

The Development of the Clinical Supervisor:
An Examination of Theories, Contributing Factors, and Measures

Keegan Barker

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Clinical Psychology

School of Psychology
Faculty of Social Sciences
University of Ottawa

© Keegan Barker, Ottawa, Canada 2014

Abstract

The development of competency in clinical supervision is becoming an increasingly important element of training and practice in professional psychology. To assist students in developing supervisory competencies, knowledge of relevant research and effective training methods is required. Three studies were designed to add to the field of supervisor development research. In the first study, I conducted a systematic review to examine the extent to which theoretical models of supervisor development have been used in empirical studies of supervisor development within professional psychology. This study revealed that studies rarely employ theory, and when they do, the Supervisor Complexity Model is the most often used theory, likely because it is the only one with an associated measure, the Psychotherapy Supervisor Development Scale (PSDS; Watkins, Schneider, Haynes, & Nieberding, 1995). Building on the findings of this systematic review, in the second study I conducted a reliability and validity generalization on the PSDS. Reliability and validity generalizations are meta-analytic methods that allow reliability and validity data from a measure to be summarized across studies. Results indicate the PSDS has a history of excellent reliability, and higher scores are found with increased supervisory training and experience, evidence of the scale's validity. The third study is a meta-analysis of the supervisor development literature, in order to assess the effect sizes of training and experience on supervisory development and self-efficacy. Results indicate that training and experience have significant positive effect sizes, though they do not differ from one another. Based on results from these studies I provide recommendations about: the samples with which the PSDS tool is appropriately used, and how researchers might increase reliability within their own future studies, and evidence based recommendations for competency based education efforts.

Statement of Co-Authorship

The three manuscripts included in this dissertation were prepared in collaboration with my dissertation supervisor. I was primary author and Dr. John Hunsley was the secondary author for the first manuscript, entitled “The Use of Theoretical Models in Psychology Supervisor Development Research from 1994 to 2010: A Systematic Review”, the second manuscript entitled “Reliability and validity of the Psychotherapy Supervisor Development Scale: A meta-analytic evaluation”, and the third manuscript entitled “The Impact of Supervisory Experience and Training on Supervisory Outcomes: A Meta-Analytic Evaluation”. As the primary author on all manuscripts, I was responsible for the conceptualization of the research question and methods, planning and execution of statistical analyses, and preparation of manuscripts. Dr. Hunsley provided guidance and assistance in all aspects of the project, especially in the refinement of the research methods, queries regarding statistical analyses, and editing of the manuscripts.

Publication Statement

Study 1, The Use of Theoretical Models in Psychology Supervisor Development Research from 1994 to 2010: A Systematic Review was published in *Canadian Psychology*: Barker, K. K., & Hunsley, J. (2013). The Use of Theoretical Models in Psychology Supervisor Development Research from 1994 to 2010: A Systematic Review. *Canadian Psychology*, 54, 176-185.

Study 2, Reliability and validity of the Psychotherapy Supervisor Development Scale: A meta-analytic evaluation, has been accepted for publication (without revisions) in *The Clinical Supervisor*: Barker, K. K., & Hunsley, J. (in press). Reliability and validity of the Psychotherapy Supervisor Development Scale: A meta-analytic evaluation. *The Clinical Supervisor*.

Acknowledgements

I'd like to express my sincere thanks to Dr. John Hunsley for his steadfast support. John—your quiet guidance, measured approach and thorough and timely feedback were reliable touchstones during my studies. Thanks also to my committee members Drs. Gosselin, Kogan and Lee for their time, feedback and encouragement.

I extend my heartfelt thanks to the clinical supervisors who have touched me in their shared passion for client care, and trainee learning and growth. A very special thanks to my first clinical supervisor, Dr. Chuck Marino, who taught about clients, psychotherapy, and about myself. Thank you to Dr. Sarah Bertrim, who provided me with the most challenging, yet safest learning experiences I have ever experienced.

Thank you to my friends in the program Natasha Carraro, Jennifer Hopton, Stephanie Wiebe and Sara Beattie. Thank you to Caley Severn, Kate Spencer, and Claire Kendall for your encouragement and patience with me. Thank you to my parents, Mickey and Steve Barker, for their never ending support, and patience with my ever vague answers to your heartfelt curious question, “are you almost done?”.

To my boys, Caplan and Toby. You came to us during my time in this program, and help me remember—every day—what is important in this life. I cherish and adore you and I am so grateful that I'm your mama. And finally, thank you to Doug. Your curiosity, dedication and patience inspire me. Thank you for being my home, my comfort, and the place I can turn—and return—to.

Table of Contents

Abstract	ii
Statement of Co-Authorship	iii
Publication Statement.....	iv
Acknowledgements.....	v
List of Tables and Figures	xi
TITLE PAGE	1
GENERAL INTRODUCTION: The Development of the Clinical Supervisor: An Examination of Theories, Contributing Factors, and Measures	2
Competencies Movement in Psychology	3
Clinical Supervision: A Core Competency	4
Clinical Supervision Training in Professional Psychology	5
The Limited Research Base on Supervisor Development	8
The Challenge of Educating for Clinical Supervision Competency	10
The Importance of Theory.....	11
Models of Supervisor Development.....	12
Rationale for Study 1.....	14
The Psychotherapy Supervisor Development Scale (PSDS).....	14
Psychometrics of the PSDS	17
Further Analysis of the Factor Structure of the PSDS	19
The PSDS In Use: Studies Employing the Scale	20
Rationale for Study 2.....	21
Supervision Experience versus Training.....	22

Survey Research.....	23
Uncontrolled Intervention Studies	27
Controlled Intervention Studies	29
Randomized Control Group Studies	30
Rationale for Study 3.....	32
Summary of the Studies in the Dissertation	36
STUDY 1: The Use of Theoretical Models in Psychology Supervisor Development Research from 1994 to 2010: A Systematic Review	
Abstract	40
Introduction	41
Method	45
Phase I: Literature Search.....	45
Phase II: Classification of articles regarding supervision focus and research type	47
Phase III: Rating extent of theory use and type of research.....	48
Results	51
Discussion.....	53
Limitations.....	57
Implications for future research and education of psychologists.....	58
Acknowledgements.....	60
References	61
Table 1.....	70
Table 2.....	72
Table 3.....	73

Figure 1.....	75
STUDY 2: Reliability and validity of the Psychotherapy Supervisor Development Scale: A meta-analytic evaluation	76
Abstract	77
Introduction	78
Reliability Generalization and Validity Generalization	80
The Present Study	82
Method	83
Literature Search.....	83
Data Extraction	85
Data Analysis: Reliability Generalization	87
Data Analysis: Validity Generalization.....	88
Results	89
Overview of Studies	89
Reliability Generalization.....	89
Analysis of Moderator Variables	91
Validity Generalization: Experience	92
Analysis of Moderator Variables	93
Validity Generalization: Training	93
Analysis of Moderator Variables	94
Discussion.....	95
Implications and Recommendations	99
References	101

Table 1.....	108
Table 2.....	109
Table 3.....	110
Figure 1.....	111
STUDY 3: The Impact of Supervisor Training and Experience on Supervisory Outcomes: A Meta-Analytic Evaluation	112
Abstract	113
Introduction	114
The Present Study	117
Hypotheses.....	118
Method	119
Literature Search.....	119
Data Extraction	122
Data Analysis	124
Results	126
Overview of Studies	126
Experience Analyses	128
Analysis of Moderator Variables	128
Training Analyses	129
Analysis of Moderator Variables	131
Comparison of Effect Sizes for Training and Experience.....	132
Discussion.....	132
Possible Limitations of the Study	136

Implications and Recommendations	138
References	142
Table 1	151
Table 2	152
Table 3	153
Figure 1	154
GENERAL DISCUSSION	155
Implications	159
Limitations of Reviewed Research	161
Limitations of Current Research	163
Recommendations for Future Research	165
REFERENCES (General Introduction and Discussion)	170
APPENDIX A: Psychotherapy Supervisor Development Scale	183
APPENDIX B: Details of the studies included in Study 3	185

List of Tables and Figures

Study 1

Table 1: PsycInfo Search.....	70
Table 2: Characteristics of study samples and foci.....	72
Table 3: Research type and extent of theory use in reviewed articles	73
Figure 1: Study 1 Flow Diagram	75

Study 2

Table 1: Study Details	108
Table 2: Meta-analysis of internal consistency values of PSDS	109
Table 3: Meta-analysis of validity coefficients for the PSDS	110
Figure 1: Study 2 Flow Diagram	111

Study 3

Table 1: Meta-Analysis of Supervision Experience and Supervisor Development/ Self Efficacy	151
Table 2: Meta-Analysis of Supervision Training and Supervisor Development/Self-Efficacy ..	152
Table 3: Meta-Analysis of Type of Supervision Training and Supervisor Development/Self- Efficacy	153
Figure 1: Study 3 Flow Diagram	154

The Development of the Clinical Supervisor:
An examination of theories, contributing factors, and measures

Keegan Barker
University of Ottawa

The Development of the Clinical Supervisor:

An Examination of Theories, Contributing Factors, and Measures

The development of competency in clinical supervision is increasingly recognized as an important element of training (American Psychological Association, 2009; Canadian Psychological Association, 2002) and practice (Johnson & Stewart, 2000) in professional psychology, and is the focus of my dissertation. In this introduction, I begin by briefly reviewing the competencies-based movement in psychology, with a particular emphasis on clinical supervision. Next, I consider the role of clinical supervision training in graduate programs in professional psychology. I then highlight the discrepancy between efforts to understand the development of the supervisee, compared to efforts to understand the supervisor, as evidenced in the variety of theoretical models of development issued for each respective group. The important role that theory plays in research is then discussed, and this is followed by a brief review of the four main theoretical models of supervisor development. The rationale for Study 1 of the dissertation is then presented. One model of supervisor development (and its corresponding measure) is the focus of most of the empirical efforts in the area, and this scale's development, its psychometrics, and studies employing it are reviewed, in order to set the stage for the rationale for Study 2 of this dissertation. Finally, literature on the impact of training in supervision, compared to supervision experience on supervisor outcomes is considered. This is followed by the rationale for Study 3. The final section of the introduction is a brief summary of the three studies that comprise this dissertation.

Competencies Movement in Psychology

Very broadly, competency-based education advocates for educational and training programs to require the explicit demonstration of competence (Falender & Shafranske, 2012) in trainees, rather than assuming that competence is achieved simply as a consequence of undertaking doctoral education and clinical training. Competency-based education is also a movement in other disciplines, including medicine and dentistry (Frank et al., 2010). In psychology, the shift toward competency-based education (CBE) emerged from three related areas: (1) a duty to protect the public by ensuring qualifications are met for license to practice, (2) a requirement on the part of training programs to meet explicit recommendations for accreditation; (3) a movement toward the scientific tracking of training outcomes (Nelson, 2007).

In 2004, the Association of Psychology Postdoctoral and Internship Centers (APPIC) initiated the Competencies Conference: Future Direction in Education and Credentialing in Professional Psychology (Kaslow et al., 2004). Contributors included the American Psychological Association (APA), and various education and training groups. In 2006, the Benchmarks Conference (American Psychological Association, 2006) was convened by the Council of Chairs of Training Councils and the APA Board of Educational Affairs. The end result was a document listing 15 competency areas (with definitions and behavioural anchors for each), which was designed to identify the minimum level of competence expected for various levels: readiness for practicum, readiness for internship, readiness for entry to practice (Falender & Shafranske, 2012). In 2009, Fouad et al. (2009) built upon the Benchmarks by considering the implications for graduate training, including individualizing and tailoring educational and training

initiatives to meet the needs of students via ongoing monitoring and feedback on the behavioural markers of competency areas.

Simply put, competency-based education in clinical supervision emphasises the ability of trainees to translate their acquired knowledge and skills into real world application, and to evaluate trainees using performance outcomes (Falender & Shafranske, 2004). Efforts to refine these competency models continue. For example, the Benchmarks have recently been revised to create a more flexible and condensed rating (Hatcher et al., 2013). The Association of State and Provincial Psychology Boards noted that many models to date applied only to academic training and thus they recently generated a competency model to include the sequence of training leading to licensure and autonomous practice (Rodolfa et al., 2013).

Clinical Supervision: A Core Competency

Across numerous countries, clinical supervision is regarded as a competency for professional psychologists (i.e., counselling, clinical, health, school, clinical neuropsychology). In Canada, the Mutual Recognition Agreement (see <http://www.cpa.ca/docs/file/MRA.pdf>) of the provincial and territorial regulatory bodies requires competency in supervision in some jurisdictions. In the United States, APA has stated that clinical supervision should be regarded as a distinct professional competency (American Psychological Association, 2006). Finally, in the United Kingdom the government-led “Improving Access to Psychological Therapies” initiative included an emphasis on the development and specification of supervision competencies (Milne et al., 2010).

Clinical supervision is defined here based on Bernard and Goodyear's widely cited work (2004), perhaps most succinctly presented by Milne and colleagues (2007):

The formal provision, (i.e., sanctioned by relevant organization/s) by senior/qualified health practitioners (or similarly experienced staff), of an intensive education (general problem solving capacity; developing capability) and/or training (competence enhancement) that is case-focused and which supports, directs and guides (including restorative and/or normative topics, addressed by means of professional methods, including objective monitoring, feedback and evaluation) the work of junior colleagues (supervisees) (Milne, 2007, p. 440)

The above definition is the most comprehensive and detailed presented thus far in the literature, as Milne and colleagues (2010) modified Bernard and Goodyear's widely used definition (Bernard & Goodyear, 2004) based on findings from their systematic literature review. There are two interrelated but distinct purposes of supervision: to ensure the integrity of clinical services and to develop service provision competence in the supervisee (Falender & Shafranske, 2004).

