

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/6299628>

# An empirical definition of supervision

**Article** in *British Journal of Clinical Psychology* · December 2007

DOI: 10.1348/014466507X197415 · Source: PubMed

---

CITATIONS

123

READS

4,686

## 1 author:



Derek Milne

Newcastle University

**189** PUBLICATIONS **2,974** CITATIONS

SEE PROFILE

## Some of the authors of this publication are also working on these related projects:



I'm writing a second edition of my main book (Evidence-based clinical supervision). With Bob Reiser, just finished writing a manual for Evidence-based/CBT supervision. [View project](#)



A Manual for Evidence-based Cognitive and Behavioral (CBT) Supervision [View project](#)



## An empirical definition of clinical supervision

Derek Milne\*

Newcastle University, and Northumberland, Tyne and Wear NHS Trust, UK

**Objectives.** The growing recognition of clinical supervision as the basis for high-quality mental health services is apparent in policy, research and clinical practice, but an empirical definition is required to progress research and practice.

**Method.** A logical analysis was used to draft a working definition, and then a systematic review of 24 empirical studies of clinical supervision produced a best evidence synthesis, which was used to test and improve this definition.

**Results.** The logical analysis indicated that the most popular definition (Bernard & Goodyear, 1992) failed all four necessary tests of a good definition: precision, specification, operationalization and corroboration. The systematic review synthesis was then used to test the working definition, which passed these tests (with two amendments).

**Conclusion.** These two complementary review approaches created a firmer basis for advancing research and practice.

Clinical supervision is increasingly recognized as a vital part of modern, effective health care systems. This is indicated in the UK by National Health Service (NHS) policy (e.g. 'A First Class Service', Department of Health, 1998), guidelines within the NHS professions (e.g. British Psychological Society, 2005) and by international research findings (e.g. Tharenou, 2001). But the exact nature of supervision is not as clear-cut, impeding research and practice. How should it be defined?

There appears to be an uncritical acceptance of the Bernard and Goodyear (2004) definition. They defined supervision as: '... an intervention provided by a more senior member of a profession to a more junior member or members of that same profession. This relationship is evaluative, extends over time, and has the simultaneous purposes of enhancing the professional functioning of the more junior person(s), monitoring the quality of professional services offered to the clients, she, he, or they see, and serving as a gatekeeper for those who are to enter the particular profession' (p. 8).

This definition is widely accepted in the USA, as indicated by a recent consensus statement (Falender *et al.*, 2004), and by its unchallenged incorporation into the Handbook of Psychotherapy Supervision (Watkins, 1997). It is argued here that the

\*Correspondence should be addressed to Dr Derek Milne, Doctorate in Clinical Psychology, Fourth Floor, Ridley Building, Newcastle University, NE1 7RU, UK (e-mail: d.l.milne@ncl.ac.uk).

Bernard and Goodyear's (2004) definition is unsatisfactory, as are the other representative definitions (i.e. British Association for Behavioural and Cognitive Psychotherapy, 2006; British Association for Counselling and Psychotherapy, 1996; British Psychological Society, 2003 and 2005; Carroll & Tholstrup, 2001; Scaife, 2001). None of them offer the requisite clarity, hampering our research and practice.

To address this definitional problem, a two-stage critical review of the concept of supervision follows. In stage one, the logical foundations of a sound definition will be clarified by deductive reasoning, the philosophy of good empirical research (Kazdin, 1998) and by reference to prior reasoning and scientific conventions. These criteria will then be applied to Bernard and Goodyear's (2004) definition. In the second part of the review, a logically derived working definition is tested, against empirical research. This is done by producing a 'best evidence synthesis', a variant of the systematic review (Petticrew & Roberts, 2006). As a result, an empirically derived definition of clinical supervision can be developed.

## Logical review

### **Four criteria for an empirical definition**

A definition needs to state the precise, essential meaning of a word or concept in a way that makes it distinct, the '*precision*' criterion (Concise Oxford English Dictionary; COED: 2004). One accepted way to do this is to draw out comparisons and cite examples (Thouless, 1930), in order to differentiate and set out boundaries (e.g. stating that supervision is not therapy or mentoring). This clarity is necessary for the three remaining criteria.

