The Use of Theoretical Models in Psychology Supervisor Development Research From 1994 to 2010: A Systematic Review

Keegan K. Barker and John Hunsley
University of Ottawa

Training in supervision is becoming an increasingly important component of professional psychology programs. To assist students in developing supervisory competencies, knowledge of relevant research and effective training methods is required and, ideally, such knowledge should be derived from an empirical literature comprising good quality, theory-based research. To evaluate the nature of the empirical literature on supervisor development, a systematic review was conducted to determine the extent to which theoretical models have been used in this literature. PsycInfo, Medline, CINAHL, and ERIC were searched for articles published between 1994 and 2010. Initially, 3,248 abstracts were reviewed, 25 of which met the criteria of being empirical, focused on supervisor development, and having participants with graduate training in counselling or professional psychology. Of these studies, only half drew upon models of supervisory development in the conceptualisation, design, or interpretation of the research study. The most often used model was Watkins’ Supervisor Complexity Model (Watkins, 1990, 1993, 1994). Approximately one third of studies examined the influence of past training and experience in supervision on current supervision practices. For competency-based training in supervision to advance, the theoretically informed evidence base needs to be greatly expanded, and much more research is needed that explicitly applies and/or evaluates models of supervisory development.

Keywords: practicum supervision, professional supervision, clinical psychology, counselling psychology, systematic review

Clinical supervision has two main purposes: to ensure the integrity of clinical services and to develop service provision competence in the supervisee (Falender & Shafranske, 2004). Although many definitions of clinical supervision have been proposed, Milne’s (2007) refinement of Bernard and Goodyear’s widely cited work (2004) is the most comprehensive and detailed presented thus far in the literature:

The formal provision, (i.e., sanctioned by relevant organisation/s) by senior/qualified health practitioners (or similarly experienced staff), of an intensive education (general problem solving capacity; developing capability) and/or training (competence enhancement) that is case-focused and which supports, directs, and guides (including restorative and/or normative topics, addressed by means of professional methods, including objective monitoring, feedback and evaluation) the work of junior colleagues (supervisees).

In Canada, competency in supervision is required for licensure by some provincial or territorial regulatory bodies (see http://www.cpa.ca/documents/MRA.pdf). Similarly, the American Psychological Association (APA) has stated that clinical supervision should be regarded as a distinct professional competency (American Psychological Association, 2006). Clinical supervision is also recognised as a key component of prelicensure training in Canada, as accreditation of professional psychology graduate programs requires doctoral students to have instruction in clinical supervision, and internship accreditation requires that interns acquire knowledge and skills in supervision (Canadian Psychological Association, 2011). As the provision of supervision is likely to be a component of future practice for many professional psychologists, graduate programs should provide training in supervision that is, ideally, built on a solid evidence base, as is the case in other areas of professional training (Hunsley & Barker, 2011).

Current perspectives on graduate training in professional psychology emphasise a competency-based approach, with competency being broadly defined as the knowledge, skills, and values required to practice effectively (Falender et al., 2004). The task for educators, then, is to identify the knowledge, skills, and values that constitute a required competency such as supervision and, further, to effectively train students and professionals to develop the competency. In this regard, it is noteworthy that there have been recent efforts to develop benchmarks for professional psychology competencies (including supervision) that incorporate specific behavioural competency criteria keyed to various developmental training stages (Fouad et al., 2009; Kaslow et al., 2009). While the move toward competency-based training is a step toward making supervision training evidence-based, there remain a number of challenges. Most importantly, early reviews of the supervision literature typically concluded that the research base was extremely...
limited and was composed of many methodologically weak studies (e.g., Ellis, Ladany, Krengel, & Schult, 1996). Additionally, it is highly likely that many psychologists currently responsible for providing supervision cannot call upon their own educational experiences to help inform current training efforts, as most have not been formally educated in this area (cf. Howes, Vallis, Wilson, Ross, & Louisy, 1996; Robiner, Saltzman, Hoberman, & Schirvar, 1997).