Clinical Supervision Training in Professional Psychology

The Competencies Conference (Kaslow et al., 2004) supervision workgroup proposed a supervision competencies framework in six key areas of supervision: knowledge, skills, values, social context, training of supervision, assessment of supervision competencies (Falender et al., 2004). The 2006 Benchmarks Conference (APA, 2006) concluded that the competency of clinical supervision requires focused development via education and clinical training. Following on these developments, the

APA's Commission on Accreditation (APA, 2009) concluded that all students should acquire and demonstrate substantial understanding of, and competency in, supervision. A similar position was taken over a decade ago by the Canadian Psychological Association (CPA), as CPA accreditation standards for professional psychology programs and internships require that graduate students receive training in clinical supervision (CPA, 2002, 2011).

Despite clear recommendations from the national psychology associations in both countries, as well as broad recommendations for educational implications (Fouad et al., 2009), specific training recommendations are lacking, as is guidance about how graduate programs and internships should share responsibility for such training (Hadjistavropoulos, Kehler, & Hadjistavropoulos, 2010). Despite this, there is evidence that some students in professional psychology receive training in how to provide clinical supervision. Haley (2002) noted that 66.2% of clinical and counselling doctoral interns at APA-accredited university counselling centre internship sites reported having taken a course in supervision, with 11.7% taking a didactic course (with an average of 45 classroom hours), and 54.5% took a didactic-experiential (with an average of 62 hours, including the supervision of, on average, 2.7 trainees). Lyon, Heppler, Leavitt and Fisher (2008) found that 39% of surveyed predoctoral interns from accredited intership sites had taken a graduate course in supervision. Of course, students cannot take a graduate course in supervision unless their academic institution or internship site offers one. A study of 332 APA-accredited academic program directors and training directors of APA-accredited intership programs found that almost one-quarter of academic programs and 34% of intership programs did not offer a didactic course or seminar in supervision

(Scott, Ingram, Vitanza, & Smith, 2000). Additionally, a survey of directors of CPA-accredited clinical training programs in clinical and counselling psychology found that although 65% reported offering course work in clinical supervision, the average time devoted to training in clinical supervision was relatively limited, ranging from 3 to 39 hours (Hadjistavropoulos et al., 2010). It thus appears that there is a wide range of course-based training experiences across accredited programs.

With respect to opportunities to engage in hands-on (i.e. practicum training) in the provision of supervision, although most students have access to such training, 20% of APA-accredited doctoral academic programs and 35% of internships did not offer this training opportunity to trainees (Scott et al., 2000). In Canada, 35% of the surveyed directors of doctoral training programs noted that students in their programs are not able to complete a practicum in supervision and practicum training was required in only five of twenty programs surveyed (Hadjistavropoulos et al., 2010). In addition, the range of hours in such a practicum was large, falling between five and 50 hours of experience providing clinical supervision.

At first glance, the data are promising in that it appears the majority of internship and doctoral programs offer supervision training to their trainees. Further, the data on supervision practica appears to have shifted since 1994, when just 4% of 559 American graduate programs in psychology required doctoral students to supervise more junior trainees (APA, 1994). There is clear recognition of the importance of training students in clinical supervision with programs responding by providing course based and/or practical training. The variability of training experiences is wide, with as little as three hours of didactic training or five hours of practicum training in academic programs

(Hadjistavropoulos et al., 2010) to as much as 62 hours of didactic-experiential graduate course work (Haley, 2002). As an emerging area of competency, training students in how to provide clinical supervision is challenging. Indeed, 40% of Canadian directors of doctoral training programs noted that a significant challenge to providing supervision training to students was the difficulty in finding faculty or supervisors appropriately trained on how to teach supervisory skills (Hadjistavropoulos et al., 2010); 63% of supervisors noted they had received no training in clinical supervision in graduate school (Johnson & Stewart, 2008).

The Limited Research Base on Supervisor Development

A further challenge in providing training in clinical supervision is the dearth of empirical literature to guide such training efforts—that is, there is a lack of research on the ways clinical supervisors progress from novice to expert supervisor. Without an understanding of the changes supervisors in training (SIT) go through as they develop competency, it is challenging to generate appropriate educational interventions to guide their development. Although clinical supervision has been a focus of study in psychology for some time (Borders, 1989; Gandolfo & Brown, 1987; A. K. Hess, 1986, 1987; K. A. Hess & Hess, 1983; Holloway, 1987; Leddick & Dye, 1987; Milne, 1989; Patton, 1987; Schwartz, 1989; Stoltenberg, 1981; Stoltenberg & Delworth, 1987; Worthington, 1987), Bernard and Goodyear (1998) noted the literature tends to emphasize both the overall process of supervision and the development and growth of supervisees. There is considerably less emphasis on models that address the development of the clinical supervisor with respect to knowledge and skills.

As early as the mid 1980s, 12 models related to supervisees' development and/or the supervision process in general had been formulated (Worthington, 1984) and, within a few years, this number had doubled (Chagnon & Russell, 1995). Falender and Shafranske (2004) characterized these broad supervision models as falling into two categories. First there are the psychotherapy-based approaches. Here, the orientation of the supervisor informs what is discussed in supervision, with the advantage that modelling of therapeutic processes can occur in supervision. For example, a cognitive-behavioural model of supervision uses processes such as collaboration and guided discovery (Padesky, 1996), and Adlerian model includes a focus on addressing countertransference with counselors in training (Tobin & McCurdy, 2006). The second category of supervision models is developmental, examining the progression of a supervisee from a novice therapist to a seasoned or master therapist. These models tend to be meta-theoretical, sequential, and progress to a fixed endpoint (Falender & Shafranske, 2004). There is a third category of models that are process-based, including Bernard and Goodyear's Discrimination Model (Bernard & Goodyear, 1992, 2014). In contrast to the various models focused on supervisee development, there are far fewer models, and research, focused on the development of supervisors (Bernard & Goodyear, 1998; Watkins, 1995a). Following a systematic review of 144 empirical studies of supervision between 1981-1993, Ellis, Ladany, Krenzel and Schult (1996) found a lack of conceptual and methodological rigour. Furthermore, Ellis and Ladany (1997) concluded that the quality of the research was substandard, and Watkins (2012) echoed this conclusion. Milne et al. (2010) noted that dearth of quality research that has resulted in the inattention paid to the training of clinical supervisors.

The Challenge of Educating for Clinical Supervision Competency

Clinical supervision plays a major role in the training of professional psychologists. For instance, in the 2012 survey of applicants to the APPIC predoctoral internship match, applicants reported receiving a median of 312 clinical supervision hours in their doctoral training (Association of Psychology Postdoctoral and Internship Centers, 2012). Accrediting bodies have recognized the importance of supervision (APA, 2009; CPA, 2002; 2011), Furthermore, licensing requirements identify supervision as an important competency for psychologists in the United States (APA, 2006), Canada (see <http://www.cpa.ca/docs/file/MRA.pdf>), and the United Kingdom (Milne et al., 2010). Providing supervision also appears to play a large role in supervising psychologists' work lives, with an average of six to seven hours of week spent on supervision-related activities (Johnson & Stewart, 2000). That is, supervisors spent almost a full day a week on supervision related activities, and yet their training in the area is varied and inconsistent, with almost a tenth of supervisors reporting no formal training or self- study in supervision (Pelling, 2008), and three-quarters having had no didactic class or practicum in supervision (Vidlak, 2002).

As the provision of supervision is likely to be a major component of future clinical practice for many psychologists, it would be ideal if graduate programs would provide evidence-based supervision training to these future supervisors. The challenge facing educators is how to best provide this training. To be consistent with other areas of professional training, supervision training should be based on the empirical literature on supervisor development and supervisor training. Ideally, evidence from good quality research should inform training efforts.

The Importance of Theory

A necessary first step in the generation of quality research is the extent that theory is used in the conceptualization and conduct of the research. A cumulative and coherent evidence base requires that theory influences how evidence is collected, analyzed, and understood (Alderson, 1998; Green, 2000). Using theories and models to inform and guide research is a concrete step that researchers can take in order to improve the overall quality of research in an area. Theories assist researchers in developing clear and concrete predictions. They also help to simultaneously focus and broaden the research efforts. Focus is 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 theoretical model, and feed those results back to further inform the development of the model.

As previously noted, models that are focused on the development of clinical supervisors have been the focus of much less inquiry than models focused on the development of supervisees (Bernard & Goodyear, 1998; Watkins, 1995a). The developmental models for supervisors share many of the same characteristics of the supervisee developmental models in that they are meta-theoretical, sequential, and progress to a fixed endpoint (Falender & Shafranske, 2004). Consistent with Milne's (2009) position that general theories are the basis for particular models, these supervisor development models are based on general lifespan development theories (Baltes, 1983; Rest, 1979). For example, the supervisor development models include key concepts from lifespan development theory such as stages, phases, and tasks. Lifespan development theories encourage attention to these stages, phases and tasks such that, if the tasks of a

particular phase are addressed, individuals progress to the next phase in their continuing growth (Baltes, 1983; Rest, 1979).

Models of Supervisor Development

There are four main models of supervisor development: Hess (A. K. Hess, 1986, 1987); Stoltenberg and Delworth's Integrated Development Model (Stoltenberg & Delworth, 1987; Stoltenberg, McNeill, & Delworth, 1998) which is a model of supervisee development, but also specifies supervisor development; Watkins' Supervisor Complexity Model (Watkins, 1990, 1993, 1994); and Rodenhauser's Multidimensional Model (Rodenhauser, 1994, 1997).

There are considerable similarities (Russell & Petrie, 1994; Watkins, 1995b; K. E. White, 1998) across the four models of supervisor development. Specifically, all the models state that supervisors begin in a state of uncertainty, insecurity, anxiety, and inexperience. At each stage supervisors grapple with a key professional issue, and progression is a product of time, experience and struggle. Ultimately, each model posits that, over time, the supervisor forms an identity as a supervisor and, attainment of the final stage reflects competence in the role of supervisor. They are thus maturational (Baltes, 1983) and stage (Rest, 1979) models with distinct, hierarchical, and sequential stages through which a supervisor moves through qualitatively different ways of thinking.

White (1998) noted that the four models all recognize the importance of: (1) the interpersonal process issues in the supervisory relationship; (2) the needs of the supervisee; (3) a flexible supervisor response to these needs (e.g., at times the supervisor acts as a didactic teacher with the supervisee and at other times the supervisor acts as a consultant); (4) an optimal level of structure and use of power in supervision; (5) accurate

supervisor self perception regarding of his/her own competence; (6) professional identification with the role of supervisor; (7) knowledge and application of theories and techniques of supervision; and (8) an awareness of and appropriate response to one's own anxiety in order to prevent defensive reactions or inappropriate use of structure. Thus the models all posit that a highly developed supervisor is confident, has a high level of skills, and is flexible in their use of structure and power (K. E. White, 1998).

Unfortunately, there are also some similarities across models in that stage theories do not specify how a supervisor progresses from one phase/stage to another (K. E. White, 1998; Worthington, 1987). Nor do the models specify what factors might influence how quickly or how well a supervisor develops. Watkins (1995b) proposed that progress occurred by way of time, experience, training and struggle, and Hess (A. K. Hess, 1986, 1987) speculated that supervisor evaluation, reading the literature on supervision, and personal factors (e.g., if the supervisor needs a lot of admiration) can affect development. Stoltenberg and Delworth (1987) suggested that level development as a clinician, supervisory training and experiences impact on supervisory development. That said, the models do not clearly articulate, for example, how much experience, training, or reading is required to progress from one stage to another. The models also do not articulate whether transitions at the initial stages of the model require the same amount of effort as those at the final stages of the model.

The empirical base for these four models of supervisor development is generally weak. One of Milne's (2009) criteria for a sound supervision model is that it is tested and supported by empirical evidence. Ellis and colleagues (Ellis et al., 1996) reviewed 144 supervision studies and evaluated them on numerous threats to validity, including

inferences about the fit of the research hypotheses with theory. They found that only 20% of the studies involved explicit tests of relevant theories, and they exhorted researchers to ensure that they were explicitly using relevant theories to address important, unanswered questions about supervision.

Rationale for Study 1

Research on supervisor development is already limited (Bernard & Goodyear, 1998; Watkins, 1995a, 2012). Having this limited research potentially spread across four models—which are essentially the same—is a lost opportunity for concentrated efforts in either developing and refining one inclusive model of supervisor development or comparing conceptually distinct models. One way to assess which model may be most useful to future research is to develop an understanding about the extent to which each model has been used in research investigations. In the first study of this dissertation, I conducted a systematic review to assess the extent to which theoretical models of supervisor development have been employed in supervisor development research in professional psychology. That is, the extent to which the research in supervisor development was informed by theoretical models, applied theoretical models, tested theoretical models or built theoretical models was assessed using a systematic review approach.

The Psychotherapy Supervisor Development Scale

In Study 1 of this dissertation, I found that The Psychotherapy Supervisor Development Scale (PSDS; C. E. Watkins, Jr., L. J. Schneider, J. Haynes, & R. Nieberding, 1995), a measure associated with Watkins' SCM (Watkins, 1990, 1993, 1994) is the only measure designed to assess supervisor development. As such, it is

important to understand its reliability and validity. To this end, the development of the PSDS, as well as its psychometrics, will be reviewed below in order to set the stage for the rationale for Study 2—a meta-analysis that summarized the reliability and validity data on the PSDS.