Secondly, we need to define any ingredients that make up supervision with similar precision, termed '*specification*' (i.e. 'A detailed description of the elements that make up something': COED, 2004). 'Hypothesis validity' refers to the extent to which research accurately reflects relationships between such constructs (Wampold, David, & Good, 1990). Definition is a prerequisite of such validity, as without it we can have 'inconsequential', 'ambiguous' or 'non-congruent' research hypotheses. Hypothesis validity threats were amongst those included in the systematic review of supervision carried out by Ellis, Ladany, Kregel, and Schult (1996). They concluded, from their review of 144 studies of clinical supervision, that this form of validity was typically not properly explicated. Ultimately, precision and specification permit manipulations such as supervision to be replicated and are the basis for the accurate interpretation of results and for practice developments (e.g. improving the way by which we train supervisors).

Thirdly, concepts such as supervision require *operationalization*, that is, they need to be stated in a form that permits measurement. This is achieved either by creating a new instrument or by relating a definition to an established, psychometrically sound measure (Kazdin, 1998). Again, an operational definition allows one to state hypotheses with validity (Wampold *et al.*, 1990) and the independent variable to be manipulated reliably (Barker, Pistrang, & Elliott, 2002). Reliable observation of the intervention (a fidelity check) and the subsequent evaluation of supervision can then follow (Borelli *et al.*, 2005; MacDonald, 2002).

The fourth and final necessary condition for an empirical definition is that it has received adequate support from research. This '*corroboration*' (i.e. something that 'confirms or gives support': COED, 2004) is required because a definition could theoretically be precise, specified and operationalized, yet be speculative or invalid.

**The four criteria applied to the Bernard and Goodyear's definition**

Unfortunately, Bernard and Goodyear's (2004) definition illustrates several of these difficulties:

- (a) The sort of 'intervention' that they allude to is not defined precisely: what exactly does the supervisor do? It is presumably an 'educational' intervention, but the reader is left to assume what this might entail. As a result, replication is impossible.
- (b) Also, their definition does not list the variables making up the supervision 'intervention'. For example, is it restorative, formative and/or normative in emphasis? How exactly are any such variables manipulated? This lack of specification makes reliable implementation (fidelity) impossible.
- (c) No instrument is suggested, to clarify what is to be measured, making operationalization problematic.
- (d) This lack of clarity would also prevent corroboration, as valid hypotheses could not be constructed, rendering findings uninterpretable.

In summary, a logical analysis of the popular Bernard and Goodyear's (2004) definition finds it deficient on all four necessary conditions.

**An improved working definition of clinical supervision**

Based on these considerations, the following is a provisional but logically improved definition of supervision (with these four necessary criteria noted in brackets):

The formal provision, by approved supervisors, of a relationship-based education and training that is work-focused and which manages, supports, develops and evaluates the work of colleague/s (precision). The main methods that supervisors use are corrective feedback on the supervisee's performance, teaching, and collaborative goal-setting (specification). It therefore differs from related activities, such as mentoring and coaching, by incorporating an evaluative component (precision by differentiation). An example is Gillam, Strike-Roussos, and Anderson (1990), who described supervision in terms of a discussion of the supervision model and of the respective roles of supervisor and supervisee, joint goal-setting, observation of the supervisee, and evaluation. They helpfully appended an outline of these meetings, consisting of 18 agenda items (precision through examples). Supervision's objectives are "normative" (e.g. quality control), "restorative" (e.g. encourage emotional processing) and "formative" (e.g. maintaining and facilitating supervisees' competence, capability and general effectiveness) (specification by identifying the functions served). These objectives could be measured by current instruments (operationalization). This definition is supported by recent reviews of the empirical literature (e.g. Falender & Shefranske, 2004; Watkins, 1997) and by consensus statements (e.g. Falender *et al.*, 2004) (corroboration). Further qualifying comments are provided in Table 1.