For training to be evidence-based, there needs to exist a solid literature base that educators can consult to understand the essential components of supervisory competence, the typical trajectories associated with supervisor development, and the best practices for supervision training. An important indicator of the quality of research in any domain is the extent to which theory is used in the conceptualisation and conduct of the research. Although one can reasonably argue that some lines of research (such as surveys of training opportunities or professional practice patterns) need not be theoretically informed, it is difficult to have a cumulative, meaningful evidence base unless most research endeavours in a domain are theoretically driven or, at least, theoretically informed. Similarly, although theories can emerge from a research study (Charmaz, 2003), their validity, credibility, or generalizability must be evaluated within a new research sample or a new study. Thus, a cumulative and cohesive evidence base requires that theory influences how evidence is collected, analysed, and understood (Alderson, 1998; Green, 2000). Moreover, a failure to explicitly describe the theoretical underpinning of a research study may result in these theoretical underpinnings going unnoticed and unappreciated (Alderson, 1998).

As educators and researchers well know, theories serve multiple functions, including explicitly describing key variables to be measured in studying the phenomenon, outlining hypothesised relations among these variables, and, most relevant for our present purposes, providing a framework on which findings from different studies can be integrated into meaningful conclusions. Indeed, trying to draw generalizable conclusions from research studies that are not theoretically informed has been likened to assembling an ever-increasing number of bricks (i.e., study findings) without having an architectural plan (i.e., theory) to guide the placement of the bricks in such a manner as to produce a solid and useful edifice (Forscher, 1963).

In this regard, the conclusions drawn by Ellis and colleagues (1996) on the state of the supervision literature are particularly important. After reviewing 144 empirical studies and evaluating them on their scientific rigour, they found that only 20% of the studies involved explicit tests of relevant supervision theories. As a result, they exhorted researchers to explicitly use relevant theories to address important, unanswered questions about supervision and they developed specific criteria for this future empirical research. One of these criteria was that the research be based clearly on theoretical foundations.

With these issues in mind, the primary goal of the current systematic review was to assess the use of theoretical models of supervisor development in the research literature on supervisor development published after the Ellis et al. (1996) review. Although not a traditional focus of systematic reviews, an examination of the use of theory has been the subject of investigation in a number of other professional literatures, including health behaviour research (Painter, Borba, Hynes, Mays, & Glanz, 2008) and the design and implementation of clinical guidelines (Davies, Walker, & Grimshaw, 2010). We chose to focus on the supervisor development literature for two reasons. First, supervisor development is important because educational efforts are, ultimately, aimed at assisting trainees to develop as supervisors. Moreover, efforts to develop competency benchmarks and to study supervisor development are complementary endeavours: competency benchmarks provide targets toward which we should strive with our training efforts, and the empirical literature on supervisor development should offer guidance on how best to develop and implement training strategies to achieve these targets. Second, understanding the quality of the literature on supervisor development is an important step in moving toward evidence-based training in supervision. Before guidance on training strategies can be gleaned from the literature, we must first have an understanding of the quality of the research that has been conducted. As previously noted, one critical indicator of the quality of research is the extent to which it is theoretically informed. Thus, in our study we systematically searched and reviewed the supervision literature to determine the extent to which theoretical models of supervisor development were used in professional psychology supervisor development research. To provide greater contextual information about the scope of this research, a secondary goal of our review was to describe key methodological characteristics of reviewed studies, such as the type of study design and the characteristics of the research participants.

**Method**

**Phase I: Literature Search**

The primary research question for the systematic review was “To what extent are theoretical models of supervisor development used in professional psychology (clinical, counselling, clinical neuropsychology, school psychology) supervisor development research?” Figure 1 provides information on the selection of the studies included in this review.