The PSDS was developed as a self-report assessment of a supervisor's perceptions of his or her development as a supervisor (C. E. Watkins, Jr. et al., 1995). Specifically, the PSDS assesses the key polarities outlined in the SCM, that is: competency versus incompetency, self-awareness versus unawareness, identity versus identity diffusion, and autonomy versus dependence (Watkins, 1993). These polarities were selected for scale development because they cut across all four stages of the SCM, they reflected core developmental concerns, and, according to SCM theory, they varied by stage (C. E. Watkins, Jr. et al., 1995). The authors generated 75 items, using a 7-point rating scale on which respondents are asked to indicate frequency with anchors ranging from 1 = never, to 4 = half of the time, and 7 = always. The item scores are summed to produce a total score. Higher scores reflect higher development.

These 75 items were assessed by a panel of five subject matter experts (licensed psychologists who supervised doctoral students providing therapeutic services). Each item was rated on a five-point scale regarding how well it assessed the issue of concern. From this initial sample of 75 items, 46 items obtained an average rating of 3.5 or higher, and thus were kept as potential scale items in the next version of the measure. These 46 items were sent as a survey to 335 members of the APA's Division of Psychotherapy, all of whom were providing psychotherapy supervision at the time of the survey. Each

participant was asked to complete the measure regarding his/her own level of development as a supervisor, using the 7-point scale.

The data from this survey were subjected to a principal-components factor analysis (PCA) with varimax rotation (it should be noted that although a PCA was reported, it appears that a principal factor analysis should have in fact been conducted, as the goal was to identify underlying factors in the data. Further, because the researchers had a model that they wished to test, a confirmatory factor analysis may have been more appropriate). After conducting PCA, nine factors emerged with eigenvalues greater than one, but when authors examined the scree tests and considered interpretability and parsimony, the authors concluded a four-factor solution provided the best fit to the data. This solution accounted for 43.5% of the variance. Eighteen of the items loaded at the level of .45 or higher on these factors and so were retained (C. E. Watkins, Jr. et al., 1995).

The first factor's eigenvalue was 13.87, and was composed of five items, all of which came from the initial competency/incompetency polarity; this factor was termed the Competence/Effectiveness factor. The second factor's eigenvalue was 2.42, and was composed of eight items, six of which came from the initial items to assess the self-awareness/unawareness polarity; the factor was named Identity/Commitment factor. The third factor's eigenvalue was 1.95, and was composed of four items, three of which came from the initial self-awareness/unawareness polarity items; this factor was termed the Self-Awareness factor. Finally, the fourth factor's eigenvalue was 1.66, made up of four items, three coming from the original identity/identity diffusion polarity items, and one from competency/incompetency; this factor was termed Sincerity in the Supervisor Role.

The authors noted that the 18 items clustered into factors that appeared to reflect key polarities as outlined in the SCM. However, because of the low number of items loading to the factors, they argued against using these four separate factors as subscales. As such, they recommended using the scale as a general assessment of development. A concern with this recommendation, however, is that the scale is composed of four factors with differing number of items, ranging from four to eight. This then means that when a total score is used, the factor with eight items (i.e., Identity/Commitment) is weighted more heavily than are the other facets. This presents a problem because this does not fit with the SCM, in that the SCM does not weight the relative importance of the model's polarities.

Psychometrics of the PSDS

In the survey of APA Division of Psychotherapy members, most participants tended to rate themselves in the upper end of the scale. That is, the mean score for the 18 items was 108.8 ($SD = 8.3$) out of a possible 136. The alpha coefficient was .90 for the total score. Using data from the sample in an analysis of variance, the researchers found a significant main effect for supervisor experience. Comparisons were made on PSDS scores between those with lower experience (less than eight years experience), moderate experience (eight through 24 years of experience), and high experience (more than 25 years of experience). Post hoc analyses demonstrated that the low experience group scored significantly lower than did the moderate and high experience groups, and that the moderate experience group scored significantly lower than did the high experience group. Thus, the authors concluded that the scale demonstrated construct validity, in that it was able to discriminate across broadly defined supervisor experience levels. These categories

were based on standard deviation (SD) from the mean of 16.1 years experience: low-experience had less than eight years (one SD below the mean), moderate-experience had eight through 24 years (plus and minus one standard deviation from mean), and high-experience 25 or more years (more than one SD above the mean). This sample had a very high mean average level of experience, and as such this may account for the finding that experience was related to development. The 18-item PSDS scale can be found in Appendix A.

An additional study was conducted by the scale's developer, with the goal to evaluate internal consistency, temporal stability (test-retest reliability), and concurrent validity values (Hillman, McPherson, Swank, & Watkins, 1998). Participants included 36 clinical supervisors at university counselling centres and seven supervisors in training who were enrolled in a clinical supervision course. The majority of the participants had doctoral degrees (74.7%), mainly in psychology or social work. Licensed psychologists made up just under half of the sample (48.8%). The participants completed the PSDS twice, four weeks apart. The resulting stability correlation was $r = .85$ ($p < .01$). Internal consistency using a split-half procedure was very high for this sample ($r = .95$). Participants were asked to describe their theory of supervision, which was used as a proxy indicator to assess the criterion validity of the PSDS because no other scale or criterion of supervisor development exists. Participants' description of their theory of supervision was then rated by three experts in the area of supervision, in terms of the complexity of the descriptions. These ratings correlated $r = .40$ ($p < .01$) with the PSDS, suggesting moderate evidence of concurrent validity.

Further Analysis of the Factor Structure of the PSDS

An independent set of researchers examined the factor structure of the PSDS eleven years after the scale's initial development (Barnes & Moon, 2006). The goal of this study was to test, using confirmatory factor analysis, the factor structure proposed by Watkins and colleagues (1995) on a new sample of clinical supervisors. The sample consisted of 255 supervisors affiliated with counsellor training programs. More specifically, the study tested two possible factor structures. The first model was a second-order factor structure, used to validate Watkins et al.'s (1995) conclusion that the PSDS is best regarded as a measure of overall development (i.e., single factor conceptualization with four factors contributing to the total factor-structure variance). The second model that was tested was a four-factor model, that is, the original factors identified by Watkins et al. (1995).

The models were tested using maximum likelihood estimation. The authors noted that the four-factor model demonstrated results nearly identical to the second order model. That said, a chi-square test determined that the four-factor model was a better fit, with $p = 0.59$. Although the difference between models was not statistically significant at .05 level, the authors noted that the four-factor model was the more parsimonious model and thus it was retained. Two items were highly correlated with one another, and because their content appeared to overlap substantially, the errors of those items were allowed to correlate in the modified structural equation model. This modified four-factor model produced the best fit, with the following factors: Competence/Effectiveness, Identity/Commitment, Self-Awareness, Sincerity in Supervisor Role. It should be noted that these differ slightly from the four

polarities originally identified by Watkins in the SCM (C. E. Watkins, Jr. et al., 1995): Competency/incompetency, Identity/Identify confusion, Self awareness/Awareness, Autonomy/dependency, but that they do indeed match the factors identified by Watkins in the initial PCA (C. E. Watkins, Jr. et al., 1995). Barnes and Moon (2006) thus concluded that the PSDS is best conceptualized as a four-factor model, in contrast to Watkins' recommendation of a second-order factor. They concluded that supervisor development, as measured by the PSDS, is indeed a multidimensional construct.

The PSDS in Use: Studies Employing the Scale

As noted above, the PSDS is the only measure of supervisor development and has thus been used in several studies of supervisor development. These findings will be reviewed briefly here. Most of the research using the PSDS has employed survey research designs. Bencivenne (1999) surveyed marriage and family therapy supervisors ($n = 136$) and their supervisees ($n = 116$), and found that supervisors with over 60 hours of training in supervision had higher PSDS scores than did supervisors with fewer hours of training. Vidlak (2002) surveyed 99 psychologists supervising at APA accredited internship sites and found that training, but not experience, was positively related to PSDS scores. Lyon and colleagues (2008) surveyed 233 predoctoral interns regarding their training and experience in supervision, and their self-perceived supervisor development and noted that number of training activities and supervision of supervision were positively related to PSDS scores. Pelling's (2008) surveyed 175 counsellors who were supervising, and found that experience and training were both predictors of PSDS. Vieceli (2007) surveyed 37 supervisors in academic and community settings and, although she found that a positive correlation between emphasis on conceptualization in

supervision and greater overall development levels of the supervisors, there were no significant associations between PSDS scores and the variables of degree type, supervisory experience, formal training in supervision, or workplace support for supervision.

One prospective study that provided a supervision training intervention to participants. Baker and colleagues' (Baker, Exum, & Tyler, 2002) prospectively studied of the impact of a training program on 12 doctoral students enrolled in a supervision practicum (intervention), and compared their PSDS scores to those not yet enrolled ($n = 7$, control). The intervention participants' PSDS scores increased significantly across the semester, and their scores were significantly greater than control participants' scores.

Rationale for Study 2

In Ellis and colleague's (1996), review of supervision research, they noted the absence of reliable and valid instruments, and suggested that it was this lack of attention to measurement validity and reliability that accounted for some of the gaps in the knowledge base at that time. Over the past decade and a half, the PSDS is the only measure to have been developed that assesses supervisor development. Yet, as Watkins (2012) himself noted, the PSDS requires more study. Summarizing the psychometric properties of the PSDS would either substantiate the measure's foothold in supervisor development research, or help guide future directions in the use of the PSDS and provide recommendations for future measurement research. Either way, such a summary would provide evidence regarding the use of the PSDS, and aid researchers in making evidence based decisions on the use of the scale, rather than it being employed simply because it is the only measure of its kind.

In light of the PSDS' use in 10 studies (including studies the three studies focused on the psychometric properties of the measure), the second study in this dissertation was a meta-analysis of the reliability and validity data from the PSDS. Even though there are a limited number of studies on which to base the meta-analysis, previous meta-analytic work suggests that there is benefit to conducting meta-analyses early in development of research in an area (Fergusson, Glass, Hutton, & Shapiro, 2005). In addition, meta-analyses are warranted when a general overall picture of the evidence in a topic area is needed to direct future research efforts (Petticrew & Roberts, 2006). This is clearly the case in supervision research, as this area of research is growing in light of competency-based education efforts. Details of the methods employed in meta-analyses are reviewed in the methods for study two.

Supervision Experience versus Training

Bernard and Goodyear (1998) noted there has been considerable interest in the development of theoretical models to explain supervisees' development—in contrast with remarkably less interest in generating models to explain the development of supervisors. They concluded that it has been largely believed that experience as a counsellor/therapist and as a supervisee is sufficient to become a supervisor. As a consequence of the paucity of supervisor development models, the corresponding research is limited (Watkins, 1995a), with only the Watkins' Supervisor Complexity Model (Watkins, 1990, 1993, 1994) having received empirical attention (Watkins, 2012).

The belief that experience as a supervisee and as a therapist is sufficient to be a supervisor does appear to be shifting, as reflected in the current emphasis on training requirements in supervision (APA, 2006; CPA, 2002), along with a competency-based

approach to this training (Falender & Shafranske, 2004, 2012). The relative impact of experience versus training on supervisors is still debated, and this is the focus of a substantial body of research in the area of supervisor development, which is reviewed below.

Survey Research

Survey research with small samples of graduate students (Caruso, Dooley, & Dabit, 1995), predoctoral interns (Haley, 2002; Lyon et al., 2008), and professionals (Johnson & Stewart, 2008; Pelling, 2008; Rodolfa et al., 1998; Stevens, Goodyear, & Robertson, 1998; Vidlak, 2002; Williams, 2012) has yielded mixed findings regarding the association between experience and supervisory outcomes (including supervisor development). Years of supervision experience were related to higher supervisor development in one study (Williams, 2012), although other research has found this link for self-efficacy only when those with five or more years were compared to those with zero to two years experience (Stevens et al., 1998). Perhaps the greatest contrast is between Pelling's (2008) research in which experience was the single best predictor of supervisor development, explaining 8.2% of the variability in PSDS scores, and other studies that have found no link between supervisory experience and supervisory development (Lyon et al., 2008; Vidlak, 2002), supervisory working alliance (Rodolfa et al., 1998), or supervisor stance during supervision (Stevens et al., 1998). Of note, Rodolfa and colleagues (1998) did not find differences in working alliance based on number of years of experience, but surprisingly they did find that the more and less experienced groups did not differ in the actual number of trainees supervised. Thus, perhaps it is not number of years that is a key consideration in determining the connection between

experience and development, but rather the amount of experience in those years (i.e., number of supervisees one has supervised).

When investigated in surveys, compared to the situation with supervisory experience, supervisory training has been found to be more consistently linked to supervisor outcomes such as self-efficacy (Johnson & Stewart, 2008; Stevens et al., 1998) and supervisor development (Lyon et al., 2008; Vidlak, 2002). Supervisory training was found in one study to account for 4.4% of the variance in PSDS scores (Pelling, 2008). More specifically, it appears that the number of training experiences is positively associated with supervisor development (Lyon et al., 2008; Vidlak, 2002) and self-efficacy (Stevens et al., 1998). Those with higher number of training hours had higher development scores (Lyon et al., 2008), as did those with three training experiences compared to one or no training experiences (Vidlak, 2002). Those with three or more training experiences had higher supervisor self-efficacy than those with no experiences (Stevens et al., 1998). Supervisors with more training felt more prepared to supervise upon graduation (Johnson & Stewart, 2008).

