

## An $N = 1$ Evaluation of Enhanced CBT Supervision

Derek L. Milne

*Newcastle University, UK*

Robert P. Reiser

*Palo Alto University, USA*

Tom Cliffe

*Newcastle University, UK*

**Background:** Clinical supervision plays an essential role in the development of mental health professionals and is increasingly viewed as a discrete professional specialization. However, research has rarely addressed core issues such as the measurement and manipulation of clinical supervision, so there are very few direct comparisons between the different supervision methods. **Aims:** To operationalize two related approaches, cognitive-behavioural (CBT) and evidence-based clinical supervision (EBCS), demonstrate their fidelity, and then evaluate their relative effectiveness in facilitating the experiential learning of one supervisee. **Method:** Within a multiple-baseline,  $N = 1$  design, we rated audiotapes of supervision on a competence rating scale. **Results:** Findings generally favoured the EBCS approach, which was associated with higher fidelity by the supervisor and increased engagement in experiential learning by the supervisee. **Conclusions:** This preliminary but novel evaluation indicated that CBT supervision could be enhanced. Implications follow for supervisor training and a more rigorous  $N = 1$  evaluation.

*Keywords:*  $N = 1$  evaluation, CBT, evidence-based clinical supervision.

### Introduction

Cognitive-behavioural (CBT) supervision was specified in seminal articles (Padesky, 1996; Liese and Beck, 1997) and updated in 2010 (Newman, 2010). However, the actual practice of CBT supervision has not demonstrated good fidelity to this model. In a repeat survey of CBT supervisors after a 5-year interval, Townend, Iannetta, Freeston and Harvey (2007) found continued evidence of poor fidelity, especially as regards the limited use of direct observation,

---

Reprint requests to Derek Milne, Doctorate in Clinical Psychology, 4<sup>th</sup> Floor, Ridley Building, Newcastle University, Newcastle NE1 7RU, UK. E-mail: d.l.milne@ncl.ac.uk

© British Association for Behavioural and Cognitive Psychotherapies 2012

few standardized competency ratings (such as the CTS), and minimal use of experiential methods (role play and behavioural rehearsal). Similarly, observational research on the CBT supervision model indicated that it had rarely been implemented with fidelity, leading to the view that manual-based,  $N = 1$  analyses are needed to evaluate its effectiveness (Milne, 2008).

Competency statements (Roth and Pilling, 2008) may improve fidelity, but the field still lacks instruments to operationalize the practice of CBT supervision (Milne and Reiser, 2011). For example, the Improving Access to Psychological Therapies (IAPT) Education and Training Group (2011) noted: "...reliable scales of supervision competence have yet to be developed..." (p.3) Fidelity may also be improved by having a supervision manual. Watkins (1998) has made the case for a manualized approach, together with other efforts to address the supervision paradox (i.e. a weak scientific knowledge-base yet huge importance within professional practice). He concluded that "something does not compute" (Watkins, 1997, p. 604). As a result of this paradox, it is thought that, in place of training, most supervisors rely on imitating how they themselves were supervised, or on transferring what they do in therapy to their supervision (Townend, Iannetta and Freeston, 2002). In their consensus statement, Falender et al. (2004) urged that, in order to support the emergence of supervision as a distinct competency, "A range of research procedures should be employed, including, for example, self-report, experimental, single-subject repeated measures..." (p. 775).

To contribute to improved fidelity, the present study adopts an  $N = 1$ , multiple baseline design to assess CBT supervision in comparison to Evidence-Based Clinical Supervision (EBCS), a science-informed approach to conducting and evaluating supervision (Milne, 2009). This new approach enhances CBT supervision by addressing the above objections through a developmentally-informed framework, explicit emotional processing of the material presented by the supervisee, higher levels of challenging and experiential learning, an integrated programme of research, systematic supervisor training (Milne, 2010), and the operationalization of EBCS through a manual and an observational instrument for rating supervisory competence (Milne, Reiser, Cliffe and Raine, 2011). This approach is consistent with Bellg et al.'s (2004) recommendations in the area of delivery of treatment: "The gold standard to ensure satisfactory delivery is to evaluate or code intervention sessions (observed in vivo or video- or audiotaped) according to a priori criteria." (p. 446).

While a limited amount of literature exists reviewing the overall outcomes of CBT supervision (Milne and James, 2000; Townend et al., 2002; Wheeler and Richards, 2007; Milne et al., 2010), the literature comparing outcomes of different supervision methods is even more restricted. In a randomly controlled trial comparing CBT and psychodynamic supervision, Bambling, King, Raue, Schweitzer and Lambert (2006) found no difference in patients' scores on a standardized depression measure at the end of an 8-session treatment period. However, analysis of outcomes of patients assigned to a third arm of the study, who were treated in a problem-solving therapy without supervision, suggested that supervision of either kind improved patient outcome. Similarly, Uys, Minnaar, Simpson and Reid (2005) found that the two supervision approaches that they compared (i.e. a developmental model, and one based on Holloway's 1995 matrix model) both produced similarly significantly improved supervisee ratings of supervision.

A number of previous studies of supervision have used a single subject ( $N = 1$ ) design in order to evaluate the effectiveness of EBCS and CBT supervision. Milne and Westerman (2001) studied the effects of fortnightly consultation (supervision-of-supervision) on the clinical supervision of three supervisees over an 8-month period, using a multiple baseline

design. Results of the study suggested that targeted supervisor behaviours (i.e. increased use of guided experiential learning) increased in frequency over time, and across supervisees and phases. In a second  $N = 1$  study, Milne and James (2002) analyzed the impact of consultancy on the effectiveness of routine CBT supervision with one supervisor and six supervisees. The results again indicated that the supervisor could develop a CBT approach with the aid of consultancy, adding corroborating evidence in terms of the supervisees' increased experiential learning. However, to our knowledge there has not yet been a direct comparison between CBT and another method of supervision.

### *Hypotheses of the study*

This paper builds on these prior  $N = 1$  studies by reporting a comparative evaluation. Our major hypotheses were as follows:

1. We would be able to reliably manipulate the EBCS and CBT supervision conditions (i.e. fidelity would be demonstrated); and
2. EBCS supervision would demonstrate significantly stronger effects, in terms of enhanced supervisee engagement in experiential learning.

## **Method**

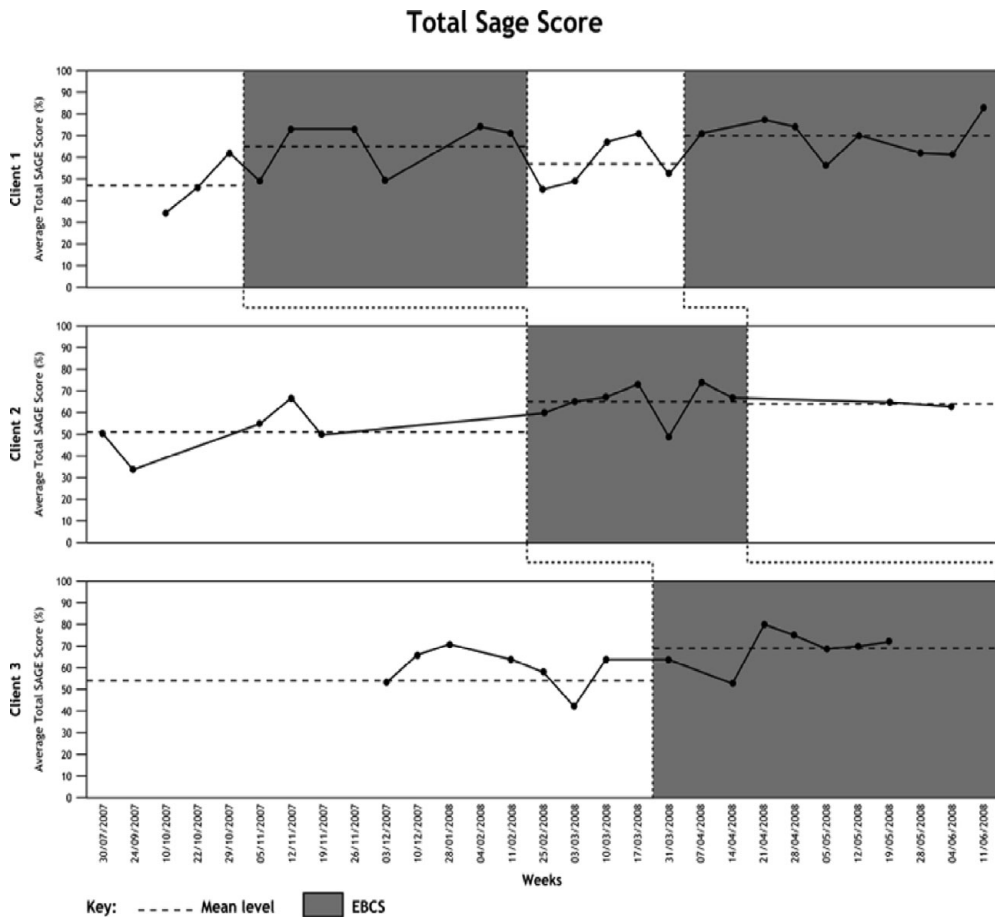
### *Design*

We utilised a multiple-baseline across participants, alternating treatments design. This design is considered appropriate for this kind of comparison (Oliver and Fleming, 1997; Borckardt et al., 2008), especially given that both approaches were still in need of systematic operationalization, and that only a few rigorous  $N = 1$  evaluations have been conducted (Milne, 2008).

To explore the hypotheses, we utilized ABA and ABAB phases across three clients (see Figure 1). These phases included 37 consecutive, audio-taped sessions of supervision over an 11-month period. A minimum of three participants is recommended, to allow for accurate visual inspection (Hersen and Barlow, 1976); whilst at least 20 data points are recommended to detect a large effect size by statistical means (Onghena and Edgington, 2005).

### *Instruments*

We rated audio-taped sessions of supervision using a direct observation instrument called SAGE (Supervision: Adherence and Guidance Evaluation: Milne, Reiser, Cliffe and Raine, 2011). SAGE is a competence rating tool, consisting of 23 supervisor and supervisee behaviours. The first 18 items of SAGE were used to rate the supervisor's competence (e.g. "agenda-setting" and "giving feedback"). Each item was rated on a 7-point competence rating scale, ranging from "incompetent" to "expert", following the revised cognitive therapy scale (Blackburn et al., 2001). These ratings were then aggregated into an overall "supervisor competence score", by summing up individual scores for items 1–18, and expressing them as a percentage of the total possible score ( $18 \times 7 = 126$ ). These data allowed us to test the first hypothesis. The final five SAGE items (19–23) reflected supervisees' engagement in experiential learning within each supervision session (for example, signs of the supervisee



**Figure 1.** The multiple baseline across participants (clients)  $N = 1$  research design, with the ratings of supervision (SAGE) during the CBT and EBCS phases of the study. This Figure displays the data for the supervisor (i.e. SAGE items 1–18)

“reflecting” and “conceptualizing”), and were similarly aggregated to give a total percentage “supervisee learning” score for each supervision session. This allowed us to test the second hypothesis. SAGE has promising inter-rater reliability ( $r = .82$ ) and validity, following Landis and Koch (1977), as detailed in Milne et al. (2010). For example, in terms of content validity, the 23 items of SAGE were derived from two existing instruments: “Teacher’s PETS” (Milne, James, Keegan and Dudley, 2002) and “CBT STARS” (James, Blackburn, Milne and Freeston, 2005). Construct validity was assessed by factor analysis, using Principal Axis Factoring. The KMO measure of Sampling Adequacy was 0.96. Examination of the Eigenvalues (19.06, 1.15, 0.71, 0.62 for the first four, respectively) and the Scree plot indicated a single factor, Supervisory Competence, accounting for 76.6% of the variance. Internal consistency was 0.98.

### *Participants*

The study site was a community-based psychology training clinic in the US serving adults presenting with complex mental health problems. The participants were one male consultant (i.e. the supervisor of the supervisor), one male supervisor, one female therapist (i.e. supervisee) and three clients (two males and one female) presenting with anxiety and depression. The consultant was a doctoral-level, licensed (i.e. registered) clinical psychologist, with over 30 years of experience in providing supervision and training to psychologists and other mental health professionals in CBT. The supervisor was a licensed psychologist certified by the Academy of Cognitive Therapy with over 20 years of clinical experience; and a training clinic director with over 6 years of experience in providing doctoral-level training. The supervisee was a post-doctoral student.

This was a convenience sample, having been selected on the pragmatic grounds that the supervisor had invited the consultant to engage in a collaborative research study. Similarly, the supervisees were currently working with the supervisor at the time of the study; and the three clients who completed all study phases (three clients discontinued therapy or did not consent) were seen by the second supervisee (i.e. the therapist). There were no exclusion criteria and the sample was considered to be representative of supervisees and clients in this clinic. The study was approved by the Palo Alto University Institutional Review Board in the USA; and by the Research and Development Department of the consultant's employing Trust in the UK. This Department deemed that additional ethical clearance was not required, as this was a "service improvement" project. Informed consent was obtained from the supervisees and the clients.

### *Procedure*

Supervisor training took place over 5 consecutive weeks immediately prior to the initial study baseline, following the consultancy procedure described below. Training practices were consistent with Bellg et al. (2004) recommendations for improving provider training, including: "... using standardized training materials, conducting role-playing, and observing actual intervention and evaluating adherence to protocol." (p. 446). The supervisor quickly acquired competence, as indicated by an improvement from a pre-training SAGE score of 55% to a score of 80% by the end of the training, indicating proficiency.

During the training, baseline and experimental phases, consultancy entailed fortnightly, phone-based, hour-long reviews of the preceding week's tape recorded supervision (i.e. supervision-of-supervision). Consultancy involved support and training utilizing didactic methods, including reference to supervision guidelines and to other supportive materials (such as a client experiencing scale, a supervisee's learning outcomes list, and illustrative video material: Milne, 2009). During the CBT training phase, supervision guidelines were based on authoritative texts (Padesky, 1996; Liese and Beck, 1997), scientific papers (e.g. Pretorius, 2006), and guidance from The British Association for Behavioural and Cognitive Psychotherapy (e.g. Lewis, 2005). During the EBCS training phase, a supervisor training manual, including supervision guidelines (Milne, 2010), was supplemented by experiential work including educational role-plays and behavioural rehearsal. Corrective feedback to the supervisor was based on the consultant's ratings of his supervision, using SAGE based on scrutinizing audiotapes of supervision. During the CBT (baseline) phases, the supervisor

adhered to the CBT supervision condition (e.g. the supervisee would not be exposed to the same emotional processing or experiential learning as in the experimental phase). The objective was to try and demonstrate experimental control over these two approaches (i.e. high fidelity).

All supervision sessions were recorded on audiotape by the supervisor. A total of 37 supervision sessions were recorded, each lasting approximately one hour. This sample was selected as it included all recorded supervision sessions in which there was a discussion of one or more of the three clients, during the entire 11-month study period.

### *Statistical analysis*

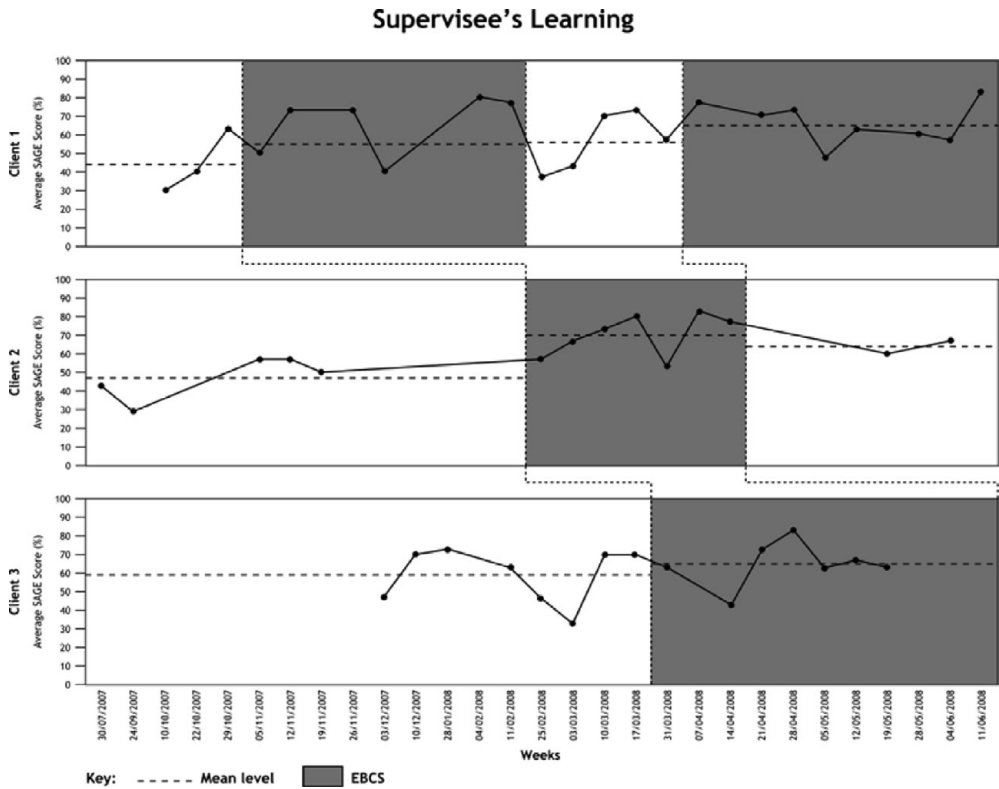
The current convention within  $N = 1$  statistical analysis is a combination of visual inspection of the descriptive (longitudinal) data and inferential statistics. However, visual inspection is prone to low reliability and Type I errors, as are the usual parametric tests. Therefore, experts recommend conservative inspection (e.g. summarizing differences in mean scores between phases) and statistical tests that are appropriate for  $N = 1$  data and that address the problem of auto-correlation (serial dependence) within longitudinal data (Borckardt et al., 2008). We therefore provide summary mean scores alongside the raw  $N = 1$  data, and apply the non-parametric PAND statistic (Percentage of All Non-Overlapping Data: Parker, Hagen-Burke and Vannest, 2007). This provides a quantitative comparison by determining how many observations overlap between the ABAB phases. The percentage overlap converts to phi, allowing significance testing by calculating confidence intervals using an online resource (Pezzullo, 2008).

## **Results**

### *Hypothesis 1*

As illustrated in Figure 1, visual inspection suggests that the EBCS phases yielded the highest SAGE scores for all three clients, though there is considerable variability. Mean values for clients in each of the phases (CBT or EBCS) is indicated by the broken horizontal lines in Figure 2, and supports this interpretation (i.e. in every instance, the EBCS phases have higher mean SAGE scores than the CBT phases), indicating good experimental control of the alternating interventions: CBT supervision and EBCS. A sub-analysis of the full data set (i.e. considering all 23 individual items within SAGE) supported this interpretation, with ratings 0.63 points higher, on average, during EBCS phases for the key SAGE variables of “emotional engagement”, “challenging”, “training” and “experiencing”. Conversely, CBT supervision was rated more highly for “conceptualizing” (mean: 0.2), reflecting the relative emphasis on case discussion within this approach.

We next analyzed the SAGE data using an inferential statistic. Participants contributed 8, 5 and 7 data points respectively in the first phase, and 14, 7, and 7 (respectively) in the second (see Figure 1). PAND was 88.2% and the non-overlap beyond chance level (50%) was 38.2%;  $\phi = 0.762$ . The confidence intervals (95%) for the phi coefficient were 0.605 and 0.862. As these confidence intervals do not include 0, this indicates a statistically significant difference between the phases ( $p < .05$ ). Although guidelines for the interpretation of effect sizes in single case research are not yet well established (Parker et al., 2007), a phi of this



**Figure 2.** Data for the supervisee's learning (i.e. SAGE items 19–23) during the CBT (white) and EBCS (shaded) phases of clinical supervision

size would elsewhere be considered a “moderate” effect size (Cohen, 1988). In summary, the quantitative (SAGE) data indicated that the two supervision approaches were implemented with fidelity.

### *Hypothesis 2*

As regards the relative effectiveness of EBCS over CBT supervision, quantitatively (i.e. SAGE ratings) we obtained trends for each successive client's supervision that indicated consistently that EBCS resulted in greater engagement in experiential learning by the supervisee. Visual inspection of Figure 2 suggests that the best results occurred during EBCS, and similarly that the mean values within the EBCS phases all exceeded those within the CBT phases. The two phases, CBT supervision and EBCS, were then compared across all three clients combined (to increase power), in terms of the supervisee's learning data (i.e. the relevant SAGE items, 19–23: see Figure 2), using the non-parametric statistic PAND.

Participants contributed the same data points as already noted, and the PAND value was again 88.2% in both cases, giving the same confidence intervals, phi coefficient and statistically significant difference between the supervisee's learning in the CBT and EBCS

phases ( $p < .05$ ). This would be considered a “moderate” effect size (Cohen, 1988). Expressed in terms of the observed differences in the SAGE learning scores between phases for all three clients, EBCS was associated with enhanced engagement in experiential learning on four of the five SAGE learning items (mean: 32% improvement).

### Discussion

Our first hypothesis was that the two supervision approaches, CBT and EBCS, could be reliably manipulated. Such fidelity is a basic logical requirement for process-outcome research, also increasing power (Lipsey, 1990). However, it is rare for studies in the supervision field to include demonstrations of fidelity (Watkins, 1997).

Using quantitative ratings of supervision, based on direct observation (i.e. SAGE), we were able to show that the CBT and EBCS approaches could be manipulated with good fidelity between the experimental phases. Our second main finding was that the CBT and EBCS approaches appeared to yield predictably different effects, as measured in terms of the supervisee’s engagement in experiential learning. Visual inspection and non-parametric statistical testing suggested that EBCS promoted significantly more engagement in experiential learning. However, it is evident from Figure 1 that the supervisor became more proficient over time across both conditions. A similar finding was reported by Milne and James (2002), who explained the progression in terms of a lag effect (i.e. the study design did not allow sufficient time for the full socialization of supervisor and supervisees to the new approach).

We should also recognize a number of additional methodological weaknesses within the present analysis, including the need to do further psychometric work on SAGE (we have commenced a generalizability study); to add complementary instruments, such as objective assessments of learning and the transfer of the supervisees’ learning to therapy; and to refine training in EBCS (see, for example, Culloty, Milne and Sheikh, 2010). Another weakness was our use of an experienced supervisor who, with an established allegiance to his usual approach (CBT supervision), might be less markedly influenced by training and consultancy, a common finding within therapist-training studies (Beidas and Kendall, 2010). A related confound was the respective guidelines: in the case of EBCS there was a formally-developed supervisor training manual (Milne, 2010), whereas in the CBT supervision we relied on a range of informal guidance (e.g. book chapters: Padesky, 1996; Liese and Beck, 1997). On the other hand, both approaches were guided by the consultant’s experience of both EBCS and CBT supervision. Finally, we acknowledge that there are inherent methodological weaknesses in  $N = 1$  designs, including significant limitations on the generalizability of the study and an inability to rule out confounding variables or alternative explanations for results, due to the inability to control for confounding variables through randomization (Harris et al., 2006).

A more substantive issue concerns the differentiation of our two conditions, CBT supervision and EBCS. It could be argued that EBCS is simply CBT supervision done properly (i.e. EBCS is not conceptually different from CBT supervision, even if there are differences in implementation). After all, authoritative accounts of CBT supervision advocate much that is in EBCS (e.g. Liese and Beck, 1997; Padesky, 1996), such as educational role-play and mutual feedback. However, in addition to a differential emphasis on some shared variables (e.g. EBCS stresses the behavioural and affective aspects of supervision), EBCS appears to be conceptually distinct in terms of drawing on ideas about adult learning from beyond the scope of CBT supervision literature (Milne, 2009).



A final reason to believe that CBT supervision and EBCS are different is the recurring finding of non-equivalence in the respective processes and outcomes across the various measures, as summarized above, and related findings from a parallel qualitative evaluation (i.e. episode and content analysis, plus interviews: Milne, Reiser, Cliffe, Breese et al., 2011). On the other hand, a more pragmatic stance is to view EBCS as an enhanced form of CBT supervision, on the logic that conceptually EBCS is compatible with the principles of CBT (e.g. the empirical emphasis) and because practically the technology that supports EBCS can readily support the development of CBT supervision (e.g. the supervisor training manual). This seems to us to be a reasonable position, given the methodological limitations of the present study and the pressing need for progress.

### *Suggestions for future research*

Future research might draw on these methodological points, so that more systematic comparisons can be undertaken. This may help identify the key conceptual and procedural features (e.g. parallel manuals for CBT and comparison approaches), relevant instruments, and the most effective elements within such approaches (Kazdin, 1998). We can envisage how this improved operationalization might build on the emergent competencies frameworks (Roth and Pilling, 2011; Falender et al., 2004), as might suitable instruments, such as SAGE. In turn, this will facilitate supervisor training. Indeed, one might helpfully view the challenge of advancing supervision research from the perspective of treatment fidelity (Bellg et al., 2004), as in requiring an integrated approach to the successive tasks of intervention design, practitioner training, delivering competent supervision, and achieving relevant short-term and longer-term outcomes (e.g. generalization of supervision to therapy). Developments within the field should be wedded to this strategy, such as clinical outcome-monitoring (e.g. Reese et al., 2009). In addition, it should be noted that other statistical approaches to the analysis of  $N = 1$  data are available and offer certain advantages, particularly for longitudinal analyses (McKnight, McKean and Huitema, 2000), and for estimating effect sizes (Parker, Vannest and Davis, 2011) and it is possible that such an analysis would yield different conclusions in the present study.

### **Conclusions**

This intensive case study indicates the potential for multiple-baseline  $N = 1$  designs to advance our understanding of CBT supervision, moving supervision towards its rightful place as a science-informed specialization within professional practice. The present analysis was novel in demonstrating the fidelity and comparative effectiveness of two closely-related supervision approaches. However, as a preliminary, small-scale analysis, it is appropriate to next undertake improved  $N = 1$  studies, before proceeding to larger sample evaluations. The present evaluation indicates that the present methodology is viable, illuminating, and is capable of fulfilling the expectations placed on it (e.g. Falender et al., 2004).

### **References**

- Bambling, M., King, R., Raue, P., Schweitzer, R. and 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.

- Beidas, R. S. and Kendall, P. C.** (2010). Training therapists in evidence-based practice: a critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, 17, 1–30.
- Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., et al.** (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH behavior change consortium. *Health Psychology*, 23, 443–451.
- Blackburn, I-M., James, I. A., Milne, D. L., Baker, C., Standart, S. H., Garland, A., et al.** (2001). The revised Cognitive Therapy Scale (CTS-R): psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29, 431–446.
- Borckardt, J. J., Nash, M. R., Murphy, M. D., Moore, M., Shaw, D. and O’Neil, P.** (2008). Clinical practice as natural laboratory for psychotherapy research. *American Psychologist*, 63, 77–95.
- Cohen, J.** (1988). *Statistical Power Analysis for the Behavioural Sciences*. New York: Lawrence Erlbaum.
- Culloty, T., Milne, D. L. and Sheikh, A. I.** (2010). Evaluating the training of clinical supervisors: a pilot study using the fidelity framework. *The Cognitive Behaviour Therapist*, 3, 132–144.
- 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.
- Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., et al.** (2006). The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Information Association*, 13, 16–23.
- Hersen, M. and Barlow, D. H.** (1976) *Single-case Experimental Designs: strategies for studying behavioral change*. New York: Pergamon.
- Holloway, E. L.** (1995). *Clinical Supervision: a systems approach*. Thousand Oaks, CA: Sage.
- Improving Access to Psychological Therapies (IAPT)** (2011). *Guidance for Commissioning IAPT Supervisor Training*. London: Department of Health.
- James, I. A., Blackburn, I. M., Milne, D. L. and Freeston, M.** (2005). Supervision training and rating scale for cognitive therapy (STARS - CT). (unpublished manuscript). Newcastle Cognitive Therapy Centre, England.
- Kazdin, A. E.** (1998). *Research Design in Clinical Psychology*. Boston: Allyn Bacon.
- Landis, J. R. and Koch, G. G.** (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lewis, K.** (2005). *The Supervision of Cognitive and Behavioural Psychotherapists*. *BABCP Magazine, Supervision Supplement*, 33, Accrington: BABCP.
- Liese, B. S. and Beck, J. S.** (1997). Cognitive therapy supervision. In C. E. Watkins (Ed.), *Handbook of Psychotherapy Supervision* (pp.114–133). Chichester: Wiley.
- Lipsey, M. W.** (1990). *Design Sensitivity*. London: Sage.
- McKnight, S. D., McKean, J. W. and Huitema, B. E.** (2000). A double bootstrap method to analyze linear models with autoregressive error terms. *Psychological Methods*, 5, 87–101.
- Milne, D. L.** (2008). CBT supervision: from reflexivity to specialization. *Behavioural and Cognitive Psychotherapy*, 36, 779–786.
- Milne, D. L.** (2009). *Evidence-Based Clinical Supervision: principles and practice*. Chichester: BPS/Blackwell.
- Milne, D. L.** (2010). Can we enhance the training of clinical supervisors? A national pilot study of an evidence-based approach. *Clinical Psychology and Psychotherapy*, 17, 321–328.
- Milne, D. L. and James, I. A.** (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology*, 39, 111–127.
- Milne, D. L. and 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., James, I. A., Keegan, D. and Dudley, M.** (2002). Teachers PETS: a new observational measure of experiential training interactions. *Clinical Psychology and Psychotherapy*, 9, 187–199.

- Milne, D. L., Reiser, R., Aylott, H., Dunkerley, C., Fitzpatrick, H. and Wharton, S.** (2010). The systematic review as an empirical approach to improving CBT supervision. *International Journal of Cognitive Therapy*, 3, 278–294.
- Milne, D. L. and Reiser, R. P.** (2011). Observing competence in CBT supervision: a systematic review of the available instruments. *The Cognitive Behaviour Therapist*, 4, 123–138.
- Milne, D. L., Reiser, R. P., Cliffe, T., Breese, L., Boon, A., Raine, R., et al.** (2011). A qualitative comparison of cognitive-behavioural and evidence-based clinical supervision. *The Cognitive Behaviour Therapist*, 4, 152–166.
- Milne, D. L., Reiser, R. P., Cliffe, T., and Raine, R.** (2011). SAGE: preliminary evaluation of an instrument for observing competence in CBT supervision. *The Cognitive Behaviour Therapist*, 4, 123–138.
- Milne, D. L. and Westerman, C.** (2001). Evidence-based clinical supervision: rationale and illustration. *Clinical Psychology and Psychotherapy*, 8, 444–457.
- Newman, C. F.** (2010). Competency in conducting cognitive-behavioral therapy: foundational, functional, and supervisory aspects. *Psychotherapy: Theory, Research, Practice, Training*, 47, 12–19.
- Oliver, J. R. and Fleming, R. K.** (1997). Applying within-subject methodology to the transfer of training research. *International Journal of Training and Development*, 1, 173–180.
- Onghe, P. and Edgington, E. S.** (2005). Customization of pain treatments: single case design and analysis. *Clinical Journal of Pain*, 21, 56–68.
- Padesky, C. A.** (1996). Developing cognitive therapist competency: teaching and supervision models. In P. M. Salkovskis (Ed.), *Frontiers of Cognitive Therapy*. London: Guilford Press.
- Parker, R. I., Hagan-Burke, S. and Vannest, K.** (2007). Percentage of all non-overlapping data (PAND): an alternative to PND. *Journal of Special Education*, 40, 194–204.
- Parker, R. I., Vannest, K. J. and Davis, J. L.** (2011). Effect size in single-case research: a review of nine non-overlap techniques. *Behavior Modification*, 35, 303–322.
- Pezzullo, J. C.** (2008). Two-way contingency table analysis: online resource for calculating Phi coefficient. Accessed on 13 May 2008 from <http://statpages.org/ctab2x2.html>.
- Pretorius, W. M.** (2006). Cognitive-behavioural therapy supervision: recommended practice. *Behavioural and Cognitive Psychotherapy*, 34, 413–420.
- Reese, R. J., Usher, E. L., Bowman, D. C., Norsworthy, L. A., Halstead, J. L., Rowlands, S. R., et al.** (2009). Using client feedback in psychotherapy training: an analysis of its influence on supervision and counselor self-efficacy. *Training and Education in Professional Psychology*, 3, 157–168.
- Roth, A. and 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)
- Townend, M., Iannetta, L. and 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.
- Townend, M., Iannetta, L., Freeston, M. and Harvey, J.** (2007). *Supervision practices of UK cognitive-behavioural psychotherapists, 2001 and 2006*. Paper presented at the 5<sup>th</sup> World Congress of Behavioural and Cognitive Therapies, Barcelona, 11–14 July.
- Uys, L. R., Minnaar, A., Simpson, B. and Reid, S.** (2005). The effect of two models of supervision on selected outcomes. *Journal of Nursing Scholarship*, 37, 282–288.
- Watkins, C. E.** (1997). *Handbook of Psychotherapy Supervision*. Chichester: Wiley.
- Watkins, C. E.** (1998). Psychotherapy supervision in the 21<sup>st</sup> century. *Journal of Psychotherapy Practice and Research*, 7, 93–101.
- Wheeler, S. and Richards, K.** (2007). The impact of clinical supervision on counselors and therapists, their practice and their clients: a systematic review of the literature. *Counseling and Psychotherapy Research*, 7, 54–65.