The definition in Table 1 was based on integrating existing definitions (esp. Bernard & Goodyear, 2004; Department of Health, 1993; Proctor, 1998; Watkins, 1997). It includes any supervision format (e.g. group supervision), profession or therapeutic orientation, and it covers pre- and post-qualification supervisions. It excludes staff training, consultancy and performance management (for details see the NICE-R coding manual, used in the present systematic review: see too Milne *et al.*, 2005). This is a working definition in the sense that the results from the systematic review are further intended to test it, using the criteria of precision, specification, operationalization and corroboration.

**Table 1.** The working definition of clinical supervision (with clarifying comments)

‘FORM’ OF SUPERVISION:
<p><b>‘The formal provision</b> (i.e. sanctioned by relevant organization/s);  <b>by senior/qualified health practitioners</b> (or similarly experienced staff)  <b>of an intensive</b> (i.e. typically 1:1 and regular/ongoing, at least, three meetings with protected time, of at least 3 hours total duration),  <b>relationship-based</b> (inc. confidential and highly collaborative, being founded on a learning alliance and featuring (e.g.) participative decision making and shared agenda setting; and therapeutic inter-personal qualities, such as empathy and warmth),  <b>education and training</b> (general problem-solving capacity or ‘capability’ aspect, not just competence enhancement)  <b>that is case-focused</b> (supervisee guides topics and tables material and supervisor typically overlays professional and organizational considerations/standards)  <b>and which supports, directs and guides</b> (inc. also ‘restorative and normative’ topics, addressed by means of professional methods, inc. objective monitoring, feedback and evaluation; and by reference to the empirical and theoretical knowledge-base)  <b>the work of colleagues (supervisees)</b> (inc. CPD/post-qualification colleagues)</p>
‘FUNCTIONS’ OF SUPERVISION:
<p>(1) <b>quality control</b> (inc. ‘gatekeeping’ and safe, ethical practice);  (2) <b>maintaining and facilitating the supervisees’ competence and capability; and</b>  (3) <b>helping supervisees’ to work effectively</b> (inc. promoting quality control and preserving client safety; accepting responsibility and mostly working independently; developing own professional identity; enhancing self-awareness and resilience/effective personal coping with the job; critical reflection and lifelong learning skills).</p>

## The systematic review

The prior systematic reviews of research on clinical supervision by Ellis *et al.* (1996), Ellis and Ladany (1997), and Milne and James (2000) illustrate this methodical scientific approach (National Health Service, 2001; Petticrew & Roberts, 2006).

Based on the logical analysis above and on the supervision literature, it was predicted that:

- (a) *Precision*: the working definition of supervision (Table 1) would capture most aspects of the definitions that are explicated in the reviewed literature (i.e. no more precise definitions are expected to be located, given the combined use of several existing definitions);
- (b) *Specification*: supervision in the sampled literature will be specified in a way that corresponds with the majority of the working definition;
- (c) *Operationalization*: the majority of studies will measure these supervision variables in a way that corresponds to the working definition (i.e. a manipulation check); and
- (d) *Corroboration*: results will indicate that supervision (as defined, manipulated and measured) is effective, in the definition-specified ways (e.g. ‘formative’, as in developing targeted competencies), in the majority of cases.

These represent the logical (a, b) and empirical tests (c, d) of the working definition, conducted with a sample of current studies, selected for their effectiveness and rigour (i.e. representing a fair test). They are tests in the sense that the working definition may fail to satisfy one or more of them, so requiring revision or rejection. An acceptable

empirical definition needs to satisfy all four tests. Table 2 summarizes these criteria, provides an illustration, and notes exactly how they were tested.

## Method

### *Inclusion criteria and coding manual*

Studies were included in the review if they satisfied nine criteria, following Milne and James (2000). These were the studies that: contained an intervention (usually consultancy and/or training) designed to improve the effectiveness of supervision; were field or effectiveness studies (external validity); were drawn from the mental health field; included any profession or theoretical orientation; recorded supervisor and supervisee behaviours (two criteria); had demonstrated effectiveness; were published in a peer-reviewed scientific journal within the past 20 years (i.e. 1986–2005).

A study coding manual and record sheet were developed to ensure that the review process can be conducted reliably (Milne, 2005). Based on the National Institute for Clinical Excellence (NICE) approach to the systematic review explicitly (NHS Centre for Reviews and Dissemination, 2001), the manual contained minimal alterations (to make the NICE method relevant for the present review). This is the 'NICE(R)' review method, where (R) stands for 'revised' (Dunkerley, Milne, & Wharton, 2004).

### *Procedure and reliability*

The articles selected for the review were identified by electronic searching of the following databases: BIDS; CINAHL; Embase; Ingenta Connect; Medline; Psychinfo; Science Direct; and StreetWise. Articles were also identified from previous reviews and books on supervision, by cross-referencing from the located articles, from browsing in promising periodicals and by asking experts to recommend papers. This process located 383 potential supervision reports for inclusion, of which  $N = 24$  satisfied the nine inclusion criteria. These studies were: Demchak and Browder (1990), Ducharme, Williams, Cummings, Murray, and Spencer (2001), Fleming, Oliver, and Bolton (1996), Gillam *et al.* (1990), Hansebo and Kihlgren (2004), Harchik, Sherman, Sheldon, and Strouse (1992), Harkness (1995, 1997), Harkness and Hensley (1991), Hundert and Hopkins (1992), Jenson, Parsons, and Reid (1998), Methot, Williams, Cummings, and Bradshaw (1996), Miller, Yahne, Moyers, Martinez, and Pirritano (2004), Milne and James (2002), Milne, Pilkington, Gracie, and James (2003), Milne and Westerman (2001), Parsons and Reid (1995), Reichelt, James, and Blackburn (2003), Reid, Parsons, Lattimore, and Reade (2005), Reid *et al.* (2003), Richman, Riorden, Reiss, Pyles, and Bailey (1988), Schepis and Reid (1994), Shlomskas *et al.* (2005) and Shore, Iwata, Vollmer, Lerman, and Zarcone (1995). These studies included 13 consultants, 72 supervisors, 499 supervisees and 711 patients. They were drawn largely from the learning disabilities field, involved a range of professional groups, and typically evaluated the effectiveness of supervision in terms of multiple impacts (including changes in the supervisees and their patients). Most studies took place in residential settings.

The author's inter-rater reliability has a mean of 76% exact agreement across all of the criteria in the NICE(R) manual (range: 72–86%) at the outset and 82% at the end of the coding exercise, indicating that coding could be conducted reliably over the review period.

**Table 2.** Summary of the four criteria for an empirical definition, with illustrative studies drawn from the surveyed sample (N = 24). Column three sets out how these criteria were tested in general, and also by specific reference to the illustrative study in column two