PsycInfo served as the main database because the topic area of the review is based in psychology. After the major constructs of the research question were identified, each was entered separately into PsycInfo to identify related search terms associated with the construct, as defined by PsycInfo’s subject headings. The subject headings were recorded, along with their respective scope notes and year of entry. This process was repeated until no new applicable search terms were identified. In addition to the subject headings, applicable keywords were developed. Keywords are important for a systematic review search for three reasons. First, the application of the subject headings is not automated, and indexers are not always experts in the article’s subject, nor in its methodology (Lefebvre, Manheimer, Glanville, & Group, 2008). As such, human error is introduced in the application of subject headings. Second, keywords are also important to expand the search’s main constructs so that the search is sensitive and yet maintains specificity. A third reason keywords are important is that the most recent articles relevant to a search may not have yet been indexed via the subject headings at the time the search is executed. Thus, the use of keywords assists in identifying these most recent articles. Please see Table 1 for the PsycInfo search terms (information on other searches are available from the first author). The
search strategy was then used as a template for the development of searches in the Medline, CINAHL, and ERIC databases. That is, subject headings and scope notes as dictated by each database’s respective system were identified and the search was individualized for each database. Separate searches for each database were required for two reasons. First, each database uses specific subject headings, thus there is variation in how any one article may be indexed from one database to another. Second, our search strategy was intended to be very broad, so it was necessary to use a wide range both of search terms and of databases in order to ensure that all potentially relevant articles would be identified.

The search of each database was conducted on the same day to maintain consistency. The start of the search was indicated as 1994 to avoid duplication with previous reviews in the area of supervision (Ellis et al., 1996; Stoltenberg, McNeill, & Crethar, 1994), and the search terminated at the end of 2010. To augment the electronic search, key journals’ tables of contents from this time were hand searched. These journals were selected based on a review by Goodyear, Bunch, and Claiborn (2005), which found the majority of supervision-related articles since 2000 were published in Professional Psychology: Research & Practice, Journal of Clinical Psychology, Journal of Consulting Psychology, and Psychotherapy: Theory, Research, Practice, Training. Two additional journals were added because of their strong focus on supervision: Clinical Supervisor and Training and Education in Professional Psychology.

Articles that were internal duplicates (i.e., those that were identified by the same search more than once) were excluded, as were duplicate articles (i.e., those identified by the other searches). The abstracts of the remaining articles were reviewed by the first author to ascertain whether they met the following criteria: (1) article type: the article was a literature review, meta-analysis, original research (qualitative), original research (quantitative), or systematic review; (2) sample: the article’s participants were from counselling or professional psychology, including clinical psychology, counselling psychology, counselling, clinical neuropsychology, or school psychology; (3) focus: the article’s focus was on clinical (vs. research) supervision. Abstracts that did not meet these criteria were excluded. The full text of the articles that met these criteria or were ambiguous with respect to meeting the criteria were obtained for further review in Phase II.

**Phase II: Classification of Articles Regarding Supervision Focus and Research Type**

The codebook used by Painter et al. (2008) for their systematic review on the use of theory in health behaviour research was used as a guide for the current project. The codebook contained the following information: the research question, definitions of the research question’s major constructs, inclusion/exclusion criteria, a guide for coding each article of the systematic review, and definitions of each of the major theoretical models of supervisor development.

In Phase II of the review, the abstracts were reviewed by the first author to determine whether their focus was on supervisor development, which was defined broadly as examining supervisors moving through specific processes (Watkins, 1995), developmental stages, or defined periods of growth (Watkins, 1993) regarding supervision. To be retained at the end of this phase required the study to have a focus on supervisor development and be original research (either quantitative or qualitative).

**Phase III: Rating Extent of Theory Use and Type of Research**

Phase III involved reading the full text of the articles retained in Phase II. At this point, articles could be excluded from the review for various reasons (see Figure 1). The reference lists of the included articles were reviewed to identify any other potentially relevant articles.

The two authors each read this final set of articles, and coded for the following information:

1. Type of research: original research articles as either intervention (randomized controlled trial, quasi-experimental, nonexperimental) or nonintervention research (descriptive/explanatory, methods/measurement research, other);
2. Aspect of supervision development studied, that is, main focus of the study;
3. List of theories used in the article; and

![Figure 1. Study flow diagram.](image-url)
4. Rating on the extent of use of theories of supervisor development. The ratings were informed by the article by Painter and colleagues (2008):

a. No theoretical framework: No theoretical model regarding supervisor development was explicitly identified. Included here are studies (n = 6) that used components of a theory but did not explicitly link them to a theoretical model.

b. Informed by theory: A theoretical model of supervisor development was identified, however there was no or limited/partial application of the model (or constructs from the model) in the study components and measures (e.g., the study only used theory to understand the findings, or in the discussion section only).

c. Applying theory: A theoretical model of supervisor development was specified and the intent of the re-
search was to use theory to understand a sample, another construct, a process, and so forth. The study was applied in that the theory was used to understand an aspect of supervision and the focus of the research was not on increased understanding or development about the theoretical model itself.

d. Testing theory: A theoretical model of supervisor development was specified and the research was designed to determine the validity, scope, or applicability of the theory. The intent of the project was increased understanding about the theory itself.

e. Building or creating theory: The intent of the project was to develop new or revised/expanded theoretical models of supervisor development using the constructs that were specified, measured, and analysed in the research study.

f. Other: Other use of theoretical models not mentioned above.

Also noted for each article was the participants’ mean age and sex, how many participants had doctoral degrees, how many came from clinical or counselling psychology (dichotomized as “minority” or “majority,” i.e., studies with a majority of participants noted as clinical psychology were subsequently categorised as “clinical”), and whether the participants were students, interns or professionals.

To examine interrater reliability, Cohen’s kappa (Cohen, 1960) was calculated for the initial independent ratings of the two reviewers at Phase III. Interrater reliability for the initial coding for research type was 0.66, which according to Cicchetti and Sparrow (1981) and Cicchetti (2001) is “good.” Reliability for the theories used in each article was 0.87, or “excellent.” Finally, the reliability of the rating of the extent of theory use was 0.59, or “fair” (the main source of rater disagreement arose from the use of the categories “informed by theory” and “applying theory”). The two authors reviewed all of these initial coding efforts so that full agreement was met on any of the initial coding that was discrepant. As such, there was 100% agreement on the final coding, and it was these final, agreed-upon ratings that were used for subsequent analyses.

Of the initial 3,256 abstracts reviewed at Phase I, 319 were examined during Phase II. After the review in Phase II, 37 articles remained. As noted above, Phase II of the review excluded papers that did not meet the criteria of being focused on supervisor development, as defined by Watkins (1993, 1995). To ensure that there were no problems of selection bias stemming from using a definition from one of the major theorists in the area (i.e., that no studies were excluded due to using Watkins’ definition), the titles of the articles that were excluded at this stage were reviewed a second time. Based on this review, no additional articles were identified as potentially relevant. We concluded that because no additional articles were included after this second review, the definition of supervisor development that was used for the review was one that is appropriately general and broad. Of the 37 articles retained after Phase II, 25 articles were retained at the end of Phase III and reviewed in full by both authors (see Figure 1). Details regarding the studies’ samples and foci are reported in Table 2.

The references of included studies are marked with an asterisk in the reference list.

Results

The studies varied in the range of content of foci (see Table 2), with the greatest percentage (36%) focusing on the influence of past training and experience in supervision on current supervision practices. Among the 25 studies, 64% were nonintervention research (i.e., not focused on evaluating the impact of supervision training). The studies in this review included a total of 3,825 participants. Female participants made up 52.7% of the total sample, though six studies did not report this information. The mean age of participants across studies was 41.8 years, based on the 13 studies that reported this information. Those with a doctorate (PhD, PsyD, or EdD) made up 30.4% of the sample, based on the 20 studies that reported this information. Participants had counselling and/or counselling psychology training in 16.8% of total participants, based on data from 12 studies. Just under half (43.9%) of participants were training in clinical psychology, based on the 16 studies that reported this information.

The supervisors who were studied came from three categories across the 25 studies that reported this information: 8.2% of total participants were students, 2.9% were interns, and 60.8% were licensed professionals (25 studies). Almost a third of the total number of participants (n = 1,079; 28.2%) did not fit into one of these three categories, had missing data, involved participants from other professional backgrounds, or involved participants who had not acted as a supervisor. In the 13 studies in which the supervisors were students or interns, seven included a consulting professional who provided supervision-of-supervision to these supervisors in training (SIT). Studies rarely included supervisees as participants. Six studies involved master’s students as the supervisees of SIT, though no data were collected from them. None of the studies included interns as supervisees. Two studies included professionals as supervisees, and one of these studies paired these supervisee participants with their supervisors in the same study.

Turning now to the use of theoretical models of supervisor development in the research articles, of the 25 articles that met our inclusion criteria, 48% used a theoretical model of supervisor development to some extent (see Table 3). No one type of study was more likely to explicitly use at least one developmental theoretical model than other types of studies. Of the articles in which a theoretical model was specified, eight articles used or referred to one theoretical model, one article used or referred to two models, and three articles used or referred to three models. Of the 12 articles that used or referred to at least one theoretical model of supervisor development, 11 involved Watkins’ Supervisory Complexity Model (SCM; Watkins, 1990, 1993, 1994), two involved Stoltenberg’s IDM (Stoltenberg & Delworth, 1987; Stoltenberg, McNeill, & Delworth, 1998), three involved theory developed by Hess (1986, 1987), one used Alonso’s psychodynamic theory of supervision (Alonso, 1983), and three used a theory not specific to supervisor development (i.e., Pederson’s Triad Training Model, Pedersen, 1994; Social–cognitive theory, Bandura, 1986, 1989, 1997).

The extent of theory use was rated along a continuum. Of the 25 retained articles, 52% specified no theoretical model of supervisor development, 16% were informed by one or more theoretical
models, 16% applied one or more theoretical models, and 16% tested one or more theoretical models. No studies involved the development of a theoretical model of supervisor development. Of the intervention studies, 22.2% were informed by or applied theory, compared with 37.5% of the nonintervention studies. Of the intervention studies, 11.1% tested theory compared to 18.8% for nonintervention studies.

Discussion

The goals of this systematic review were to assess the extent of use of theoretical models in supervisor development research and to provide descriptive information on the methodological characteristics of studies on this topic. Various findings stemming from our review warrant discussion. We begin by examining the extent...
to which the available research utilizes current theoretical models and covers aspects of supervisor development relevant to professional psychologists and their service activities. Next, we consider the limited research on supervisor development and then move to discuss the nature of the available research and problems with the reporting of participant characteristics in published articles. Finally, implications for future research and for the education of professional psychologists are considered.

One of the most concerning findings from our review was that just over half (52%) of the studies did not appear to use any theoretical models of supervisor development in formulating their research design or questions. This finding echoes that of an early systematic methodological critique of 144 studies in supervision, where a general absence of conceptual rigour was noted and a greater focus on theoretically informed research was advocated (Ellis et al., 1996). Using theoretical models to inform and guide research is a concrete step that researchers can take to improve the overall quality of research. Theoretical models assist researchers in developing clear and concrete predictions. They also help to simultaneously focus and broaden the research efforts. Focus can be achieved by allowing researchers to make specific hypotheses or predictions. Breadth can be achieved when numerous investigators, using different methods, measurements, and samples, all conduct research on elements of the same theory.

Among the articles in which theory was used to some extent, there was an equal distribution of those that were informed by theory, applied, or tested theory (four studies each). Watkins’ SCM (Watkins, 1990, 1993, 1994) was the most commonly used theoretical model, as it was cited in 11 of the 12 studies that identified at least one theoretical model of supervisor development. It is important to note that we did not undertake an evaluation of the quality of the theories (including Watkins’ SCM), or the findings from the studies included in the review, as these assessments went beyond the scope of the current project. That said, it may well be that the frequency of use of the SCM, at least in part, is attributable to the fact that it is the only model of which we are aware that has a measure associated with it: the Psychotherapy Supervisor Development Scale (PSDS; Watkins, Schneider, Haynes, & Nieberding, 1995). The highest rating in our coding system for theory use was “testing theory,” and all four of these studies used the SCM, as did all four of the studies in the second highest category (“applying theory”). Quite simply, having a measure keyed to the theory makes it much easier for researchers to both apply and test the theory.

Based on the articles we reviewed, it appears that Watkins’ SCM is the “only game in town” with respect to the empirical study of supervisor development theories. Watkins deserves credit for establishing a measure with some documented psychometric properties (Barnes & Moon, 2006; Hillman, McPherson, Swank, & Watkins, 1998; Watkins et al., 1995). However, having one theory dominate the limited literature base, although laudable as potential evidence of the clearly established and tested superiority of a particular model, is evidently premature in this area. Echoing Borders (1989) and Ellis (1991), we encourage researchers to direct their efforts at empirically evaluating the assumptions and premises of existing theoretical models of supervisor development. Possible reasons that other models have not been empirically evaluated include perceptions that the existing models are not useful (especially because they are largely descriptive and do not account for how change occurs) and the lack of corresponding measures for most models. In addition, a belief that general experience and training as a clinician is adequate preparation for supervising another clinician has predominated the profession for many years (Bernard & Goodyear, 1998). Such a belief would certainly limit the formulation and evaluation of models of supervisor development.

Perhaps the most striking finding from our review is the dearth of empirical studies focused on supervisor development: only 25 met the inclusion criteria of being empirical investigations of the development of clinical supervisors in professional psychology. In other words, a mean of only 1.6 studies per year were published between 1994 and 2010. It should be noted that this review did not include research on clinical supervision in other disciplines, nor did it include nonempirical articles on supervision. Nonetheless, Green and Dye’s (2002) assertion that the research output in this area is generally “miserly” (p. 108) still seems apt.

Even with our liberal definition of intervention research in the current study, only a third of the articles located in our review examined the impact of supervision training on the development of supervisors. Strikingly, there was only one article in which a randomized controlled trial design was used to evaluate the effects of supervision training. Although not particularly surprising when we consider the dearth of research in the area, it is discouraging. When taken together, the paucity of research, and lack of intervention research in particular, indicate the supervisor development literature currently has an emphasis on descriptive, explanatory research. In essence, the literature on supervisor development in professional psychology is in its empirical infancy.

This poses a significant challenge for psychology educators. Educators are, increasingly, required to train for supervision competency—that is, to help their trainees develop substantial knowledge and skills as supervisors—but there exists negligible evidence regarding this development in the literature. Given this, educators must rely on panel consensus statements, rather than extensive, replicated research findings, to develop training goals and programs. Indeed, Fouad and colleagues used a panel consensus approach when generating competency benchmarks (Fouad et al., 2009) and a corresponding competency assessment toolkit.

Table 3

<table>
<thead>
<tr>
<th>Research Type and Extent of Theory Use of Reviewed Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characteristics of articles</td>
</tr>
<tr>
<td>All articles in final sample</td>
</tr>
<tr>
<td>Intervention research</td>
</tr>
<tr>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>Quasi-experimental</td>
</tr>
<tr>
<td>Non-experimental</td>
</tr>
<tr>
<td>Non-intervention research</td>
</tr>
<tr>
<td>Descriptive/explanatory research</td>
</tr>
<tr>
<td>Methods/measurement research</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Extent of use of Supervisor Development Model</td>
</tr>
<tr>
<td>No theoretical model</td>
</tr>
<tr>
<td>Informed by theory</td>
</tr>
<tr>
<td>Applying theory</td>
</tr>
<tr>
<td>Testing theory</td>
</tr>
<tr>
<td>Building theory</td>
</tr>
</tbody>
</table>

Based on the articles we reviewed, it appears that Watkins’ SCM is the “only game in town” with respect to the empirical study of supervisor development theories. Watkins deserves credit for establishing a measure with some documented psychometric properties (Barnes & Moon, 2006; Hillman, McPherson, Swank, & Watkins, 1998; Watkins et al., 1995). However, having one theory dominate the limited literature base, although laudable as potential evidence of the clearly established and tested superiority of a particular model, is evidently premature in this area. Echoing Borders (1989) and Ellis (1991), we encourage researchers to direct their efforts at empirically evaluating the assumptions and premises of existing theoretical models of supervisor development. Possible reasons that other models have not been empirically evaluated include perceptions that the existing models are not useful (especially because they are largely descriptive and do not account for how change occurs) and the lack of corresponding measures for most models. In addition, a belief that general experience and training as a clinician is adequate preparation for supervising another clinician has predominated the profession for many years (Bernard & Goodyear, 1998). Such a belief would certainly limit the formulation and evaluation of models of supervisor development.

Perhaps the most striking finding from our review is the dearth of empirical studies focused on supervisor development: only 25 met the inclusion criteria of being empirical investigations of the development of clinical supervisors in professional psychology. In other words, a mean of only 1.6 studies per year were published between 1994 and 2010. It should be noted that this review did not include research on clinical supervision in other disciplines, nor did it include nonempirical articles on supervision. Nonetheless, Green and Dye’s (2002) assertion that the research output in this area is generally “miserly” (p. 108) still seems apt.

Even with our liberal definition of intervention research in the current study, only a third of the articles located in our review examined the impact of supervision training on the development of supervisors. Strikingly, there was only one article in which a randomized controlled trial design was used to evaluate the effects of supervision training. Although not particularly surprising when we consider the dearth of research in the area, it is discouraging. When taken together, the paucity of research, and lack of intervention research in particular, indicate the supervisor development literature currently has an emphasis on descriptive, explanatory research. In essence, the literature on supervisor development in professional psychology is in its empirical infancy.

This poses a significant challenge for psychology educators. Educators are, increasingly, required to train for supervision competency—that is, to help their trainees develop substantial knowledge and skills as supervisors—but there exists negligible evidence regarding this development in the literature. Given this, educators must rely on panel consensus statements, rather than extensive, replicated research findings, to develop training goals and programs. Indeed, Fouad and colleagues used a panel consensus approach when generating competency benchmarks (Fouad et al., 2009).
The evidence from the current review indicates that the approach thus far taken by educators such as Fouad and colleagues is indeed appropriate in light of the very limited empirical evidence base and is, therefore, current “best practice.” However, the challenge in any professional competency domain is to develop and implement initiatives based on current best practice, while simultaneously developing the empirical evidence base. It is the extremely limited development of this evidence base over the past 16 years that is concerning. Research linking theoretical models to competency frameworks could provide invaluable guidance, with the models providing an overall “road map” about supervisors’ development and the competency frameworks providing behavioural “road signs” that indicate development is occurring.

Also striking is that the reviewed studies often lacked basic demographic information about participants. For instance, six studies did not report the sex of participants and 12 studies did not report data about participants’ age. Specific information regarding participants’ subdiscipline within professional psychology (e.g., clinical or counselling psychology) was missing in numerous studies. Reporting the demographics of research participants is essential, for without this information it is difficult to accurately interpret research results and meaningfully speculate on the generalizability of the findings. This, in turn, makes it challenging for educators to use the literature to guide their training plans and programs.

**Limitations**

Limitations of the current review include our use of a nominal scale to rate many of the constructs of interest, thereby limiting the power of our statistical analyses. In addition, two of the ratings regarding extent of theory use (i.e., “informed by theory” and “applying theory”) accounted for most disagreement between raters. Although the authors of the reviewed papers may not be explicit enough about how theoretical formulations influenced the design of their research, it also possible that these categories may not optimally map on to the current state of the supervisory development literature. To address this, future systematic reviews in this area may need to take a more “bottom up” approach whereby classification categories are developed based on what is found in the literature, as opposed to our use of a “top down” approach that involved the application of preestablished set of categories to the reviewed literature. It should be noted that even if a study focused on supervisor development as a topic area, if it used a theory that was not specifically focused on supervisor development (e.g., Bandura’s Social Cognitive Theory; Bandura, 1986, 1989, 1997), it was categorised as “other” in our classification scheme. Indeed, as an example, this was the case with Johnson and Stewart’s (2008) paper.

**Implications for Future Research and Education of Psychologists**

Currently there are four main models of supervisor development cited in the literature (Hess, 1986, 1987; Rodenhauser, 1994, 1997; Stoltenberg & Delworth, 1987; Stoltenberg et al., 1998; Watkins, 1990, 1993, 1994). The choice of future research questions, study foci, and educational interventions and their evaluations should be guided by the nature and components of these theoretical models (or newly formulated models), and this connection should be explicitly noted in articles reporting on these initiatives. In this regard, researchers would do well to explicitly consider that both broad theories of competency development (see Johnson & Stewart, 2008) and specific models of supervisor development could be applied to this domain. Measurement research in this literature would, ideally, focus on the development of psychometrically sound tools to assess change that occurs as a function of a supervisor’s development, and the tools would ideally be “keyed” to a corresponding theoretical model. These tools would assist psychology educators in assessing the impact of their supervision training efforts. That said, the question of whether training and/or experience as a supervisor has any perceptible impact on supervisor development remains to be answered, as previous research indicates mixed results (Baker, Exum, & Tyler, 2002; Borders, 1996; Haley, 2002; Johnson & Stewart, 2008; Lyon, Heppler, Leavitt, & Fisher, 2008; McMahon & Simons, 2004; Pelling, 2008; Rodolfa et al., 1998; Stevens, Goodyear, & Robertson, 1998; Vidlak, 2002).

It is noteworthy that all of the supervisor development models focus on the supervision of psychotherapy competencies. Professional psychologists provide a wide array of services, not just psychotherapy, and models of supervisor development that are used to guide training efforts in professional psychology should encompass this broad array of professional activities. Furthermore, as supervisees progress in their training during and after their graduate program, there is likely an increasing emphasis on issues such as ethical decision-making and professional roles and identity. These aspects of professional activity should be included in any model that claims to address the development of supervisory skills and competencies. Supervision researchers and psychology educators would also do well to consider how supervisors’, or supervisor-educators’, own development as a supervisor affects their capacity to facilitate supervisees’ growth in these important professional realms. It is our hope that the above suggestions will aid in professional psychology’s quest to provide evidence-based training aimed at developing trainees’ supervision competency.

**Résumé**

La formation en supervision occupe une place grandissante dans les programmes d’études en psychologie. Pour aider les étudiants à acquérir les compétences nécessaires, il faut connaître les recherches pertinentes ainsi que les méthodes de formation efficaces et, idéalement, ces connaissances doivent provenir d’une littérature empirique réunissant des travaux de qualité s’appuyant sur des bases théoriques. En vue d’évaluer la nature de la littérature empirique consacrée à la formation en supervision, on a procédé à une revue systématique pour déterminer dans quelle mesure les recherches recensées reposaient sur des modèles théoriques. On a extrait des bases de données PsycINFO, Medline, CINAHL et ERIC les articles publiés entre 1994 et 2010. Au départ, 3248 sommaires ont été lus, parmi lesquels 25 satisfaisaient aux critères suivants : la nature empirique de la recherche, la formation en supervision comme sujet et le recours à des participants ayant fait des études supérieures en psychologie à des fins professionnelles ou en psychologie du counseling. Parmi ces articles, la moitié seulement se fondait sur des modèles pour l’apprentissage de la supervision...
pour la conceptualisation, la conception ou l’interprétation de l’étude. Le modèle le plus souvent utilisé est le Supervisor Complexity Model, de Watkins (Watkins, 1990, 1993, 1994). Environ le tiers examinait l’influence d’une formation antérieure et d’une expérience sur les pratiques actuelles de supervision. Pour que progresse la formation en supervision axée sur les compétences, la base théorique doit être grandement élargie et il faudra beaucoup plus de recherches qui se fondent explicitement sur des modèles pour la formation en supervision, ou encore des études qui les évaluent.

Mots-clés : stage en supervision, supervision professionnelle, psychologie clinique, psychologie du counseling, revue systématique.

References

References included in the systematic review are indicated with an asterisk.


*Vidhak, N. W. (2002). Identifying important factors in supervisor development: An examination of supervisor experience, training, and attributes. University of Nebraska, Lincoln, US. ProQuest Dissertations and Theses database. (72743591)


Received October 25, 2011
Revision received June 28, 2012
Accepted July 16, 2012