haworth Press, 1989.
- Steinert, Y., Mann, K., Centeno, A., Dolmans, D., Spencer, J., Geleula, M., & Prideaux, D. (2006). A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education. *Medical Teacher*, *28*, 497–526.
- Stiles, W. B., & Shapiro, D. A. (1994). Disabuse of the drug metaphor: Psychotherapy process-outcome correlations. *Journal of Consulting and Clinical Psychology*, *62*, 942–948.
- Sundin, E. C., Ogren, M-L., & Boethius, S. B. (2008). Supervisor trainees' and their supervisors' perceptions of attainment of knowledge and skills: An empirical evaluation of a psychotherapy supervisor training programme. *British Journal of Clinical Psychology*, *47*, 381–396.
- Townend, M., Iannetta, L., & Freeston, M. H. (2002). Clinical supervision in practice: A survey of UK cognitive behavioural psychotherapists accredited by the BABCP. *Behavioural and Cognitive Psychotherapy*, *30*, 485–500.
- Turpin, G. (2008). *Supervision: The key to successful patient outcomes by delivering IAPT interventions of proven quality and fidelity. Executive summary: Clinical supervision principles and guidance*. London: NHS. Retrieved from <http://www.iapt.nhs.uk>
- United Kingdom Council for Psychotherapy (UKCP). (2008). *Standards for education and training: The minimum core criteria for psychotherapy with adults. SETS document number 1PwA*. London: Author.
- Wampold, B. E., David, B., & Good, R. H. (1990). Hypothesis validity of clinical research. *Journal of Consulting and Clinical Psychology*, *58*, 360–367.
- 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. 11–30). New York: Wiley.
- Watkins, C. E. (1995). Psychotherapy supervisor development: On musings, models, and metaphor. *Journal of Psychotherapy Practice and Research*, *4*, 150–158.
- Wheeler, S., & Richards, K. (2007). *The impact of clinical supervision on counsellors and therapists, their practice and clients: A systematic review of the literature*. Lutterworth: BACP.
- Whitman, S. M., Ryan, B., & Rubenstein, D. F. (2001). Psychotherapy supervisor training: Differences between psychiatry and other mental health disciplines. *Academic Psychiatry*, *25*, 156–161.