Elements of an empirical definition:	Details of illustrative studies from review sample:	How elements of definition were tested (link to <i>illustrative study</i> in Column 2):
<p><b>Precision</b> (i.e. stating the precise, essential meaning of supervision in a distinct way)</p>	<p>Harkness (1997) studied one social worker's supervision of four supervisees (psychologists and social workers), working with adult out-patients in a community mental health setting, defined in terms of considering the links between clinical presentations, therapeutic interventions and outcomes, within a positive working relationship</p> <p>Miller <i>et al.</i> (2004) studied training in motivational interviewing amongst 140 licensed, multi-disciplinary mental health professionals working with substance abuse clients. Their supervision specification included: collaborative problem solving of practice difficulties; direct observation; needs-led role plays (to both model and rehearse skills); asking for and answering questions; and providing feedback</p> <p>Fleming <i>et al.</i> (1996) measured their four supervisors' competence with staff working in learning disability group homes, using a nine-item observational checklist (e.g. 'participative goal-setting'; 'provides feedback').</p> <p>Milne and James (2002) used a micro-outcome approach to evaluate the effectiveness of supervision for six mental health professionals training in CBT. They reported restorative and formative impacts, contributing to quality control and the development of competence in CBT.</p>	<p>Check if the working definition incorporates most of the definitional material in the reviewed sample of studies. (<i>Illustrative study</i>: Shows relationship emphasis, case focus, functions, etc.)</p>
<p><b>Specification</b> (i.e. a detailed description of the elements making up supervision)</p>	<p>Miller <i>et al.</i> (2004) studied training in motivational interviewing amongst 140 licensed, multi-disciplinary mental health professionals working with substance abuse clients. Their supervision specification included: collaborative problem solving of practice difficulties; direct observation; needs-led role plays (to both model and rehearse skills); asking for and answering questions; and providing feedback</p> <p>Fleming <i>et al.</i> (1996) measured their four supervisors' competence with staff working in learning disability group homes, using a nine-item observational checklist (e.g. 'participative goal-setting'; 'provides feedback').</p> <p>Milne and James (2002) used a micro-outcome approach to evaluate the effectiveness of supervision for six mental health professionals training in CBT. They reported restorative and formative impacts, contributing to quality control and the development of competence in CBT.</p>	<p>Compare the working definition with the definitions within this sample. (<i>Illustrative study</i>: Indicates close correspondence with working definition, in terms of: problem-solving effort; objectively monitoring practice; providing feedback; etc.).</p>
<p><b>Operationalization</b> (i.e. the definition permits measurement of supervision)</p>	<p>Fleming <i>et al.</i> (1996) measured their four supervisors' competence with staff working in learning disability group homes, using a nine-item observational checklist (e.g. 'participative goal-setting'; 'provides feedback').</p> <p>Milne and James (2002) used a micro-outcome approach to evaluate the effectiveness of supervision for six mental health professionals training in CBT. They reported restorative and formative impacts, contributing to quality control and the development of competence in CBT.</p>	<p>Assess whether the ways in which supervision is measured within the sample corresponds with the majority of the working definition. (<i>Illustrative study</i>: These competencies are consistent with the 'training' and 'directing' parts of the definition, etc.)</p>
<p><b>Corroboration</b> (i.e. the definition is supported by research findings)</p>	<p>Milne and James (2002) used a micro-outcome approach to evaluate the effectiveness of supervision for six mental health professionals training in CBT. They reported restorative and formative impacts, contributing to quality control and the development of competence in CBT.</p>	<p>Determine the overall and specific effectiveness of supervision within the sample (i.e. were the studies, which corresponded to the working definition, generally successful; did individual studies achieve the outcomes specified in the definition?). (<i>Illustrative study</i>: These findings are in keeping with at least two of the three functions of supervision within the definition.)</p>



## Results

### **(a) The working definition will capture most of the aspects defined explicitly in the reviewed literature**

Definitions in this literature were generally absent. Of the six studies (25% of sample), five satisfied the NICE(R) manual criteria by explicating at least two methods and one function of supervision. However, none of these authors differentiated their definition from related educational activities. Therefore, on the precision criterion, this sample of literature is unsatisfactory and fails to corroborate or falsify the working definition.

### **(b) Supervision will be specified in a way that corresponds with the majority of the variables within the working definition**

Twenty-three of the 24 studies were rated as having specified distinct variables within their supervision manipulations. Some of these were within relatively simple interventions, whereas other researchers had detailed more complex accounts of supervision (e.g. Demchak & Browder, 1990). These specifications agreed with the working definition as regards the formative function and the supervision methods. However, the normative and restorative aspects were not reported to any significant extent. Therefore, the working definition fails this part of the test and requires amendment to exclude these two functions.

### **(c) The majority of supervision manipulations will be measured in a way that corresponds to the working definition**

For a study to be credited with satisfactory dependent variable measurement, the NICE(R) manual requires that 'all or most of the variables are actually measured'. Sixteen (67%) of the 24 studies satisfied this criterion. These 16 studies utilized measures that agreed with the working definition (see Table 2 for an example).

### **(a) Results from the majority of the studies will indicate that supervision (as defined and manipulated) is effective (corroboration)**

The outcomes obtained by the 24 reviewed studies were expressed in terms of a standardized rating of the effectiveness of supervision (following Milne & James, 2000). This simple 7-point rating is based on a summary of the studies' reported outcomes, where -3 is ascribed to a study reporting very negative impacts, 0 to no effect at all and 3 to the most positive effect. These rating categories were defined in terms of the proportion of participants who were affected by the supervision intervention (see the NICE-R manual). For example, if more than 75% of supervisees were found to have achieved significantly higher scores on a quiz following supervision, then this study's effectiveness would be rated as a 3. As noted above, these ratings have been made reliably. This procedure is detailed in Milne, Aylott, Dunkerley, Wharton, and Fitzpatrick (2007). This simple overall rating (across studies and all outcome variables therein) was used in place of meta-analysis because the relevant criteria for a meta-analysis were not satisfied (e.g. hypotheses within sampled studies were not identical, nor were their dependent variables).

The obtained effectiveness ratings indicated a mean value of 2.4 for supervisees (therapists' learning) and 2.3 for patients (clinical outcomes), which indicates that positive results were obtained overall, equivalent to 80% and 77% supervisory



effectiveness, respectively. These results indicate that supervision, as defined and measured according to the present working definition, is associated with positive outcomes. As illustrated in Table 2, this is supported by the effectiveness of specific supervisory interventions, in relation to predicted study outcomes.

## Discussion

The logical analysis proposed four necessary conditions for a definition: precision, specification, operationalization and corroboration. Using these criteria, it was possible to improve the popular Bernard and Goodyear's (2004) definition in a working definition (Table 1). This definition was then tested against the best available evidence synthesis, yielding generally favourable results. Twenty-three of the review sample of 24 empirical studies specified their independent variable (supervision) as per the working definition (though the normative and restorative functions were not reported), most went on to measure it in compatible terms (67%), and the great majority furnished evidence supporting the definition (e.g. 77% generalization to patients' clinical outcomes). But these studies were very poor at providing a precise definition, which meant that testing the working definition relied on the three other criteria.

The poor precision is unsurprising, as prior reviews have noted this point (e.g. Hansebo & Kihlgren, 2004; Lyth, 2000). Similarly, surveys of practitioners have revealed confusion and inconsistency over the nature of supervision (e.g. Lister & Crisp, 2005; White *et al.*, 1998). However, the findings on the remaining three criteria for a sound definition (specification, operationalization and corroboration) were satisfied. As there was also no better definition within the reviewed literature, these findings allow us to accept the working definition of clinical supervision (Table 1), with the amendment that the normative and restorative aspects of supervision are deleted.

While this logical analysis and linked systematic review appear to have merit, there are several weaknesses. The best evidence synthesis overcomes the sample heterogeneity that confounds other systematic reviews (i.e. Ellis & Ladany, 1997; Ellis *et al.*, 1996), but it fails to characterize the wider supervision literature, making generalizations inappropriate. Another concern is that the test of corroboration is based on a simple, summary association between the definition and the overall effectiveness of the 24 studies in the review. It is possible that another sample of similar studies that did not correspond with the present definition would achieve even better results. However, these would probably present other interpretational difficulties, being less rigorous, etc. A more promising, comparative strategy would be to arrange the present sample by degrees of success, checking for adherence to the definition. If the definition is corroborated by this criterion, greater adherence to the definition would be expected to be associated with greater success. This might be feasible in the near future, when a larger research study sample exists, and when there is greater specificity about the way supervision was manipulated.

In conclusion, this two-stage review proposed a logically derived working definition of clinical supervision then tested it against a synthesis of the best available evidence. This test indicated that the reviewed studies furnished satisfactory information on the specification, operationalization and corroboration of supervision. This information permitted the working definition of clinical supervision to be refined, and an empirical definition of clinical supervision clarified (i.e. as set out in Table 1, minus the 'restorative' and 'normative' aspects). This provides the basis for dealing with various scientific and practical challenges.

## Acknowledgements

I am grateful to Helen Aylott for her help in completing the search work started by Chris Dunkerley, Helen Fitzpatrick and Sarah Wharton; and thank them all for their input to the reliability assessments. Helpful comments on drafts were made by John Ormrod, Joyce Scaife and Peter Bower. Thanks to Nasim Choudhri for help with the references.

## References

- Barker, C., Pistrang, N., & Elliott, R. (2002). *Research methods in clinical psychology*. Chichester: Wiley.
- Bernard, J. M., & Goodyear, R. K. (1992 & 2004). *Fundamentals of clinical supervision*. London: Pearson.
- Borelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., *et al.* (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behaviour research. *Journal of Consulting and Clinical Psychology, 73*, 852–860.
- British Association for Behavioural and Cognitive Psychotherapy (2006). *BABCP training*. Accrington: BABCP.
- British Association for Counselling and Psychotherapy (1996). *What is supervision? Information sheet 6*. Rugby: BACP.
- British Psychological Society (2003). *Policy guidelines on supervision in the practice of clinical psychology*. Leicester: BPS.
- British Psychological Society (2005). *DCP policy on continued supervision*. Leicester: BPS.
- Carroll, M., & Tholstrup, M. (2001). *Integrative approaches to supervision*. London: Jessica Kingsley.
- Concise Oxford English Dictionary (2004). Oxford: Oxford University Press.
- Demchak, M., & Browder, D. M. (1990). An evaluation of the pyramid model of staff training in group homes for adults with severe handicaps. *Education and Training in Mental Retardation, 25*, 150–163.
- Department of Health (1993). *A vision for the future*. London: NHS Executive.
- Department of Health (1998). *A first class service*. Leeds: Department of Health.
- 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. *Behaviour Modification, 25*(2), 233–254.
- Dunkerley, C., Milne, D. L., & Wharton, S. (2004). *A NICE(R) systematic review of evidence-based clinical supervision*. Conference of the European Association for Behavioural and Cognitive Therapies: University of Manchester.
- 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.
- Ellis, M. V., Ladany, N., Kregel, M., & Schult, D. (1996). Clinical supervision research from 1981–1993: A methodological critic. *Journal of Counselling Psychology, 43*, 35–40.
- Falender, C. A., 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: American Psychological Association.
- 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 Behaviour Management, 16*, 3–25.
- Gillam, R. B., Strike-Roussos, C., & Anderson, J. L. (1990). Facilitating changes in supervisees' clinical behaviours: An experimental investigation of supervisory effectiveness. *Journal of Speech and Hearing Disorders, 55*, 729–739.

- Hansebo, G., & Kihlgren, M. (2004). Nursing home care: Changes after supervisions. *Journal of Advanced Nursing*, 45, 269-279.
- Harchik, A. E., Sherman, J. A., Sheldon, J. B., & Strouse, M. C. (1992). Ongoing consultation as a method of improving performance of staff members in a group home. *Journal of Applied Behaviour Analysis*, 25, 599-610.
- Harkness, D. (1995). The art of helping in supervised practice: Skills, relationships and outcomes. *Clinical Supervisor*, 13, 63-76.
- Harkness, D. (1997). Testing interactional social work theory. *Clinical Supervisor*, 15, 33-50.
- Harkness, D., & Hensley, H. (1991). Changing the focus of social work supervision. *Social Work*, 36, 506-512.
- Hundert, J., & Hopkins, B. (1992). Training supervisors in a collaborative team approach to promote peer interaction of children with disabilities in integrated pre-schools. *Journal of Applied Behaviour Analysis*, 25, 385-400.
- Jenson, J. M., Parsons, M. B., & Read, D. H. (1998). Supervisory training for teachers: Multiple, long term effects in an education programme for adults with severe learning disabilities. *Research in Developmental Disabilities*, 19, 449-463.
- Kazdin, A. E. (1998). *Research design in clinical psychology*. London: Allyn and Bacon.
- Lister, P. G., & Crisp, B. R. (2005). Clinical supervision in child protection for community nurses. *Child Abuse Review*, 14, 57-72.
- Lyth, G. M. (2000). Clinical supervision: A concept analysis. *Journal of Advanced Nursing*, 31, 722-729.
- MacDonald, J. (2002). Clinical supervision: A review of underlying concepts and developments. *Australian and New Zealand Journal of Psychiatry*, 36, 92-98.
- Methot, L. L., Williams, W. L., Cummings, A., & Bradshaw, B. (1996). Measuring the effects of a manager's supervisor training programme through the generalised performance of managers, supervisors, front line staff and clients in a human service setting. *Journal of Organizational Behavioural Management*, 16, 3-34.
- Miller, W. R., Yahne, C. E., Moyers, T. B., Martinez, J., & Pirritano, M. (2004). A randomized trial of methods to help clinicians learn motivational interviewing. *Journal of Consulting and Clinical Psychology*, 72, 1050-1062.
- Milne, D. L., Aylott, H., Dunkerley, C., Fitzpatrick, H., & Wharton, S. (2007). Towards evidence-based training for clinical supervisors. MS in preparation, available from the first author (d.l.milne@ncl.ac.uk).
- Milne, D. L., Dunkerley, C., Wharton, S., Fitzpatrick, H. & Aylott, H. (2005). The NICE(R) manual for systematic reviews of the clinical supervision literature. Unpublished document, available from Derek Milne (d.l.milne@ncl.ac.uk).
- Milne, D. L., & 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., Pilkington, J., Gracie, J., & James, I. A. (2003). Transferring skills from supervision to therapy: A qualitative and quantitative  $N = 1$  analysis. *Behavioural and Cognitive Psychotherapy*, 31, 193-202.
- Milne, D. L., & Westerman, C. (2001). Evidence-based clinical supervision: Rationale and illustration. *Clinical Psychology and Psychotherapy*, 8, 444-457.
- National Health Service (Centre for Reviews and Dissemination) (2001). *Undertaking systematic reviews of research on effectiveness*. York: University Press.
- Parsons, M. B., & Reid, D. H. (1995). Training residential supervisors to provide feedback from maintaining staff teaching skills with people who have severe disabilities. *Journal of Applied Behaviour Analysis*, 28, 317-322.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell.

- Proctor, B. (1998). Supervision: A co-operative exercise in accountability. In: M. Marken and M. Payne (eds). *Enabling and Ensuring*. Leicester: Leicester National Youth Bureau and Council for Education and Training in Youth and Community Work.
- Reichelt, F. K., James, I. A., & Blackburn, I.-M. (2003). Impact of training on rating competence in cognitive therapy. *Journal of Behaviour Therapy and Experimental Psychiatry*, 34, 87-99.
- Reid, D. H., Parsons, M. B., Lattimore, L., & Reade, K. K. (2005). Improving staff performance through clinical application of outcome management. *Research in Developmental Disabilities*, 26, 101-116.
- Reid, D. H., Rotholz, D. A., Parsons, M. B., Braswell, B. A., Green, C. W., & Schell, R. M. (2003). Training human service supervisors in aspects of PBS. *Journal of Positive Behaviour 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 Behavioural Analysis*, 21, 401-409.
- Scaife, J. (2001). *Supervision in the mental health professions*. Hove: Brunner-Routledge.
- Schepis, M. M., & Reid, D. H. (1994). Training direct service staff in congregate settings to interact with people with severe disabilities: A quick, effective and acceptable programme. *Behavioural Interventions*, 9, 13-26.
- Shlomskas, D. E., Syracuse-Sewart, G., Rounsaville, B. J., Ball, S. A., Nuro, K. E., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in CBT. *Journal of Consulting and Clinical Psychology*, 73, 106-115.
- Shore, B. A., Iwata, B. A., Vollmer, T. R., Lerman, D. C., & Zarcone, J. R. (1995). Pyramidal staff training in the extension of treatment for severe behaviour disorders. *Journal of Applied Behaviour Analysis*, 28, 323-332.
- Tharenou, P. (2001). The relationship of training motivation to participation in training and development. *Journal of Occupational and Organizational Psychology*, 74, 599-621.
- Thouless, R. H. (1930). *Straight and crooked thinking*. London: Pan.
- Wampold, B. E., David, B., & Good, R. H. (1990). Hypothesis validity of clinical research. *Journal of Consulting and Clinical Psychology*, 58, 360-367.
- Watkins, C. E. (Ed.). (1997). *The handbook of psychotherapy supervision* (Chapter 1). Chichester: Wiley.
- White, E., Butterworth, T., Bishop, V., Carson, J., Jeacock, J., & Clements, A. (1998). Clinical supervision: Insider reports of a private world. *Journal of Advanced Nursing*, 28, 185-192.

Received 3 August 2006; revised version received 8 March 2007