Sustained training appears to be related to supervisor development, in that training taken over a period of time was linked with higher development, whereas concentrated workshop training was not (Vidlak, 2002). Two studies (Lyon et al., 2008; Williams, 2012) found that workshop training was related to development, though it had the lowest correlation with development of the various training methods (Lyon et al., 2008). Format of the supervisory training may also to play a role, in that training that includes a practicum (Vidlak, 2002), combined didactic-practicum format (Haley, 2002; Vidlak, 2002), and supervision of supervision (Lyon et al., 2008; Vidlak, 2002) were linked with

higher supervisor development scores (Lyon et al., 2008; Vidlak, 2002) and higher overall supervisory self-efficacy compared to intern supervisors with no supervision training (Haley, 2002). Only one study (Williams, 2012) failed to find significant correlations for supervisor development and supervision of supervision (SOS), number of hours in a didactic course, and number of hours in a supervision practicum. Finally, weekly SOS was listed by 62% of doctoral interns as the most effective supervisory training activity in a supervision training course (Caruso et al., 1995).

In general, survey research is limited methodologically because it does not allow conclusions of causality to be drawn. Other limits are apparent in the eight studies reviewed above. First, they are all retrospective inquiries, and as such are susceptible to recall bias on the part of participants. Second, most of the studies had limited sample sizes, with a range of $n = 27$ (Caruso et al., 1995) to $n = 298$ supervisors (Williams, 2012). Sample size can impact the study's power and thus its ability to detect significant patterns in the data.

A third limit of the survey research reviewed here is that many of the studies relied entirely on self-report tools, which are susceptible to biases in respondents' self-assessments. Meta-analytic research has found that physicians have a limited ability to accurately self-assess (Davis et al., 2006), and it is possible that this phenomenon applies to other health professionals. It is also possible that those who have participated in training experiences may be strongly invested in the view that this training was of use to them, and therefore report higher development and/or self-efficacy levels than do those with no training. The self-assessments in the current research were not linked to any

objective data (e.g., a behavioural observation of competency in supervision, ratings from supervisees), and as such limited in their contribution to our understanding of supervisor development.

A fourth limitation of these survey data is the nature of the measures used. For instance, one study used a tool that was adapted from other measures and reported only test-retest reliability of the adapted tool (Haley, 2002), others used a measure developed specifically for the survey, with either no (Caruso et al., 1995) or limited evidence of psychometric adequacy (Johnson & Stewart, 2008), and one relied on a single item measure (Stevens et al., 1998), which limits the reliability and validity of the associated construct. In sum, measurement problems evident in these studies weaken the conclusions that can be drawn from this research.

A fifth limitation is the varied nature with which experience and training were operationalized. For instance, experience was noted as time since licensure (Rodolfa et al., 1998), number of years supervising (Haley, 2002; Stevens et al., 1998) both independently and with SOS (Pelling, 2008), or number of hours providing supervision (Lyon et al., 2008). Training was operationalized as number of formal instruction activities such as courses or workshop (Stevens et al., 1998) or total number of training activities (Lyon et al., 2008), but training was also further delineated in other studies (Pelling, 2008; Vidlak, 2002) as including both formal and informal (e.g., self-study, teaching a class) efforts, or were classified as didactic, experiential, or combination training (Haley, 2002). Training was also assessed via a 7-point scale regarding its adequacy (Johnson & Stewart, 2008), which is open to different interpretations because participants may not have the same criteria for evaluating the same training experience

(i.e., for some, an eight-hour training workshop may feel “adequate” whereas for others the same workshop may be “inadequate”).

Uncontrolled Intervention Studies

Training and experience have also been investigated with uncontrolled intervention studies that are designed to prospectively assess an intervention’s impact on supervisor outcomes. Borders and Fong (Borders, 1994) employed a within-subject, repeated measures design and found no change in supervisory outcomes in $N = 8$ trainees after they completed a 15-week training program that provided them with experiential (practicum) supervisor training. Specifically, there was no change in cognitions about supervision and no change in participants’ choices of supervisory interventions. Overall, the task of supervision was still regarded as relatively difficult and participants reported high stress and negative assessments regarding their capacity to cope with, and be successful at, the supervision task (Borders & Fong, 1994).

The research team followed this up two years later when they initiated a didactic-experiential 15-week supervisor training program (Borders, 1996), again using a within-subject, repeated measures design. They found support for a positive link between the supervisory training and participants’ skills in conceptualization. Compared to their initial ratings, at the end of the program participants rated supervision as less difficult and rated themselves as better able to cope with the supervision task. Milne and colleagues (Milne, James, Keegan, & Dudley, 2002) employed a longitudinal (8-month), single subject, multiple phase study that involved the supervisor in training providing supervision and receiving SOS him/herself. They found a four-fold increase in advanced supervisory behaviour (e.g., over time the supervisor in training used more guided

experiential learning interventions with supervisees). These findings suggest that didactic-experiential programs (Borders, 1996) or a longer term practicum with SOS (Milne et al., 2002) may have more discernible impact than a program which offers time-limited practical experience on participants' supervisor outcomes including self-assessments of difficulty of supervision and their ability to perform supervision (Borders, 1996), the use of more advanced supervisory interventions (Milne & James, 2002). It should be noted, though, that the small sample sizes of these studies impact the generalizability of their findings.

The impact of a five-day classroom-based (didactic) training program was assessed using a within-subject, repeated measures design with $N = 28$ supervisors in training (Taylor, Gordon, Grist, & Olding, 2012). Participants evaluated themselves on an 18-item measure developed by the researchers and reported an increase in overall supervision competency. These findings stand in some contrast to those reported by others (Borders, 1996; Borders & Fong, 1994; Milne & James, 2002) and Milne and James (2002) in that Newman-Taylor's study is comparatively very short duration (five days, compared to 15-weeks, one year). However, Newman-Taylor's intervention totaled approximately 40-hours, which is comparable to the number of hours in the studies by Borders (1996) and Borders and Fong (1994) research (i.e., three hours a week for 15-weeks, or 45-hours).

The methodology of the intervention studies is an improvement from survey research because the investigators prospectively assessed an intervention's impact on supervisor outcomes. However, there are limits to these designs. First and foremost, the studies lack control groups. Without a control group, the question of whether any

observed improvement or change could occur without the intervention cannot be answered (i.e., is the effect simply due to maturation). The second limit to the above studies is their very small numbers of participants. The three quantitative studies employed a mean of seven participants, thus limiting the power of the studies and their generalizability. The $n = 1$ design of Milne and James (2002), although obviously limited with respect to sample size, does deserve special notice in that it employed an outcome measure that included observation and rating of supervisor's behaviours using a standardized observation tool.

Controlled Intervention Studies

The research from three controlled intervention studies generally indicates training has a perceptible effect on supervisor outcomes. Baker and colleagues (2002) used the PSDS to evaluate a 15-week training program for $n = 12$ doctoral students, as compared to $n = 7$ controls. Significant differences on development were found after the training program between intervention and controls. Supervisory confidence, knowledge, and skills (McMahon & Simons, 2004) increased steadily over the four day course of the training program for $n = 16$ intervention participants, all of whom were practicing professional counselors. K.E. White (1998) used a mixed-methods study to assess changes in $N = 4$ supervisor trainees over a year as they engaged in didactic and experiential/practicum training with SOS. Trainees reported increased confidence in their role as supervisors. However, White (1998) also noted changes in the control group, leading the researcher to conclude that factors external to the supervision experience may affect PSDS scores.

The use of control groups greatly increases the methodological rigour of the studies reviewed above. That said, because the participants were not randomly assigned to conditions, pre-existing differences between the two groups may account for some of the observed changes. For example, the intervention participants may have self-selected into this group, based on an interest in supervision. Small sample sizes are again a concern for these studies, in that they may not possess adequate power to detect changes due to the interventions. There are statistical problems with White's study (1998) in particular, as she did not assess any of the changes via tests of significance, likely due to limited sample size. In addition, the participants in White's study (1998) did not all receive the same intervention. Taken together, these limitations mean that the conclusions that can be drawn from White's (1998) research are restricted. McMahon et al.'s study (2004) used a tool constructed by the research team. Although they report that experts judged the tool as having content and face validity, as well as reporting a Cronbach's alpha of .96, this can only be considered preliminary psychometric evaluation. As such, findings from this study should be interpreted with caution.

Randomized Control Group Studies

Only two studies in this literature employed a randomized controlled design. White and colleagues (E. White & Winstanley, 2010) investigated the impact of a four day classroom-based (didactic) training in clinical supervision on $N = 24$ mental health nurses' clinical supervision practice using a mixed methods design. The Manchester Clinical Supervision Scale (MCSS) (Winstanley, 2000) was administered only to intervention participants before and after the training, so the data related to this scale are effectively derived from a repeated measures, within-subject design study. The MCSS

assesses trust/rapport in supervision, supervisor advice/support, improvement of care/skills, value of supervision, timing and personal issues. Participants' total MCSS scores increased significantly after the course compared to baseline, and this increase was maintained after 12-months. During that 12-month period, trainees were providing supervision, and were supported by visiting study co-coordinators who provided advice and support.

A randomized trial with 46 supervisor/supervisee pairs of mental health clinicians was conducted (Kavanagh et al., 2008) where participant pairs were randomly assigned to one of three interventions (immediate, delayed, and split: pairs did not take the workshop together) in order to assess the impact of supervisor and supervisee joint attendance. The intervention was a didactic two-day workshop. The researchers assessed the impact of the training on: extent the supervision agreement between supervisor and supervisee covered various topic areas, the frequency with which various supervision strategies were employed, self-efficacy to provide effective supervision and address potential challenges, and frequency of supervision problems. All measures were constructed by the researchers for this study. Of interest to the review at hand, there was no overall effect of the intervention on supervisor self-efficacy. Due to these findings, the authors questioned whether a brief workshop format has the potential to bring about change in supervisory skills. Of note, although participants were randomly assigned to one of three intervention formats, participants self-nominated for inclusion in the study and as such may have had a pre-existing interest in supervision. The researchers employed various tools the authors developed to assess the study outcomes. The reliability for self-efficacy of supervisors providing supervision had a Cronbach's alpha of .92.

Rationale for Study 3

A major limitation of the above studies is that the majority of the outcomes were assessed using self-report measures, with only one study assessing change in observed supervisory behaviours (Milne et al., 2002). This reliance on self-report could have various effects. It is possible that those who have participated in training experiences may be more invested in the concept that this training was of use to them, and as such report higher development and/or self-efficacy levels. Two studies (Johnson & Stewart, 2000; Stevens et al., 1998) used a single-item to assess the construct of self-efficacy, which likely limits the reliability and validity of the measurement. Several authors (Caruso et al., 1995; Haley, 2002; Johnson & Stewart, 2008; Kavanagh et al., 2008; McMahon & Simons, 2004) used measures developed by the research team and reported only limited (Haley, 2002; Johnson & Stewart, 2008; Kavanagh et al., 2008; McMahon & Simons, 2004) or no (Caruso et al., 1995) psychometric data on these measures. As noted previously, there is very little consistency across studies in how researchers' operationalized experience and training, thus making it difficult to compare findings across the studies. In sum, measurement problems evident in many studies weaken the conclusions that can be drawn from this research. Nearly twenty years ago, Ellis and colleague's (1996), noted the absence of reliable and valid instruments in the supervision research literature. As noted recently by Watkins (2012), little appears to have changed.

Only two studies used a randomized design to study the impact of training on supervisor development (Kavanagh et al., 2008; E. White & Winstanley, 2010). Those who participated in supervision training across the other studies may have been a select group who were already interested in issues of supervision, which may in turn affect the

study outcomes. Even in the RCT design, participants were not randomly selected, and in one of the studies only those in the intervention group completed post-test measures. Finally, most of the research reviewed above was based on very small sample sizes, which raises the concern about whether the studies had adequate power to detect effects.

With these caveats in mind, a number of patterns and questions do emerge from this line of research. First, there are mixed findings with respect to the relation between supervisory experience and supervisory development and self-efficacy. Some studies found no significant relation (Lyon et al., 2008; Rodolfa et al., 1998; Vidlak, 2002), although both Stevens et al. (1998) and Pelling (2008) reported significant findings, as did Williams (2012). No immediately discernable patterns are apparent that can readily explain why some studies found significant associations whereas others did not.

Second, the picture for supervisory training appears more consistent, in that it appears that training has been linked to higher self-efficacy (Borders, 1996; Johnson & Stewart, 2008; Stevens et al., 1998), and confidence (McMahon & Simons, 2004), although one study found the link between training and self-efficacy only for those who had no prior supervision experience (Haley, 2002). Differences in these findings may be due to measurement: Stevens et al. (1998) used a single item to assess self-efficacy and Johnson and Stewart (2008) used nine items, whereas Haley (2002) used a 42-item measure. It may be that the 42-item measure was able to discern variability across participants in a way a one or nine item measure could not. Other work has also found support for a link between supervisory training and skills in conceptualization (Borders, 1996; Haley, 2002), as well as overall supervisor development (Baker et al., 2002; Lyon

et al., 2008; Vidlak, 2002), and training, and training has been found to account for 4.4% of the variance of PSDS scores (Pelling, 2008).

When the findings for the link between training, supervisory practices, and supervisory development are examined more closely, a subtler picture emerges. Specifically, there appear to be three characteristics that may moderate the relationship between training and supervisory outcomes. First, it appears that the specific type, or format, of the training may impact outcomes. Various studies have found that training which includes a combination of didactic and experiential components, or experiential components with SOS is linked to: increased supervisor self-efficacy (Borders, 1996) for those without any previous supervision experience (Haley, 2002); supervisory confidence (McMahon & Simons, 2004), conceptualization skills (Borders, 1996; Haley, 2002), and supervisor development (Baker et al., 2002; Lyon et al., 2008; Vidlak, 2002). In fact, when findings from a practicum/experiential format alone (Borders & Fong, 1994) were compared with findings from a didactic-practicum course (Borders, 1996), there was a shift toward greater self-efficacy for participants in the didactic-practicum course. Supervision with SOS was also found to predict higher development levels (Pelling, 2008), lending support to the idea that an experiential training combined with a didactic component, or SOS is important. These results for the importance of hands-on training in supervision development are echoed in a meta-analysis (Davis et al., 1999) which found that there were no significant effects of didactic sessions on physician behaviour, but that interactive and didactic-interactive educational sessions were associated with a significant effect on physicians' practices.

Second, it appears that the number of training experiences (i.e., number of training exposures) may moderate the impact on supervisory practice and development. Specifically, three or more training experiences have been linked to higher supervisory self-efficacy (Stevens et al., 1998) and supervisor development (Vidlak, 2002). The third possible moderator is the duration of training, that is, whether it is sustained or not. Vidlak (2002) found that it was training taken over a period of time (e.g., semester course) that was linked to higher development scores. In contrast, there is limited or no apparent impact of concentrated, workshop training on supervisor development (Vidlak, 2002), or self-efficacy—even when assessed using the rigorous RCT methodology (Kavanagh et al., 2008).

Interestingly, we can see the impact of *type* of training, *number* of, and *length* (*duration*) of training experiences when we examine the results of the study by Milne and colleagues (2002) compared to the research by Borders and Fong (1994). Both studies were similar in that they involved hands-on, experiential supervision training, with concomitant SOS. Borders and Fong (1994) found little impact of their supervision practicum, though Milne and colleagues (2002) found greater use of advanced supervisory interventions. Two main differences between the programs were length of time and number of supervisory interactions. Borders and Fong's (1994) intervention was over the course of 15-weeks, whereas Milne et al. (2002) studied a supervisor in training (SIT) over 13 months and 1387 interactions between the SIT and his supervisees.

Although the narrative review above notes some possible patterns in the literature, a meta-analysis can quantitatively evaluate whether these apparent patterns are in fact statistically evident in the included research studies. As such, the focus of the third

study in this dissertation is to examine, via meta-analysis, the relative contribution of supervisory training and experience on supervisor development and self-efficacy. Meta-analyses are an effective tool to assess the consistency of effect sizes across studies, regardless of whether each of the studies' findings in the analysis is significant or not. The meta-analysis includes both significant and nonsignificant findings because *p* values do not allow conclusions to be drawn about the magnitude of effect sizes. This point is often missed in narrative reviews (such as the one above), where non-significant findings are often interpreted to mean there is no effect (Borenstein, 2009). Details of the methods employed in meta-analyses are reviewed in the methods of study three.

Summary of the Studies in the Dissertation

This dissertation is comprised of three studies. The first study is a systematic review that examined the extent to which theoretical models of supervisor development have been used in empirical studies of supervisor development within professional psychology. The research base on supervisor development is important in guiding initiatives in supervision training, because it outlines various changes that supervisors are expected to move through as they grown and learn as supervisors. The current systematic review contributes to the literature on supervisor development by determining whether studies are explicitly theory driven. To that end, the systematic review addressed the question: "To what extent are theoretical models of supervisor development used in professional psychology (clinical, counselling, clinical neuropsychology, school psychology) supervisor development research?".

The second study of this dissertation is an extension of the systematic review (Study 1), which revealed that one theory (Watkins' SCM) dominates the literature on

supervisor development, and that the SCM's associated measure (the PSDS) is the only measure of supervisor development. The second study of this dissertation was designed to assess (1) the reliability and (2) convergent validity of the PSDS across administrations to various study populations, and (3) to explore what study features predicted variance in the reliability and validity. These goals were met via a reliability generalization (RG) and validity generalization (VG) conducted on studies that have employed the PSDS scale. RG is a meta-analytic method that allows reliability data from a measure to be summarized across studies, and VG is a meta-analytic method to assess whether the validity of scores of a given measure are generalizable across samples and studies. The SCM predicts training and experience will impact PSDS scores, thus the current VG study examined the convergent validity of the PSDS scale by analyzing available data on these two constructs. The findings from this RG/VG will aid psychology educators in their decision to use this tool to assess change in supervisors' development as a result of supervision training efforts. It will also aid researchers in the field of supervisor development in understanding more about the samples with which the PSDS tool is appropriately used, and how they might increase reliability within their own future studies.

The third study of this dissertation also builds on the systematic review conducted for Study 1. In the Study 1 systematic review, I found that the majority of research included in the review considered the impact of training and experience in clinical supervision on supervisor development. Therefore, the third study of this dissertation assessed the relative impact of supervision training and experience on supervisor development and self-efficacy, via a meta-analysis. Specifically, I hypothesized that: (1)

training experiences would have a greater effect on supervisory outcomes than will supervisory experience. This was hypothesized because, overall, studies examining training have found a more consistently positive relationship than those examining experience, (2) in keeping with meta-analytic evidence (Davis et al., 1999) in other areas of health care, I hypothesized that prolonged training (e.g., courses, practicum experience) would have a larger effect on supervisory outcomes, when compared with concentrated training interventions (e.g., workshop), (3) larger effect sizes would be associated with training that included an applied component (practicums and/or supervision of supervision; SOS), as compared to classroom based/didactic training. This hypothesis is in keeping with meta-analytic evidence that interactive continuing education sessions with applied components can effect change in professional practice (Davis et al., 1999). The findings from Study 3 will provide valuable information about the relative contribution of experience and training, which, in turn, has implications for how psychology educators might enhance their training efforts.

The Use of Theoretical Models in Psychology Supervisor Development Research from
1994 to 2010:

A Systematic Review*

Keegan Barker John Hunsley

University of Ottawa

*This study was published in *Canadian Psychology*: Barker, K. K., & Hunsley, J. (2013). The Use of Theoretical Models in Psychology Supervisor Development Research from 1994 to 2010: A Systematic Review. *Canadian Psychology, 54*, 176-185. doi: <http://dx.doi.org/10.1037/a0029694>

Abstract

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 comprised of 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 conceptualization, 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

**The Use of Theoretical Models in Supervisor Development Research: A
Systematic Review of the Literature from 1994 to 2010**

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

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 recognized as a key component of pre-licensure 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 emphasize 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 towards 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, Krenzel, & 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 (c.f. 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 in order 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 conceptualization 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, analyzed, 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 hypothesized 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 towards 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 towards 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 in order 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 in order to maintain consistency. The start of the search was indicated as 1994 in order 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 Counseling Psychology*, *Journal of Clinical 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 if 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 if 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, non-experimental) or non-intervention research (descriptive/explanatory, methods/measurement research, other);
2. Aspect of supervision development studied, i.e. main focus of the study;
3. List of theories used in the article; and
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 research was to use theory to understand a sample, another construct, a process, etc. 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 analyzed 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 categorized as "clinical"), and whether the participants were students, interns or professionals.

To examine inter-rater reliability, Cohen's kappa (Cohen, 1960) was calculated for the initial independent ratings of the two reviewers at Phase III. Inter-rater 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). In order 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 manuscripts that were excluded at this stage were reviewed a second time.

Based on this review, no additional manuscripts were identified as potentially relevant. We concluded that because no additional manuscripts 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 non-intervention 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 had 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, 7 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 to 37.5% of the non-intervention studies. Of the intervention studies, 11.1% tested theory compared to 18.8% for non-intervention 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 is 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 due 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 (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 non-empirical 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 employed a panel consensus approach when generating competency benchmarks (Fouad et al., 2009) and a corresponding competency assessment toolkit (Kaslow 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 sixteen 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’ sub-discipline 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 is 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 pre-established 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 categorized 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 and 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 his/her 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.

Acknowledgements

We thank Lee-Anne Ufholz of the University of Ottawa for her expertise in systematic review database searches.

References

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

- Alderson, P. (1998). The importance of theories in health care. *BMJ*, 317(7164), 1007-1010. doi: 10.1136/bmj.317.7164.1007
- Alonso, A. (1983). A developmental theory of psychodynamic supervision. *The Clinical Supervisor*, 7, 23-36.
- American Psychological Association. (2006). APA taskforce on the assessment of competence in professional psychology (Final Report). Retrieved from <http://www.apa.org/ed/resources/competency-revised.pdf>
- *Baker, S. B., Exum, H. A., & Tyler, R. E. (2002). The developmental process of clinical supervisors in training: An investigation of the supervisor complexity model. *Counselor Education and Supervision*, 42, 15-30.
- Bandura, A. (1986). Self efficacy. In A. Bandura (Ed.), *Social foundations of thought and action: A social cognitive theory* (pp. 390-453). Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, 44, 1175-1184.
- Bandura, A. (1997). *Self efficacy: The exercise of control*. New York: Freeman.
- *Barnes, K. L., & Moon, S. M. (2006). Factor structure of the Psychotherapy Supervisor Development Scale. *Measurement and Evaluation in Counseling and Development*, 39, 130-140.
- Bernard, J. M., & Goodyear, R. K. (1998). *Fundamentals of clinical supervision (2nd ed.)*. Needham Heights, MA: Allyn & Bacon; US.

- Bernard, J. M., & Goodyear, R. K. (2004). *Fundamentals of clinical supervision* (3rd ed.). Boston: Pearson Education.
- Borders, L. D. (1989). A pragmatic agenda for developmental supervision research. *Counselor Education and Supervision, 29*, 16-24.
- *Borders, L. D. (1996). Impact of a counseling supervision course on doctoral students' cognitions. *Counselor Education and Supervision, 35*, 204-217.
- *Borders, L. D., & Fong, M. L. (1994). Cognitions of supervisors-in-training: An exploratory study. *Counselor Education and Supervision, 33*, 280-293.
- Canadian Psychological Association. (2011). Accreditation standards and procedures for doctoral programmes and internships in professional psychology (Fifth revision). Ottawa: Canadian Psychological Association.
- *Caruso, M., Dooley, K., & Dabit, J. (1995). *An evaluation of a clinical supervision course format*. Paper presented at the Annual Conference of the Midsouth Educational Research Association.
- Charmaz, K. (2003). Grounded theory. In J. A. Smith (Ed.), *Qualitative Psychology: A Practical Guide to Research Methods* (pp. 81-110). Thousand Oaks, California: Sage Publications, Inc.
- Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology, 23*, 695-700.

- Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- Davies, P., Walker, A. E., & Grimshaw, J. M. (2010). A systematic review of the use of theory in the design of guideline dissemination and implementation strategies and interpretation of the results of rigorous evaluations. *Implementation Science, 5*, 14. doi: 10.1186/1748-5908-5-14
- Ellis, M. V. (1991). Research in clinical supervision: Revitalizing a scientific agenda. *Counselor Education and Supervision, 30*, 238-251.
- Ellis, M. V., Ladany, N., Kregel, M., & Schult, D. (1996). Clinical supervision research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology, 43*, 35-50. doi: 10.1037/0022-0167.43.1.35
- Falender, C. A., Cornish, J. A., Goodyear, R., Hatcher, R., Kaslow, N. J., Leventhal, G., . . . Grus, C. (2004). Defining competencies in psychology supervision: A consensus statement. *Journal of Clinical Psychology, 60*, 771-785. doi: 10.1002/jclp.20013
- Falender, C. A., & Shafranske, E. P. (2004). *Clinical supervision: A competency-based approach* (1st ed.). Washington, DC: American Psychological Association.
- Forscher, B. K. (1963). Chaos in the Brickyard. *Science, 142*, 339. doi: 10.1126/science.142.3590.339

- Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B., . . . Rodolfa, E. R. (2009). Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology, 3*, S5-S26. doi: <http://dx.doi.org/10.1037/a0015832>
- Goodyear, R. K., Bunch, K., & Claiborn, C. D. (2005). Current supervision scholarship in psychology: A five year review. *Clinical Supervisor, 24*, 137-147. doi: 10.1300/J001v24n01_07
- *Green, D., & Dye, L. (2002). How should we best train clinical psychology supervisors? A Delphi survey. *Psychology Learning & Teaching, 2*, 108-115.
- Green, J. (2000). The role of theory in evidence-based health promotion practice. *Health Education Research, 15*, 125-129.
- *Haley, S. J. (2002). *The influence of supervision training on supervisor self-efficacy among doctoral interns at university counseling centers*. Haley, Sarah Jane: West Virginia U , US. ProQuest Dissertations and Theses database. (725930541)
- Hess, A. K. (1986). Growth in supervision: Stages of supervisee and supervisor development. *Clinical Supervisor, 4*, 51-67.
- Hess, A. K. (1987). Psychotherapy supervision: Stages, Buber, and a theory of relationship. *Professional Psychology: Research and Practice, 18*, 251-259.
- *Hillman, S. L., McPherson, R. H., Swank, P. R., & Watkins, C. E., Jr. (1998). Further validation of the psychotherapy supervisor development scale. *The Clinical Supervisor, 17*, 17 - 32. doi: 10.1300/J001v17n01_02

- Howes, J. L., Vallis, T., Wilson, A. R., Ross, M. A., & Louisy, H. J. (1996). Predoctoral internship training in Canada: Internship settings and supervisory issues. *Canadian Psychology, 37*, 173-179. doi: 10.1037/0708-5591.37.3.173
- Hunsley, J., & Barker, K. K. (2011). Training for competency in professional psychology: A Canadian perspective. *Australian Psychologist, 46*, 142-145.
- *Johnson, E. A., & Stewart, D. W. (2000). Clinical supervision in Canadian academic and service settings: The importance of education, training, and workplace support for supervisor development. *Canadian Psychology, 41*, 124-130.
- *Johnson, E. A., & Stewart, D. W. (2008). Perceived competence in supervisory roles: A social cognitive analysis. *Training and Education in Professional Psychology, 2*, 229-236. doi: 10.1037/1931-3918.2.4.229
- Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency Assessment Toolkit for professional psychology. *Training and Education in Professional Psychology, 3*(4, Suppl), S27-S45. doi: 10.1037/a0015833
- *Kavanagh, D. J., Spence, S., Sturk, H., Strong, J., Wilson, J., Worrall, L., . . . Skerrett, R. (2008). Outcomes of training in supervision: Randomised controlled trial. *Australian Psychologist, 43*, 96-104. doi: 10.1080/00050060802056534
- Lefebvre, C., Manheimer, E., Glanville, J., & Group, C. I. R. M. (2008). Searching for studies. In J. Higgins, S. Green & Cochrane Collaboration. (Eds.), *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. Chichester, UK: Wiley-Blackwell.

- *Lyon, R. C., Heppler, A., Leavitt, L., & Fisher, L. (2008). Supervisory training experiences and overall supervisory development in predoctoral interns. *Clinical Supervisor, 27*, 268-284. doi: 10.1080/07325220802490877
- *Magnuson, S., Norem, K., Jones, N., McCrary, J. C., & Gentry, J. (2000). The triad model as a cross-cultural training intervention for supervisors. *The Clinical Supervisor, 19*, 197-210. doi: 10.1300/J001v19n02_12
- *Majcher, J.-A., & Daniluk, J. C. (2009). The process of becoming a supervisor for students in a doctoral supervision training course. *Training and Education in Professional Psychology, 3*, 63-71. doi: 10.1037/a0014470
- *McMahon, M., & Simons, R. (2004). Supervision training for professional counselors: An exploratory study. *Counselor Education and Supervision, 43*, 301.
- Milne, D. L. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology, 46*, 437-447.
- *Nelson, K. W., Oliver, M., & Capps, F. (2006). Becoming a supervisor: Doctoral student perceptions of the training experience. *Counselor Education and Supervision, 46*, 17-31.
- Painter, J. E., Borba, C. P., Hynes, M., Mays, D., & Glanz, K. (2008). The use of theory in health behavior research from 2000 to 2005: A systematic review. *Annals of Behavioral Medicine, 35*, 358-362.
- Pedersen, P. (1994). *A handbook for developing multicultural awareness (2nd ed)*. Alexandria, VA: American Counseling Association.

- *Pelling, N. (2008). The relationship of supervisory experience, counseling experience, and training in supervision to supervisory identity development. *International Journal for the Advancement of Counselling*, 30, 235-248.
- *Robiner, W. N., Saltzman, S. R., Hoberman, H. M., & Schirvar, J. A. (1997). Psychology supervisors' training, experiences, supervisory evaluation and self-rated competence. *Clinical Supervisor*, 16, 117-144. doi: 10.1300/J001v16n01_07
- Rodenhauser, P. (1994). Toward a multidimensional model for Ppsychotherapy supervision based on developmental stages. *Journal of Psychotherapy Practice and Research* 3, 1-15.
- Rodenhauser, P. (1997). Psychotherapy supervision: Prerequisites and problems in the process. In C. E. J. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 527-548). New York Wiley.
- *Rodolfa, E. R., Haynes, S., Kaplan, D., Chamberlain, M., Goh, M., Marquis, P., & McBride, L. (1998). Supervisory practices of psychologists--Does time since licensure matter? *The Clinical Supervisor*, 17, 177-183. doi: 10.1300/J001v17n02_12
- *Ronnestad, M. H., Orlinsky, D. E., Parks, B. K., Davis, J. D., & The Society for Psychotherapy Research (SPR) Collaborative Research Network. (1997). Supervisors of psychotherapy: Mapping experience level and supervisory confidence. *European Psychologist*, 191-201.
- *Stevens, D. T., Goodyear, R. K., & Robertson, P. (1998). Supervisor development: An exploratory study in changes in stance and emphasis. *Clinical Supervisor*, 16, 73-88.

- Stoltenberg, C. D., & Delworth, U. (1987). *Supervising counselors and therapists*. San Francisco, CA: Jossey-Bass.
- Stoltenberg, C. D., McNeill, B., & Delworth, U. (1998). *IDM supervision: An integrated developmental model for supervising counselors and therapists*. San Francisco, CA: Jossey-Bass.
- Stoltenberg, C. D., McNeill, B. W., & Crethar, H. C. (1994). Changes in supervision as counselors and therapists gain experience: A review. *Professional Psychology: Research and Practice*, 25, 416-449. doi: 10.1037/0735-7028.25.4.416
- *Vidlak, N. W. (2002). *Identifying important factors in supervisor development: An examination of supervisor experience, training, and attributes*. Vidlak, Nicole White: U Nebraska - Lincoln, US. ProQuest Dissertations and Theses database. (727413951)
- *Vieceli, A. E. (2007). *A comparison of supervisor development and supervision emphasis of academic and on-site supervisors*. Vieceli, Ada E.: Southern Illinois U Carbondale, US. ProQuest Dissertations and Theses database. (1212791611)
- Watkins, C. E., Jr. (1990). Development of the psychotherapy supervisor. *Psychotherapy: Theory, Research, Practice, Training*, 27, 553-560. doi: 10.1037/0033-3204.27.4.553
- Watkins, C. E., Jr. (1993). Development of the psychotherapy supervisor: Concepts, assumptions, and hypotheses of the supervisor complexity model. *American Journal of Psychotherapy*, 47, 58-74.
- Watkins, C. E., Jr. (1994). The supervision of psychotherapy supervisor trainees. *American Journal of Psychotherapy*, 48, 417-431.

- Watkins, C. E., Jr. (1995). Psychotherapy supervisor development: On musings, models, and metaphor. *Journal of Psychotherapy Practice & Research*, 4, 150-158.
- *Watkins, C. E., Jr., Schneider, L. J., Haynes, J., & Nieberding, R. (1995). Measuring psychotherapy supervisor development: An initial effort at scale development and validation. *The Clinical Supervisor*, 13, 77-90. doi: 10.1300/J001v13n01_06
- *White, K. E. (1998). *Becoming a supervisor: An intensive study of the early development of clinical supervisors*. White, Karen E.: U Manitoba, Canada. ProQuest Dissertations and Theses database. (736787841)

Table 1

PsycInfo Search

Search term number	Search term(s)
1.	clinical psychologists/
2.	exp clinical psychology/
3.	exp counseling psychology/
4.	mental health personnel/ or school psychologists/
5.	exp psychological assessment/
6.	neuropsychology/
7.	psychologists/ or educational psychologists/
8.	postgraduate students/ or graduate students/
9.	psychology graduate student*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
10.	practicum supervision/
11.	professional supervision/
12.	clin* supervis*.mp.
13.	practic* supervis*.mp.
14.	prof* supervis*.mp.
15.	supervis*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
16.	10 or 11 or 12 or 13 or 14 or 15
17.	clinical psychology graduate training/ or clinical psychology internship/ or counselor education/
18.	internship programs/
19.	graduate psychology education/
20.	educational program accreditation/
21.	postgraduate training/
22.	psychotherapy training/
23.	professional consultation/
24.	professional development/ or continuing education/ or mental health inservice training/ or mentor/ or professional competence/
25.	career development/ or curriculum development/ or program development/
26.	inservice training/ or mental health inservice training/
27.	postgraduate training/
28.	training/
29.	clinical methods training/ or psychotherapy training/
30.	graduate train*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
31.	postgraduate train*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
32.	profession* develop*.mp. [mp=title, abstract, heading word, table of contents, key concepts]

33. continuing education*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 34. program* develop*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 35. education* curric*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 36. 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35
 37. Counseling Psychologists/
38. clinical psych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 39. neuropsych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 40. school psych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 41. education* psych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 42. health psych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 43. medic* psych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 44. clinical psych* intern*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 45. psych* practicum.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 46. counsel* psych*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 47. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46
 48. develop*.mp. [mp=title, abstract, heading word, table of contents, key concepts]
 49. 36 or 48
 50. 16 and 47 and 49
 51. limit 50 to yr="1994 -Current"
 52. limit 51 to English language
-

Table 2

Characteristics of Study Samples and Foci

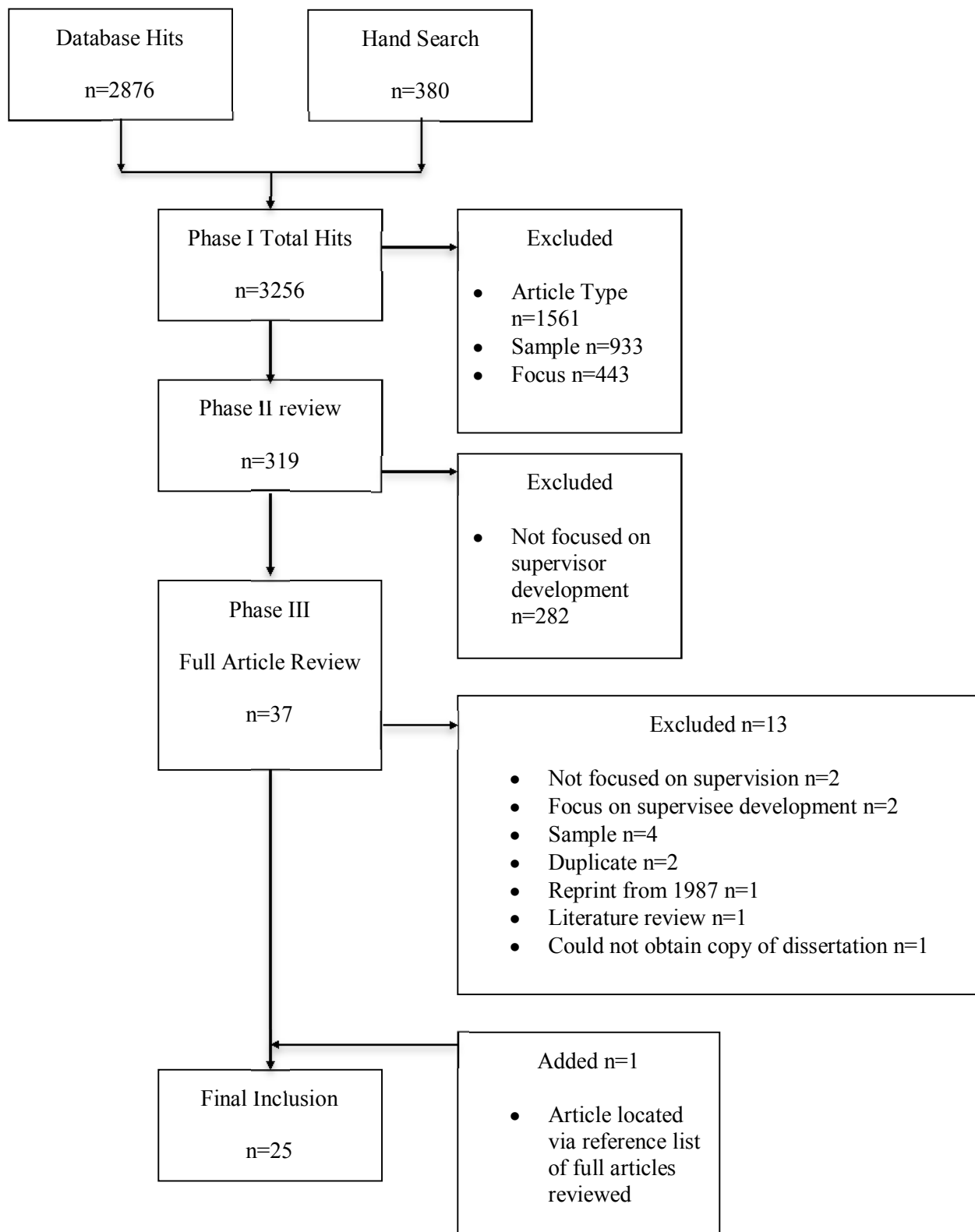
Study	Sample		Study's major area of focus
	<i>n</i>	Characteristics	
Borders & Fong (1994)	9	doctoral level students; CACREP approved program Counsellor education; enrolled in required supervision course	Experiences of supervisors-in-training
Borders et al. (1996)	11	doctoral level students; CACREP approved program Counsellor education; enrolled in required supervision course	Experiences of supervisors-in-training
Caruso et al. (1995)	27	post/doctoral level students; counsellor education program; completed a supervision course	Impact of a training program
Hillman et al. (1998)	43	clinical supervisors: university counselling centres; supervisor trainees enrolled in supervision courses	Scale Development (PSDS)
Baker, Exum, & Tyler (2002)	16	doctoral level participants in supervision practicum	Impact of a training program
Barnes & Moon (2006)	225	members of CACREP counsellor education programs conducting supervision or enrolled in supervision course	Scale Development (PSDS)
Green & Dye (2002)	50	expert panel members (Delphi survey)	Other: Delphi approach to identify components of supervisor training
Haley (2002)	145	Pre-doctoral interns; APA accredited university counselling sites	Influence of past training & experience on supervisory practice
Johnson & Stewart (2000)	156	supervisors from 45 Canadian university or clinical settings	Influence of past training & experience on supervisory practice
Hillman et al. (1998)	43	Clinical supervisors: university counselling centres; supervisor trainees enrolled in supervision courses	Scale Development (PSDS)
Johnson & Stewart (2008)	156	Psychologists providing supervision at academic & internship programs	Other: Supervisor self-efficacy
Kavanagh et al. (2008)	150	Supervision pairs: allied health practitioners in Mental Health service of Queensland Health	Impact of a supervision training program
	46		

Study	Sample		Study's major area of focus
	<i>n</i>	Characteristics	
Lyon et al. (2008)	233	Pre-doctoral interns at APA accredited sites	Influence of past training & experience on supervisory practice
Magnuson et al. (2000)	4	Doctoral level internship course participants	Impact of a supervision training program
Majcher & Daniluk (2009)	6	Second year doctoral students; counselling psychology; enrolled supervision course	Experiences of supervisors-in-training
McMahon & Simons (2004)	99	Australian counsellors	Impact of a supervision training program
Nelson, Oliver, & Capps (2006)	18	Doctoral students in supervision practicum/internship	Experiences of supervisors-in-training
Pelling (2008)	175	Members: Association for Counsellor Education and Supervision	Influence of past training & experience on supervisory practice
Robiner et al. (1997)	62	Supervisors APA accredited clinical psychology internships	Influence of past training & experience on supervisory practice
Rodolfa et al. (1998)	131	State licensed psychologists	Influence of past training & experience on supervisory practice
Ronnestad et al. (1997)	1639	Psychotherapists from 12+ countries	Influence of past training & experience on supervisory practice
Stevens, Goodyear, & Robertson (1998)	60	Practicing psychotherapists; 12 cities Southern California	Influence of past training & experience on supervisory practice
Vieceli (2007)	37	Clinical supervisors CACREP accredited universities; community-based clinical supervisors	Other: Comparison of academic and community supervisors
Vidlak (2002)	99	Pre-doctoral intern supervisors at APA internship sites	Influence of past training & experience on supervisory practice
Watkins et al. (1995)	335	Members APA Division of Psychotherapy	Scale Development (PSDS)
White (1998)	8	PhD student supervisee trainees	Other: Self perceptions of early supervisors

Table 3
Research type and extent of theory use of reviewed articles

Characteristics of Articles	Number (%) of articles
All articles in final sample	n=25
Intervention research	n=9 (36%)
Randomized controlled trial	n=1 (4%)
Quasi-experimental	n=2 (8%)
Non-experimental	n=6 (25%)
Non-intervention research	n=16 (64%)
Descriptive/explanatory research	n=13 (52%)
Methods/measurement research	n=3 (12%)
Other	n=0
Extent of use of Supervisor Development Model	
No theoretical model	n=13 (52%)
Informed by theory	n=4 (16%)
Applying theory	n=4 (16%)
Testing theory	n=4 (16%)
Building theory	n=0

Figure 1: Study Flow Diagram



Reliability and validity of the Psychotherapy Supervisor Development Scale:

A meta-analytic evaluation*

Keegan Barker John Hunsley

University of Ottawa

*This study has been accepted for publication (without revisions) in *The Clinical Supervisor*: Barker, K. K., & Hunsley, J. (in press). Reliability and validity of the Psychotherapy Supervisor Development Scale: A meta-analytic evaluation. *The Clinical Supervisor*.

Abstract

The Psychotherapy Supervisor Development Scale (Watkins, Schneider, Haynes, & Nieberding, 1995) is a measure developed to assess supervisor development as outlined by the Supervisor Complexity Model (Watkins, 1990, 1993, 1994). There is preliminary evidence of adequate psychometric properties for the measure in the published literature (Barnes & Moon, 2006; Hillman, McPherson, Swank, & Watkins, 1998; Watkins et al., 1995), but it remains a scale that requires more evaluation. We conducted reliability and validity generalization meta-analyses in order to better understand the PSDS's psychometric properties. The mean internal reliability generated from 1, 163 participants (6 studies) was $r = 0.93$, 95% CI [0.91, 0.94]. As for the scale's convergent validity, the mean effect related to supervision experience derived from 948 participants (5 studies) was $g = 0.40$, 95% CI [0.17, 0.62], and the mean effect related to supervision training derived from 448 participants (4 studies) was $g = 1.13$, 95% CI [0.53, 1.73]. For each analysis, we characterize the variance in scores across studies, and consider sample and study characteristics that are predictive of the reliability and validity coefficients. We discuss the importance of conducting sample-specific reliability analyses and make recommendations for future research on the development of supervisor competency.

Keywords: professional supervision, practicum supervision, professional development, meta-analysis, reliability, validity

Reliability and Validity of the Psychotherapy Supervisor Development Scale: A meta-analysis

Providing high quality supervision is regarded as an important competency for professional psychologists in many countries, including the United States, (American Psychological Association, 2006), Canada (see <http://www.cpa.ca/docs/file/MRA.pdf>), and the United Kingdom (Milne et al., 2010). Ideally, the supervisor—who is indeed ultimately responsible for both the service that is provided to the client and for the supervisory process itself—is functioning at a high level of supervisory competency.

Supervisory competency is defined as the requisite knowledge, skills and values/attitudes required to deliver effective supervision (Falender et al., 2004). Supervisor development is defined as supervisors moving through specific processes (Watkins, 1995), developmental stages, or defined periods of growth (Watkins, 1993) regarding their role as a supervisor, with the ultimate goal of becoming a “seasoned, master practitioner” (Watkins, 2012, p. 48).

--in other words, a practitioner who possesses the knowledge, skills and attitudes necessary to competently provide supervision.

Watkins’ Supervisor Complexity Model (SCM; Watkins, 1990, 1993, 1994) is the model of supervisor development that has received the most empirical attention to date (Barker & Hunsley, 2013; Watkins, 2010). The SCM (Watkins, 1990, 1993, 1994) outlines four stages through which supervisors are believed to progress in their trajectories toward becoming “seasoned, master” (Watkins, 2012, p. 48) supervisor: role shock, role recovery and transition, role consolidation, and role mastery. In each stage, supervisors must develop so that they are more competent, autonomous, and have greater

self-identity and self-awareness than in previous stages. The SCM identifies variables that are posited to influence movement from one stage to another, including: personality factors (openness, flexibility, motivation), and environmental supports (such as peer and senior colleagues to help one process and learn from the supervisory experience). The SCM notes that the developmental process of supervisors can be impacted by the training and experience in supervision, as it is posited that this training will allow them to more “optimally reflect upon and draw from their (supervisory) experiences” (Watkins, 1993, p. 68), thereby facilitating growth through the various stages of the model.

The SCM is the only model of supervisor development that has a measure associated with it: the Psychotherapy Supervisor Development Scale (PSDS; Watkins et al., 1995). To our knowledge, none of the other models of supervisor development (Hess, 1986, 1987; Rodenhauser, 1994, 1997; Stoltenberg & Delworth, 1987; Stoltenberg, McNeill, & Delworth, 1998; Watkins, 1990, 1993, 1994) have measures associated with them, which may explain, in part, the paucity of research on these models (Barker & Hunsley, 2013) and in the area in general (Ellis, Ladany, Krenzel, & Schult, 1996). The PSDS is a self-report measure designed to assess the key polarities outlined in the SCM: (1) competency versus incompetency, (2) self-awareness versus unawareness, (3) identity versus identity diffusion, (4) autonomy versus dependence (Watkins, 1993). These polarities were selected for scale development because they cut across all four stages of the SCM, reflect core developmental concerns and should, according to SCM theory, vary by stage (Watkins et al., 1995). The scale has 18 items, and uses a 7-point response scale (1 = never, 4 = half of the time, 7 = always). Its four factor structure has been supported by factor analytic research (Barnes & Moon, 2006). Moreover, there is some

preliminary evidence of adequate psychometric properties for the PSDS in the published literature (Barnes & Moon, 2006; Hillman et al., 1998; Watkins et al., 1995). Yet as Watkins (2012) himself noted, the PSDS remains a largely unproven scale that requires more study.

Considering that the PSDS is the only measure of its kind, it is thus central in the research now being done on supervisor development. Indeed, a recent systematic review (Barker & Hunsley, 2013) noted that Watkins' SCM is the "only game in town" with respect to the empirical study of supervisor development theories. As such, it is important that its basic psychometric properties are further understood. The primary goal in this study, therefore, is to examine the evidence of the reliability and validity associated with scores on the PSDS measure.

Reliability Generalization and Validity Generalization

Reliability refers to the repeatability or stability of scores. Because reliability coefficients are based on data obtained from a specific sample, the reliability coefficient is a property of the scores on that measure for that particular sample, rather than a property of the measure itself. That is, reliability is a function of both the measure, and the sample on which the measure is used. An instrument will likely yield scores with varied reliabilities across different administrations if the samples across these administrations differ (Henson & Thompson, 2002). Accordingly, each time a measure is used, investigators should report the reliability values for their specific samples, rather than relying on the estimates reported in the manual or from previous research using the instrument (Haynes, Smith, & Hunsley, 2011).

Reliability generalization (RG) is a meta-analytic method that allows reliability data from a measure to be summarized across studies. The findings from an RG do not replace the need for authors to report study-specific reliability values; rather, the RG estimates provide a guideline for future authors to consider in deciding whether the tool may be one that is reliable for their particular sample. The results of an RG can provide an estimate of the typical reliability scores across studies, the amount of variability in these reliability coefficients, and the sources of variability that influence these reliability coefficients (Vacha-Haase, 1998). An RG study can (a) indicate which sample and study features are relevant to obtaining higher score quality for the measure (Henson & Thompson, 2002) and (b) provide estimates of the level of reliability likely to be found in data from samples similar to those included in the RG.

Validity refers to the degree to which scores on a measure reflect the phenomenon it purports to measure. Convergent validity refers to the degree to which the data from a measure are coherently related to data on other measures of the same construct, as well as to other theoretically related variables (Haynes et al., 2011). For example, the model on which the PSDS is based—the SCM (Watkins, 1990, 1993, 1994)—predicts that supervisors who are more developed will show higher levels of competency, autonomy, identity and self-awareness in their supervisory activities. It also predicts that a supervisor's developmental level will be affected by supervisory training and experience.

Validity generalization (VG; Schmidt & Hunter, 1977) is a meta-analytic method to assess whether the validity of scores of a given measure are generalizable across samples and studies. In VG, the means, standard deviations, and other descriptive statistics are computed for the validity coefficients across a sample of studies that all

employ the same measure. Further, as in a RG study, the features of the studies (e.g., sample composition and sample size) are used to understand characteristics of the reviewed studies that moderate the variability observed across the obtained validity coefficients (Vacha-Haase, 1998).

The Present Study

The first purpose of the current study was to assess the typical PSDS reliability coefficients and the variability of these coefficients across administrations. We also explored what study features predict variance in the reliability coefficients for the PSDS. These goals were met via an RG conducted on studies that have employed the PSDS scale. The second purpose of the current study was to assess the convergent validity of the PSDS. The SCM predicts training and experience will impact PSDS scores, thus the current VG study examined the convergent validity of the PSDS scale by analyzing available data on these two constructs.

Meta-analytic evidence can be used for many purposes, one of which is obtaining a general overall picture of the evidence in a topic area in order to direct research efforts (Petticrew & Roberts, 2006). Although there are a relatively small number of studies that have used the PSDS, it is important to understand the extent of the evidence of reliability and validity, given that the PSDS is currently the only measure available to assess supervisor development. Meta-analytic procedures increase the utility of individual studies, as a function of the “borrowed strength” (Rosenthal, 1994, p. 131) that results when even a small number of studies are combined. Further, small scale meta-analyses can guide future research by providing researchers with emerging information about a

subject area and informing them about the reporting of study characteristics and details in future research (Dewey & Hunsley, 1990).

Method

Literature Search

The search strategy for a recent systematic review on supervisor development by Barker and Hunsley (2013) was used as the basis for the current search strategy. Briefly, this involved developing a comprehensive search strategy for PsycINFO using the subject headings of supervision and development/training. Related search terms associated with the constructs, as defined by PsycINFO's subject headings, were recorded until no new applicable search terms were identified. In addition to the subject headings, applicable keywords were developed. The search strategy was then used as a template for the development of additional searches in the MedLine, CINAHL and ERIC databases, thereby individualizing the search for each database. This general search strategy was used to identify potentially relevant studies (including both published studies and unpublished dissertations) available as of the end of 2011. No discipline limits were included in the search so that any study using the PSDS could be identified. Additionally, a specific keyword search for "PSDS" was implemented to supplement this main search strategy.

The abstracts identified in the search were reviewed in various phases. In Phase I, one author (KB) reviewed abstracts to exclude theoretical papers, book reviews, and letters to the editor. In Phase II, one author (KB) reviewed the abstracts to exclude those not focused on supervisor development, defined broadly as examining supervisors

moving through specific processes (Watkins, 1995), developmental stages, or defined periods of growth (Watkins, 1993) regarding supervision.

Additionally, a hand search was conducted to augment the electronic search. Key journals' tables of contents 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 *Journal of Counseling Psychology*, *Journal of Clinical Psychology*, *Professional Psychology: Research & Practice*, 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*. Reference lists of included articles and dissertations were searched to identify any relevant publications. Finally, the author of the PSDS scale was contacted to obtain any additional references of studies that had used the scale.

Duplicates and reprints were excluded, as were papers that could not be obtained. Figure 1 is the flow diagram of the study. In summary, 5,509 abstracts were reviewed, with 84 documents meeting our criteria to warrant review of the full document. Of these, 71 were excluded, so that the final sample of documents included 13 studies. Upon closer review, two studies (Barnes, 2002; Crook-Lyon, Presnell, Silva, Suyama, & Stickney, 2011) were based on the same data set as another study (Barnes & Moon, 2006; Lyon, Heppler, Leavitt, & Fisher, 2008), and thus the latter studies were retained because they contained more complete information. One study (Vieceli, 2007) did not report reliability information, and although it reported PSDS data regarding experience and training, the

PSDS data were combined with another measure in the analyses. As a result, this study was excluded from our analyses.

The final sample (indicated in the reference list with an asterisk) was based on 10 studies (including 3 unpublished dissertations). To be included in the final sample, studies had to report reliability data of PSDS scores and/or report PSDS data of different training and/or experience levels. Due to the limited number of studies in this area, all study designs were included, as were all study populations. Further, both published and unpublished studies were included. The wide inclusion criteria was regarded as appropriate to the study's goal of presenting a general overall picture of the reliability and validity evidence of the PSDS, in order to direct future research efforts. Because of the inclusion of unpublished studies, evaluating the possibility of publication bias affecting results was not necessary. We did not, therefore, conduct analyses to determine the fail safe N or generate funnel plots of effects.

Data Extraction

When reported in a study, both mean PSDS scores (coded as a continuous variable) and standard deviation of PSDS scores (coded as a continuous variable) were extracted. Any reliability data that were reported in the studies were extracted, as were data relating to convergent validity. Information on participants' experience in providing supervision (mean and standard deviation of years supervising) was extracted, as well as any statistical results pertaining to PSDS scores and supervisory experience. The second aspect of convergent validity data that was extracted involved data related to PSDS scores and supervisory training (type and extent of training; defined as efforts aimed at improving a participant's knowledge, skills or attitudes towards supervision).

As noted above, because reliability estimates are largely a function of total score variance, issues of group homogeneity become central in explaining potential fluctuations in scores (Henson & Thompson, 2002). Accordingly, the following information was extracted from the studies (as continuous variables) so that they could be examined as possible moderating variables in any observed heterogeneity of effect sizes: year of publication, sample size, participants' mean age, participants' gender (percentage female), and percentage of participants with doctoral degrees. Whether an author on the study being assessed was an author of the PSDS was coded, as was whether the majority of study participants were in professional psychology (coded dichotomously). When any of the above data were not included in the article, the study author was contacted to obtain this information. Emails requesting information were sent to 10 authors concerning reliability coefficients and 5 responses were received (response rate of 50%). Regarding validity coefficients, emails requesting information were sent to 8 authors, and 4 responses were received for a response rate of 50%. It should be noted that some authors were contacted regarding questions of both reliability and validity coefficients. Of all the queries that were sent (emails to 10 different authors), only one author provided information that was ultimately included in the study. The others either did not have the information available or did not respond to the requests.

In any meta-analysis, the issue of non-independence of observations is a concern. As is frequently done in meta-analysis (e.g., Parker, Hanson, & Hunsley, 1988), we controlled for this by having each study be allowed to contribute one observation for each measurement category (alpha coefficient, test-retest coefficient, convergent validity coefficient) and statistic used. If more than one observation was available for a

measurement category, the observation that contained the most complete information for calculating an effect size was used. A random effects model was used for all computations and analyses, as it allows for the possibility that the effect size may be influenced by participant variables (e.g., age, training, experience) or variations in the intensity of any interventions offered to participants (Borenstein, 2009).

Data Analysis: Reliability Generalization

There were three main steps in conducting the RG. The first involved the characterization of the reliability information on the PSDS scale across studies. At this step, the effect sizes for each study were calculated with the program Comprehensive Meta Analysis, version 2.0 (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005). Because the reliability data were correlational in nature, the following information was entered for each study: r value, sample size, effect direction. The standard error of r , Fisher's z , and its standard error were calculated for each study. The subsequent computations were carried out using the Fisher's z scale transformed values. Meta-analyses are not generally performed on the correlation coefficient itself because the variance depends strongly on the correlation (Borenstein, 2009). However, in order that the reliability coefficient findings make intuitive sense, the final results were converted back to r values. Following these transformations, the second step was to calculate the overall effects across studies. In the third step, the heterogeneity of these overall effect sizes was examined, followed by regression analyses to explore how well the coded study features predicted variations in each of the RCs.

Data Analysis: Validity Generalization

The three main steps outlined for the RG were followed for the VG. That is, effect sizes for individual studies were calculated, followed by calculation of overall summary effects, and then an assessment of heterogeneity and subsequent moderator analyses. These steps were carried out separately for the coefficients related to the constructs of supervisory experience and for supervisor training. In order to maintain independence of observations, each study was allowed to contribute one coefficient for the analysis related to experience and one for the construct of training.

For the VG, we used the Campbell Collaboration effect size calculator (Wilson, 2013) to calculate each study's effect size. This was done because the CMA program does not handle multi-group ANOVA design, which was used in two studies (Bencivenne, 1999; Watkins et al., 1995). For these two studies, the "low experience" group was treated as the control condition, and the "mid" and "high" experience groups were considered as treatment conditions when using the effect size calculator. One study (Bencivenne, 1999) did not report the number of participants in each of the experience and training subgroups, so the total sample size of the study was divided by the number of subgroups to obtain an estimate of the number of participants per subgroup.

Each study's summary effect size was then entered into CMA in order to calculate overall effect sizes (random effects model) from the 6 studies that examined experience and 4 studies that examined training. Any applicable meta-regressions (random effects model, method of moments) were conducted using CMA. For the VG, Hedge's g was used because both d and r values can be translated into this summary effect index.

Results

Overview of Studies

The final sample included 10 studies and a total of 1,649 participants, 918 (55.7%) of whom were female. Mean age of participants was 44.2 years. Of the 8 studies that reported information on degrees, 830 (64%) participants had doctorates. The participants varied in their professional backgrounds, and included supervisors who were psychologists (Vidlak, 2002; Watkins et al., 1995), supervisors from university counseling centers, (Hillman et al., 1998), marriage and family therapists (Bencivenne, 1999), doctoral students (Baker, Exum, & Tyler, 2002), predoctoral interns (Lyon et al., 2008), counseling education supervisors (Barnes, 2002; Pelling, 2008), substance abuse counselor supervisors (Culbreth & Cooper, 2008), and genetic counselors (Lee, Veach, & LeRoy, 2009). Mean years of supervision experience was 10.2 across the seven studies that reported this data. The mean (baseline) PSDS score from the nine studies that reported it was 102, with an average standard deviation of 10.9. Table 1 provides an overview of the studies.

Reliability Generalization

Of the 10 studies reviewed, 5 (50%) (Barnes, 2002; Lee et al., 2009; Lyon et al., 2008; Pelling, 2008; Watkins et al., 1995) studies reported internal consistency statistics (Cronbach's alpha) based on the study sample's scores. Of the 5 studies that did not report an alpha value (Baker et al., 2002; Bencivenne, 1999; Culbreth & Cooper, 2008; Hillman et al., 1998; Vidlak, 2002), each referenced the original PSDS alpha value (Watkins et al., 1995) and 2 of these studies also referenced other reliability statistics of an early follow-up study on the PSDS scale (Hillman et al., 1998). Table 2 shows the

results of the reliability generalization. The summary reliability coefficient is also reported after a sixth study's (Hillman et al., 1998) split half reliability coefficient was added to the analysis, in order to provide an overall summary reliability coefficient of the internal consistency of the PSDS. Whether the split half reliability coefficient was included or not had little effect on the overall results: in both case the mean reliability value was greater than .90.

Two of the 10 studies reported test-retest reliability coefficients. The first study (Hillman et al., 1998) reported $r = 0.85$, $n = 38$ based on a test-retest interval of 4 weeks, and the second study (Bencivenne, 1999) reported $r = 0.86$, $n = 73$ (also over an interval of 4 weeks). Thus, based on a random effects analysis, the summary RC for test-retest RCs was $r = 0.86$, $p < 0.001$, 95% CI [0.85, 0.90].

We used a Q test of homogeneity to assess the degree of dispersion of reliability estimates around the six internal consistency reliability coefficients (i.e., Cronbach's alpha and split-half reliability coefficients). This statistic indicates whether there is sufficient evidence to reject the null hypothesis that all of the studies share a common effect size. The I^2 statistic was also calculated, which indicates the extent to which the observed heterogeneity in reliability estimates is due to true variability across samples. The Q value for the internal consistency summary reliability coefficient was 18.36, $p = 0.003$, and the I^2 value was 72.77. The significant Q statistic and high I^2 coefficient indicates that the PSDS has noteworthy variance across samples, so it is appropriate to examine if any of the coded sample and study characteristics account for this variability.

Analysis of Moderator Variables

We conducted a series of meta-regression analyses (mixed effects model, method of moments) to examine whether sample and study characteristics could predict the summary reliability coefficient. The following variables were used as moderators: year of publication, number of participants who were female, mean age of participants, and sample size. None were significant moderators. Five of the six studies reported the percent of participants with a doctoral degree; a meta-regression with this as a moderator variable was nonsignificant. Four of the six studies reported participants' mean years supervising, but an analysis examining this variable as a moderator was also nonsignificant.

A subgroup analysis was conducted in order to assess whether having an author of the PSDS scale (coded dichotomously) as an author on the study could explain any variance. This analysis was conducted twice: once in which we did not assume there was common study variance (i.e., tau-squared was not pooled) and a second time assuming common study variance (i.e., using tau-squared pooled). Results of both analyses were not significant, though it should be noted that only two studies included a PSDS author and thus the test cannot be considered to have been particularly strong.

Finally, a subgroup analysis was conducted to assess whether the profession of study participants (coded dichotomously as whether or not professional psychologists were the majority of study participants) could explain any variance in the summary reliability coefficient. A mixed effects analysis (tau-squared not pooled) found that the point estimate for the 4 studies where the majority of participants were not psychologists was $r = 0.94$, $p < 0.001$, 95% CI [0.92, 0.95], whereas the point estimate for the 2 studies

where the majority of the participants were psychologists was $r = 0.90$, $p < 0.001$, 95% CI [0.89, 0.92]. The two sets of studies were found to be heterogeneous, with a Q statistic of 6.01, $p = 0.014$. The results remained significant when tau-squared was pooled, though the Q statistic dropped to 4.79, $p = 0.029$. This drop is predictable because study-to-study variance (tau-squared) is assumed to be the same for all subgroups, thus the study groups are assumed to be less heterogeneous. However, both analyses remained significant, indicating that the discipline of participants explained some variance of the reliability coefficients across this sample of studies.

Validity Generalization: Experience

Of the 10 studies in the final sample, 7 studies (Barnes, 2002; Bencivenne, 1999; Culbreth & Cooper, 2008; Lee et al., 2009; Pelling, 2008; Vidlak, 2002; Watkins et al., 1995) included data on supervision experience and PSDS scores. One study (Pelling, 2008) reported data on experience, but it was not useable in the VG because it did not contain the information required to calculate an effect size. Table 3 presents the details of the studies included in the VG analyses.

Based on a random effects analysis of the 6 studies, the summary validity coefficient was $g = 0.69$, $p = 0.013$, 95% CI [0.15, 1.23]. Data from the Culbreth and Cooper study (2008) were based on a multiple regression, with supervision experience entered as the last variable in the regression. Because no zero order correlations involving supervision experience were reported, we used the total R^2 value of 0.56 from the regression analysis to derive the estimated r value of 0.75. As this value is likely to overestimate the correlation between supervision experience and PSDS scores, for subsequent analyses each result will be reported both with and without this study's

estimate. The summary validity coefficient when the Culbreth and Cooper (2008) data were excluded was $g = 0.40$, $p = 0.001$, 95% CI [0.17, 0.62]. The Q test of homogeneity was used to assess the degree of dispersion of validity coefficients. The Q test value for the six studies was 71.56, $p < 0.001$, and $I^2 = 93.01$. When based on five studies (i.e. excluding Culbreth & Cooper, 2008), it was 8.59, $p = 0.07$, and $I^2 = 53.43$

Analysis of Moderator Variables

As with the reliability coefficient analysis, we conducted a series of meta-regression analyses (mixed effects model, method of moments) to examine whether the following variables could predict the summary validity coefficients that included Culbreth and Cooper (2008): year of publication, number of participants who were female, mean age of participants, sample size, percentage of participants with doctoral degrees, and mean number years supervising. There were no significant results for these analyses. Subgroup analyses were run to assess for the impact of having a PSDS author involved with the study. This was the case for only one of the six studies, and the moderator was not significant both when tau-squared was pooled and when it was not, and whether or not data from Culbreth and Cooper (2008) were included. A subgroup analysis was run to assess whether having a majority of participants being professional psychologists could explain any variance. When the six studies were examined this moderator was not significant, both when tau-squared was and was not pooled.

Validity Generalization: Training

Of the 11 studies in the final sample, 4 studies (Baker et al., 2002; Bencivenne, 1999; Lyon et al., 2008; Vidlak, 2002) reported data regarding supervision training on PSDS scores. Table 3 provides details of the included studies. For one study

(Bencivenne, 1999), the number of participants in each training group (18-59 hours training or 60-16 hours of training) was not reported, so the total study sample was divided by two to estimate the sample size for each group. Baker and colleagues (2002) reported multiple PSDS scores across time points of a training intervention, but only the values from the end of the training period are used for the current analysis. