Evaluating the training of clinical supervisors: a pilot study using the fidelity framework

Tonia Culloty[1], Derek L. Milne[2]* and Alia I. Sheikh[2]

[1]*Northumberland, Tyne & Wear Mental Health NHS Trust, St George's Park Hospital, Morpeth, Northumberland, UK*
[2]*School of Psychology, Newcastle University, Newcastle upon Tyne, UK*

Abstract. Theevaluation of supervisor training has featured weak measurement and lacked a coherent framework, limiting effectiveness. A literature review was first conducted to clarify the current status of supervisor workshop evaluations, related to the promising fidelity framework. This consists of five criteria: the workshop's *design*, the *training* (competence of the trainer), the *delivery* of the workshop, the learning of the participants (*receipt*), and the clinical practice outcomes (*enactment*). Second, we applied this framework to the training of supervisors ($n = 17$) in a cognitive-behavioural therapy (CBT) approach, by analysing one trainer leading two successive supervisors' workshops. The review of the literature indicated that there were significant psychometric and conceptual deficiencies in the current evaluation of supervisor training. The data from the case-study analysis suggest that the manual-based workshop could be delivered with high fidelity by this trainer (e.g. the CBT approach to supervision received 89% approval). The fidelity framework provided a systematic, feasible and coherent rationale for the evaluation of supervisor training. Our preliminary findings indicated that the workshop was successful. To fulfil its promise as an improved way of evaluating supervisor training, the framework should be piloted with other trainers, instruments and workshops, using controlled designs.

Key words: Evaluation, fidelity framework, supervision, training.

Introduction

Clinical supervision has been empirically defined as: 'The formal provision, by senior/ qualified health practitioners, of an intensive, relationship-based education and training that is case-focussed and which supports, directs and guides the work of colleague/s (supervisees)' (Milne, 2007*a*; p. 440). It has grown in prominence internationally, due to the emergence of government policies regarding high-quality care (DoH, 1998), an improving literature on what constitutes effective supervision (Bambling *et al.* 2006), and the commitment of professional bodies to promote evidence-based mental health practice (American Presidential Task Force,

---

*Author for correspondence: Dr D. L. Milne, School of Psychology, Newcastle University, Newcastle upon Tyne, UK. (email: d.l.milne@ncl.ac.uk)

2006). However, supervisor training represents an obstacle to policy implementation, partly because 'few rigorously controlled evaluations of the effectiveness of supervisor training have been conducted' (Kavanagh *et al.* 2002, p. 250). Although a systematic review located 11 controlled evaluations, practical application of the findings was compromised by the lack of consensus on the workshop methods (in total, some 20 methods were reported) and by the surprising lack of emphasis on established training methods (e.g. educational needs assessment and observation; Milne *et al.* unpublished data).

Training is also impeded by the general shortage of instruments, particularly psycho-metrically sound tools that can assess competence in supervision (Ellis & Ladany, 1997). For example, the Bambling *et al.* (2006) study relied on an *ad-hoc* instrument to measure adherence in supervision (to either a cognitive-behavioural or a psychodynamic approach). Although this had good reliability, no validity data were presented. Furthermore, a review by Lomax *et al.* (2005) noted that important components like feedback and evaluation 'are all too often given only cursory attention and handled in a haphazard fashion' (p. 501). In recognition of this problem, Townend *et al.* (2002) argued that measurement should assess changes in clinical supervisors' knowledge, skills and attitudes (KSAs) with the aid of psychometrically sound tools, linked to an evaluation of their supervisees' competence, and in turn to client outcomes. Their appraisal echoes the review of CBT supervision by Milne & James (2000), whose sample of 28 empirical studies included 11 instruments that evaluated learning (the KSAs), although there were many more instruments used to assess the generalization of this learning to the workplace. However, these 28 studies mostly utilized *ad-hoc* quizzes and simple checklists with no psychometric validation, and they were heavily reliant on direct observation. Finally, treating such outcome data as the sole basis for evaluation omits other vital criteria, such as training process and workshop content, as these are logically necessary to forming a sound judgement about outcomes (Rossi *et al.* 2003).

In summary, our focus is on the important but incomplete task of evaluating supervisor training. We next consider a promising way forward, the fidelity framework, a measurement strategy that is particularly compatible with CBT because of its empirical emphasis (including stepwise mediators of behaviour change and observable causal mechanisms). We then review the relevant literature, to assess how extensively it has been applied.

The fidelity framework

Treatment fidelity can be defined as the different methodological strategies that researchers use to monitor and improve the reliability and validity of their interventions (Borrelli *et al.* 2005). Fidelity includes familiar concepts such as competence, adherence and generalization, but integrates these, encouraging researchers (or clinicians) to think systematically about their interventions. It is also a necessary condition for inferring that an intervention is indeed responsible for an obtained effect. For example, non-significant findings may be due to the faulty or weak administration of CBT, rather than being attributable to a specific cognitive intervention that is inherently ineffective (Waller, 2009). Thus, ensuring that an intervention has high fidelity can prevent the illogical abandonment of an effective intervention. It can also increase the plausibility that a high-fidelity treatment is indeed the reason for significant findings. Additionally, fidelity has merit in terms of increasing statistical power, theory

Table 1. *A summary of the fidelity framework*

Fidelity framework	Questions addressed
Design of supervision	What is the right thing to do? Which supervision model was used?
Training of supervisors	Has the right thing been done? Was there adherence to the model?
Delivery of supervisor training	Has it been done right? Was the intervention manipulated correctly?
Receipt of training	Did it result in the right outcome? Is it having the intended initial effect on the supervisor?
Enactment of training	Did it result in the right impact? Was there the expected transfer to supervisees?

refinement, and promoting dissemination (e.g. enabling the writing of more precise and detailed treatment manuals; Bellg *et al.* 2004).

The fidelity framework consists of five intervention criteria, arranged in a stepwise manner. Specifically, high fidelity first requires information on the way that an intervention is *designed*, followed by data on the *training* of clinicians to be competent in applying the intervention. If fidelity has been demonstrated at these initial two levels then the *delivery* of the intervention becomes the next step in the evaluation, followed by an assessment of the *receipt* or mini-impact of the intervention on the client. If the intervention has proved to have fidelity thus far, the final analysis concerns the client's *enactment* (generalization) of any such receipt. Individually, these are not novel criteria, and CBT has traditionally given significant emphasis to them over the years, albeit in a partial fashion (see review below), especially generalization. Therefore, we believe that their collective formulation as a taxonomy for evaluating interventions such as CBT is a promising conceptualization. This is partly because the fidelity framework is content-free, and can therefore be applied legitimately to CBT practice in all its diversity, including clinical supervision and supervisor training – our present focus. However, we are aware that there is as yet no consensus regarding this or any related approach, so our stance is to treat it as a promising formulation, to be evaluated pragmatically (can it be applied?) and empirically (does it work?). Table 1 outlines the framework, defining the five criteria in terms of the basic questions that are posed. Next, we consider the extent to which prior research on supervisor training has addressed the five criteria within this framework, before illustrating its use in a pilot study.

Literature review: to what extent has the fidelity framework been applied?

The relevant literature was searched using Medline, PsycINFO, and ASSIA databases from 1988 to 2008, with the key words 'clinical supervision' or 'supervision training'. Ninety-four articles were located and then examined to evaluate their relevance. Of these, 13 papers were judged to be relevant, and are summarized in Table 2. The data indicate that only one of the 13 studies addressed all five levels of fidelity (i.e. Henggeler *et al.* 2008). By contrast, the majority of studies in this review addressed one or two fidelity criteria, most frequently

Table 2. *Use of the fidelity framework within a sample of the supervisor training literature*

	Design	Training	Delivery	Receipt	Enactment	No. of fidelity dimensions addressed (out of 5)
Frisch (1989)	Yes	No	No	Yes	No	2
Smith *et al.* (1992)	Yes	No	No	Yes	No	2
Diwan *et al.* (1996)	Yes	No	No	No	No	1
Borders *et al.* (1996)	Yes	No	No	Yes	No	2
Fleming *et al.* (1996)	No	No	No	Yes	No	1
Barrow & Domingo (1997)	No	No	No	Yes	No	1
Ducharme *et al.* (2001)	Yes	No	No	Yes	No	2
Spence *et al.* (2002)	Yes	No	No	No	No	1
O'Brien *et al.* (2003)	No	No	No	No	No	0
Bedward & Daniels (2005)	Yes	No	No	Yes	No	2
Busari *et al.* (2006)	Yes	No	No	Yes	No	2
Henggeler *et al.* (2008)	Yes	Yes	Yes	Yes	Yes	5
McDonnell *et al.* (2008)	Yes	No	Yes	Yes	No	3
Total: 13 studies	**10 (77%)**	**1 (8%)**	**3 (21%)**	**10 (77%)**	**1 (8%)**	

design and/or *receipt*. As can be seen from Table 2, these two levels were addressed by 77% of the studies. With a frequency of 8% each, *training* (adherence/competence) and *enactment* (generalization) were the two aspects of fidelity that were least frequently addressed in this sample of studies.

Therefore, this literature review, which includes CBT training, suggests that fidelity is rarely treated comprehensively, with only one of 13 studies explicitly reporting data that were relevant to all five levels of fidelity. This corroborates the view that few coherent evaluation frameworks have been applied to supervisor training. On the other hand, these studies do indicate that the fidelity framework can be applied and enjoy acceptance. We therefore next outline a case study in which we attempt to apply the fidelity framework to an example of supervisor training.

Pilot study of the fidelity framework

Method

Participants

Completion of the sampled part-time M.Sc. in CBT for Psychosis and Recovery in Complex Mental Health entailed an annual workshop on clinical supervision, which was delivered to a total of 17 mental health professionals in two consecutive cohorts, one year apart. These participants included four males and 13 females. One participant was a social worker, one was a psychiatrist, and the remaining participants ($n = 15$) had a professional background in mental health nursing. The age range was 28–55 years (mean 40.5 years) and the number of years of post-qualification experience ranged from 3 years to 27 years (mean 13.6 years). The workshop leader was a part-time clinical lecturer (the first author) who had 5 years of clinical

experience as an occupational therapist, followed by 4 years as a psychological therapist. During this 9-year clinical period, she had 8 years of experience as a clinical supervisor.

Design

A quasi-experimental, longitudinal design was utilized. Self-report data were collected at the close of the two consecutive workshops, each lasting for 3 days, and then again after a follow-up period (between 6 and 12 weeks). In addition, direct observations were made of a sample of two sessions from day 2 of the second workshop. In total, four measures (quantitative and qualitative) of the training were applied in order to operationalize the full fidelity framework. A further major influence on the study design was the 'indirect evidence' approach (Elliott, 2002). In the absence of controls, this draws on a combination of evaluation methods to create a network of evidence, in order to clarify the extent to which there is a plausible causal link between the intervention and the anticipated outcomes. In particular, the network includes a searching, sceptical style of enquiry into whether or not any apparent benefits of training are actually due to the training. For example, by using an interview approach, it encourages the evaluator to assess directly whether the participants' have a tendency to respond to assessments in a socially desirable way, and to check whether alternative explanations for obtained changes are equally plausible (e.g. life events, other potential causes of improved scores). To strengthen this feature, we invited a colleague who was not directly involved in the training (the third author) to conduct this interview.

Instruments

We now describe the four instruments used to operationalize the fidelity framework. Note that some instruments were used to assess more than one fidelity level (by matching the most relevant items in instruments with the given fidelity level, as described below).

The first of these instruments was the Training Acceptability Rating Scale (TARS; Davis *et al.* 1989), which was one of two instruments for evaluating whether the *design* of the supervision workshop was appropriate. There are six items within this self-report 'reaction' questionnaire, each assessing the workshop's 'acceptability' (including items on appropriateness and social validity). An example item is: 'This approach to supervision would be appropriate for a variety of staff' (item 1: General acceptability). TARS items 1 and 4–6 were treated as assessments of *design*, items 12 and 13 as assessments of the training, items 2, 7 and 9 as assessments of receipt of training, and item 3 contributed to the assessment of *enactment of training.* The TARS items are rated on a six-point, bipolar Likert scale, ranging from 'strongly agree' to 'strongly disagree'. TARS has good test–retest reliability ($r = 0.83$) and internal consistency (0.99), as well as sound construct and concurrent validity.

The second instrument was an *ad-hoc*, semi-structured interview. This consisted of seven main items, each with a number of sub-items or prompts to facilitate the interviewees' responding (all participants were interviewed together, as a group of trainee supervisors). It was designed to reflect the 'indirect evidence' method (Elliott, 2002). Items covered topics like the workshop's *design* by asking open-ended questions and prompting for alternative explanations, etc. (e.g. 'what happened during the training – what kind of workshop was it? Prompts: did the workshop rely on lectures? Was the guideline distributed? Were you asked to read or discuss this material?). More specifically, interview item 2 assessed the training itself, items 2, 4 and 7 assessed the delivery of the training, item 3 assessed the receipt,

and item 5 assessed enactment of training. Answers were recorded live by the interviewer, in terms of the essence of the interviewees' replies. The only ratings included were for the first and last questions, which entailed the interviewer making a rating of social desirability at the outset (following a suitable ice-breaker question, designed to clarify the interviewees' socially desirable responding), and then again at the close, when the interviewees are asked to rate how 'free and frank' their reflections were. Both of these social desirability ratings were on a ten-point scale, ranging from 'completely absent' to 'strongly present'.

Third, we measured adherence to the training model by direct observation, with Teachers' Process Evaluation of Training and Supervision (PETS; Milne *et al.* 2002). This is a 25-item rating scale, designed to measure the training approach that was used (*training adherence*) and the associated mini-impacts on the participants' experiential learning (the latter were treated as part of our assessment of *delivery*). The training approach items (1–19) cover the main ways that trainers facilitate experiential learning, with items such as 'supporting', 'questioning' and 'providing feedback'. The remaining six items are treated as a measure of the effect of the training on the participants' learning. Example items are '*Reflecting*' and '*Conceptualizing*' (see Table 3). Each of the 25 items was rated on the seven-point competence scale, ranging from 'incompetent' to 'expert', based on the Supervision Adherence and Guidance Evaluation rating scale (SAGE; unpublished instrument by Milne & Reiser, available from the corresponding author). This departed from the momentary time-sampling procedure that is normally used with PETS so that a competence assessment could be made. PETS has demonstrated good inter-rater reliability ($\kappa = 0.84$) and promising empirical and concurrent validity (Milne *et al.* 2002).

The fourth measure was the Marlowe–Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960). This is a widely used instrument to test for a social desirability bias. It is a self-report questionnaire, consisting of 33 items, each one answered on a 'true-false' basis. The maximum score is 33 and scores of $\geqslant 19$ have been interpreted as indicative of socially desirable responding amongst professionals (Andrews & Meyer, 2003). An illustrative item is, '*I'm always willing to admit it when I make a mistake*'. Internal consistency and test–retest reliability have been reported as sound by a succession of researchers (i.e. values between 0.72 and 0.89; Crowne & Marlowe, 1960), and the MCSDS's criterion-related validity also appears to be acceptable (value 0.83; for details see Andrews & Meyer, 2003).

Procedure

The two workshops were held over three consecutive days, structured into six sessions, and guided by a 185-page manual for training supervisors in the evidence-based clinical supervision approach (EBCS; Milne & Westerman, 2001; Milne, 2009). This approach is thought to enhance traditional CBT supervision (Milne, 2008). The part-time nature of the Masters course meant that the three workshop days were delivered at fortnightly intervals. On each day, one session was delivered in the morning and one in the afternoon (with a coffee break within each session and a lunch break splitting the sessions). Each session tackled one section of the EBCS manual: *orientation to supervision*; *goal-setting*; *facilitating learning*; *the supervisory relationship*; *evaluation*; and *the supervision system*. In terms of the workshop methods, each session included an introductory powerpoint slideshow presentation; guided reading of four NICE-style supervision guidelines; group discussion, viewing and discussing video clip illustrations, and educational role plays with feedback. All participants

Table 3. *Competence ratings of the workshop leader, based on direct observation* (*teachers' PETS*)

		Tape 1	Tape 2	Tape 3	Mean
Workshop leader's behaviours					
1	Listening/observing	3	5	5	4.3
2	Supporting	3	5	4	4
3	Closed questioning	3	5	4	4
4	Open questioning	4	5	5	4.7
5	Needs assessing	3	3	3	3
6	Goal-setting	3	4	4	3.7
7	Restating	3	3	4	3
8	Reflecting	2	2	3	2.3
9	Interpreting	2	3	4	3
10	Formulating	3	4	4	3.7
11	Managing	3	3	2	2.7
12	Informing	3	5	4	4
13	Guided experiential learning	0	3	4	2.3
14	Self-disclosing	2	0	3	1.7
15	Challenging	3	4	3	3.3
16	Disagreeing	2	0	0	0.7
17	Evaluating	2	3	3	2.7
18	Feeding back	2	4	5	3.7
19	Other	0	0	0	0
Effects on learners					
20	Experiencing	3	4	3	3.3
21	Reflecting	4	4	5	4.3
22	Conceptualizing	3	3	4	3.3
23	Planning	3	5	5	4.3
24	Experimenting	0	4	4	2.7
25	Other	0	0	0	0

Ratings key: 0, absent; 1, 'incompetent'; 2–4, 'competent'; 5–6, 'expert'.

were provided with an information sheet and consent form prior to data collection, and all consented to participate in this study, which had received ethical approval from the local Research and Development Department. The TARS was completed at the end of the 3-day workshop; Teachers' PETS observations were based on three professional-quality video recordings that were made of the second day of the workshop (i.e. tapes 1–3 in Table 3), to provide a representative sample of two of the workshop sessions ('facilitating learning' and 'supervision alliance': a total of 2 h and 48 min were taped and rated). The tapes were coded with the PETS instrument by an experienced independent observer, blind to the study objectives. The Social Desirability Questionnaire and Interview were administered 6–10 weeks after the end of the workshop.

Results

The findings are summarized below as they correspond to each fidelity level. As the findings were similar, we have combined the data from the two workshops.

Design

The results suggest that both workshops were associated with excellent acceptability ratings (89% endorsement of the EBCS approach by the 17 participants), although in interview the participants noted a need to more clearly distinguish 'normative' supervision (which addresses management issues, such as waiting lists and service audits) from the primarily 'formative' function (i.e. the educative aim; Proctor, 1988). This information suggests that the workshop *design* was appropriate.

Training

These participants were also satisfied that the trainer had remained true to the EBCS approach, in the way that she delivered the workshop. This was indicated by both the TARS data (79%) and the group interview. For instance, the balance between experiential and didactic work was perceived to have been faithful to the model ($n = 17$). The observational data also indicated that the *training* had adhered to the model, as at least 'competent' ratings were made for almost all of the relevant PETS items (1–18). Indeed, eight ratings fell in the 'expert' range (e.g. for appropriate listening, supporting, and questioning on tape 2), but there were also some poor ratings (e.g. of item 16, 'disagreeing', on tapes 2 and 3). However, the overall profile suggests adherence to the EBCS model (mean percent rating for these items was 94%), with the ratings falling firmly within the competent range on the distinctive EBCS items (e.g. 'conducting a needs assessment', 'goal-setting', and 'guiding experiential learning'; these received mean ratings of 3, 3.7, and 2.3, respectively). This indicates good adherence, as assessed by self-report and by direct observation. The direct observation data are detailed in Table 3.

Delivery

In terms of the workshop *delivery*, the TARS and interview data indicated that it had a good structure and used a comprehensive approach ($n = 6$). For example, the mean TARS score for the relevant items was 88%. However, some improvements were proposed, such as focusing on the knowledge base in the morning, so that the afternoon could be devoted to experiential work ($n = 11$). The mean Social Desirability Scale score of 17.5 is quite high, but is consistent with the interpretation that we received acceptably candid communication from the delegates, rather than a 'grateful testimonial'. The observational data supports the view that the workshop was delivered in a competent way, as the mean rating for PETS items 1–18 was at the centre of the 7-point competence scale (3.25).

Receipt

Following on from the above data, one might reasonably expect that the workshop also facilitated the supervisors' learning. Consistent with this assumption, the TARS findings were favourable (i.e. 78%), and the observational data indicated that the workshop leader had been effective in facilitating these learning modes, with a mean PETS score for her competence in fostering the relevant experiential learning items of 3.6.

Enactment

Finally, at the close of the workshop, the group believed strongly that the training was transferable, and by the time of the follow-up interview (6–12 weeks after the workshop), at least six of them reported having transferred some of the workshop material to their routine NHS work with their supervisees. Examples included introducing a clearer structure to their supervision (e.g. collaborative agenda setting), being more goal-oriented (e.g. ensuring that the full experiential learning cycle occurs within each session), and relating supervision to the supervisees' learning styles (e.g. increasing the emphasis on reflection). The remainder felt that they were not yet ready to start transferring material, but were optimistic of so doing.

Discussion

In order to pilot the systematic evaluation of supervisor training, this study operationalized the fidelity framework, applying four instruments to assess the five evaluation criteria within the fidelity framework: the approach taken to supervision (the *design*; EBCS); the consistency between this model and its competent presentation within the workshop (*training*); whether the workshop process was appropriate (*delivery*); the consequent learning impacts on the participants (*receipt*); and last, whether this was transferred to their supervisees post-workshop (*enactment*). At the procedural level, this evaluation was achieved without exceptional demands on either evaluators or participants, despite entailing a variety of instruments (self-report and direct observation). Therefore, consistent with the one study in our review that had attempted it (i.e. Henggeler *et al.* 2008), we found that the framework could be applied readily to supervisor training. Because it was fairly successful, the present workshop is perhaps not the best illustration of the fidelity framework. One of the advantages of this stepwise approach to evaluation is that it has the potential to pinpoint where a workshop is failing (e.g. a sound *design* might be spoiled by poor *delivery*). In the present example the identified problem arose at the final *enactment* stage, a common problem that would probably require a significant complementary effort at the organizational level (Beidas & Kendall, 2010).

In terms of evaluating the present workshop, our preliminary findings indicated that the training had been fairly effective at all five fidelity levels. However, this aspect of the study should be treated as purely illustrative. Methodological limitations include the small sample size (one trainer and 17 supervisors), the onus we placed on self-reported 'reaction' data, and the absence of any controls. Future research might therefore include data concerning the participants' learning, such as a quiz, and more objective data on their transfer of any learning, such as an audit of supervision (Kirkpatrick, 1967; Milne, 2007*b*). These criticisms clearly limit our capacity to infer that the workshop was the reason for the 'high fidelity' reported by the supervisors.

In addition to these methodological caveats, we should also recognize that the concept of the fidelity framework has practical and theoretical limitations. Practically, the operationalization of this multi-faceted framework presents methodological challenges. There are not only five linked steps, but also a related emphasis on applying multiple measures (both in order to strengthen the validity of the measurements). Furthermore, the framework is ideally conceptualized as the basis of a feedback system, in which clinical outcomes inform the earlier steps (e.g. the future design of supervisor training). Therefore, the framework can seem impractical, with multiple feasibility and resource challenges.

Theoretically, although the five levels of analysis constitute a coherent evaluation taxonomy, this may not ultimately prove to be the best way to construe the measurement task, at least in relation to workshop evaluation. This is because it is possible that some critical changes are missed between or within the five levels (e.g. attention to the 'receipt' of declarative knowledge could obscure important changes in the participants' procedural knowledge). Another example is that the time-frame for the overall fidelity framework is too brief (e.g. negative 'side-effects' of training may only emerge years later, due to important contextual developments; Beidas & Kendall, 2010). For these kinds of reasons, we recognize the need for caution in the application and interpretation of the framework.

However, as a case study in applying the fidelity framework, this pilot workshop evaluation indicates how a battery of instruments can be fitted to the framework, and can then yield potentially useful information. For instance, the *receipt* data pinpointed the unusual degree of challenge that the material (and the workshop leader) created, but suggested that it was probably optimal, as it was ultimately associated with the desired learning episodes (see Zorga, 2002, for the reasoning behind this interpretation).

In conclusion, this case study represents a preliminary example of the application of the fidelity framework, which offers a coherent, stepwise approach to the evaluation of supervisor training (it may also have merit in relation to evaluating other training, and to CBT interventions generally). Our case study evaluation should be viewed purely as illustrative and be improved upon, for example by adding more objective data on learning and its transfer, and by introducing larger samples, more trainers, and controlled designs. An example of possible improvements is a controlled study of the transfer of communication skills reported by Heaven *et al.* (2006), who provided a 3-day workshop to 61 clinical nurse specialists. They evaluated the effectiveness of the training (in terms of both learning and transfer) using real and simulated patient interviews, as conducted by these nurses. A rating scale, tapping 10 interviewing skills (e.g. exploration and responding to disclosure), was applied to tape recordings of these interviews and inter-rater reliability with the scale was demonstrated. The findings indicated that all the nurses had improved their communication skills significantly. However, this improvement was only maintained at the 2-month follow-up assessment by the group that received supervision.

Summary

- The literature shows that evaluations of supervisor training are at present incomplete in that they omit important information about the training and its stepwise impacts, thus limiting our understanding of the effectiveness of such training.
- The fidelity framework is a coherent conceptualization of training, one that incorporates information about competence, adherence, and generalization; this can enhance training evaluations (e.g. by pinpointing the step at which training is failing).
- One preliminary step is to ensure that any training seems appropriate to the participants (i.e. has social validity). In our illustrative evaluation, we found that our approach to supervision, which has been likened to 'super-CBT supervision' (Milne, 2008), was highly acceptable.
- In terms of the workshop also developing supervision skills, we believe that there was some plausible but decidedly preliminary evidence suggesting that the workshop resulted in

relevant learning and transfer given the replication of the findings in the second workshop, the use of multiple measures, and the plausibility interview data).

This promising work provides a basis for more rigorous evaluations of the vital business of supervisor training (e.g. controlled designs).

Acknowledgements

Declaration of Interest

None.

Recommended follow-up reading

On how to train supervisors
Kaslow NJ, Borden KA, Collins FL, Forrest L, Illfelder-Kaye J, Nelson PD, Rallo JS, Vasquez MJJ, Willmuth ME (2004). Competencies conference: future directions in education and credentialing in professional psychology. *Journal of Clinical Psychology* **60**, 699–712.

On the competences that training might target
Falender C, Cornish JAE, Goodyear R, Hatcher R, Kaslow NJ, Leventhal G, Shafranske E, Sigmon ST, Stoltenberg C, Grus C (2004). Defining competencies in psychology supervision: a consensus statement. *Journal of Clinical Psychology* **60**, 771–785.

A controlled example of supervisor training
Barrow M, Domingo RA (1997). The effectiveness of training clinical supervisors in conducting the supervisory conference. *The Clinical Supervisor* **16**, 55–78.

On the effectiveness of supervisor training
Kilminster SM, Jolly BC (2000). Effective supervision in clinical practice settings: a literature review. *Medical Education* **34**, 827–840.

References

American Presidential Task Force (2006). Evidence-based practice in clinical psychology. *American Psychologist* **61**, 271–285.

Andrews P, Meyer RG (2003). Marlowe–Crowne Social Desirability Scale and short form C: forensic norms. *Journal of Clinical Psychology* **59**, 483–492.

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 RA (1997). The effectiveness of training clinical supervisors in conducting the supervisory conference. *The Clinical Supervisor* **16**, 55–77.

Bedward J, Daniels HRJ (2005). Collaborative solutions – clinical supervision and teacher support teams: reducing professional isolation through effective peer support. *Learning in Health and Social Care* **4**, 53–66.

Beidas RS, Kendall PC (2010). Training therapists in evidence-based practice: a critical review from a systems-contextual perspective. *Clinical Psychology: Science and Practice* **17**, 1–30.

Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Ogedegbe G, Orwig D, Ernst D, Czajkowski S (2004). Enhancing treatment fidelity in health behaviour change studies: best practises and recommendations from the NIH Behaviour Change Consortium. *Health Psychology* **23**, 443–451.

Borders LD, Rainey LM, Crutchfield LB, Martin DW (1996). Impact of a counseling supervision course on doctoral students' cognitions. *Counselor Education and Supervision* **35**, 204–217.

Borrelli B, Sepinwall D, Ernst D, Bellg AJ, Czajkowski S, Greger R *et al.* (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across ten years of health behaviour research. *Journal of Consulting and Clinical Psychology* **73**, 852–860.

Busari JO, Scherpbier AJJA, Van Der Vleuten CPM, Essed GGM (2006). A two-day teacher-training programme for medical residents: investigating the impact on teaching ability. *Advances in Health Sciences Education* **11**, 133–144.

Crowne DP, Marlowe D (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* **24**, 349–354.

Davis JR, Rawana EP, Capponi DR (1989). Acceptability of behavioural staff management techniques. *Behavioural Research and Treatment* **4**, 23–44.

DoH (1998). A First Class Service: Quality in the New NHS. London: Department of Health.

Diwan S, Berger C, Ivy C (1996). Supervision and quality assurance in long-term-care case management. *Journal of Case Management* **5**, 65–71.

Ducharme JM, Williams L, Cummings A, Murray P, Spencer T (2001). General case quasi-pyramidal staff training to promote generalization of teaching skills in supervisory and direct-care staff. *Behavior Modification* **25**, 233–254.

Elliott R (2002). Hermeneutic single-case efficacy design. *Psychotherapy Research* **12**, 1–21.

Ellis MV, Ladany N (1997). Inferences concerning supervisees and clients in clinical supervision: an integrative review. In: *Handbook of Psychotherapy Supervision* (ed. C. E. Watkins), pp. 447–507. New York: Wiley.

Fleming RK, Oliver JR, Bolton DM (1996). Training supervisors to train staff: a case study in a human service organization. *Journal of Organizational Behavior Management* **16**, 3–25.

Frisch MB (1989). An integrative model of supervisory training for medical center personnel. *Psychological Reports* **64**, 1035–1042.

Heaven C, Clegg J, Maguire P (2006). Transfer of communication skills training from workshop to workplace: the impact of clinical supervision. *Patient Education and Counselling* **60**, 313–325.

Henggeler SW, Schoenwald SK, Liao JG, Letourneau EJ, Edwards DL (2008). Transporting efficacious treatments to field settings: the link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child & Adolescent Psychology* **31**, 155–167.

Kavanagh DJ, Spence SH, Wilson J, Crow N (2002). Achieving effective supervision. *Drug and Alcohol Review* **21**, 247–252.

Kirkpatrick DL (1967). Evaluation of training. In: *Training and Development Handbook* (ed. R. L. Craig and L. R. Bittel), pp. 87–112. New York: McGraw-Hill.

Lomax JW, Andrews LB, Burruss JW, Moorey S (2005). Psychotherapy supervision. In: *Oxford Textbook of Psychotherapy*, pp. 495–503. Oxford University Press.

McDonnell A, Sturmey P, Oliver C, Cunningham J, Hayes S, Galvin M, Walshe C, Cunningham C (2008). The effects of staff training on staff confidence and challenging behaviour in services for people with autism spectrum disorders. *Research in Autism Spectrum Disorders* **2**, 311–319.

Milne D, James I (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology* **39**, 111–27.

Milne DL (2007*a*). An empirical definition of clinical supervision, *British Journal of Clinical Psychology* **46**, 437–447.

Milne DL (2007*b*). Evaluation of staff development: the essential 'SCOPPE'. *Journal of Mental Health*
16, 389–400.

Milne DL (2008). CBT supervision: from reflexivity to specialisation. *Behavioural and Cognitive Psychotherapy* **36**, 779–786.

Milne DL (2009). *Evidence-Based Clinical Supervision: Principles and Practice*. Oxford: Wiley/Blackwell.

Milne DL, James IA, Keegan D, Dudley M (2002). Teachers PETS: a new observational measure of experiential training interactions. *Clinical Psychology and Psychotherapy* **9**, 187–199.

Milne DL, Westerman C (2001). Evidence-based clinical supervision: rationale and illustration. *Clinical Psychology and Psychotherapy* **8**, 444–445.

O'Brien A, Price C, Burns T, Perkins R (2003). Improving the vocational status of patients with long-term mental illness: a randomised controlled trial of staff training. *Community Mental Health Journal* **39**, 333–347.

Proctor B (1988). A cooperative exercise in accountability. In: *Enabling and Ensuring* (ed. M. Marken and M. Payne), pp. 21–34. Leicester: Leicester National Youth Bureau and Council for Education and Training in Youth and Community Work.

Rossi PH, Freeman HE, Lipsey MW (2003). *Evaluation: A Systematic Approach*, 7th edn. London: Sage Publications.

Smith T, Parker T, Taubman M, Lovaas OI (1992). Transfer of staff training from workshops to group homes: a failure to generalize across settings. *Research in Developmental Disabilities* **13**, 57–71.

Spence C, Cantrell J, Christie I, Samet W (2002). A collaborative approach to the implementation of clinical supervision. *Journal of Nursing Management* **10**, 65–74.

Townend M, Iannetta L, Freeston MH (2002). Clinical supervision in practice: a survey of UK cognitive behavioural psychotherapists accredited by the BABCP. *Behavioural and Cognitive Psychotherapy* **30**, 485–500.

Waller G (2009). Evidence-based treatment and therapist drift. *Behaviour Research and Therapy* **47**, 119–127.

Zorga S (2002). Supervision: the process of lifelong learning in social and educational professions. *Journal of Inter-Professional Care* **16**, 265–276.

Learning objectives

By the end of this paper, the reader should be able to:

(1) Understand the fidelity framework concept.
(2) Identify the five levels associated with the fidelity framework.
(3) Summarize the current status of the literature in relation to applications of this framework.
(4) Describe the present example of supervisor training, in terms of the workshop's content and methods.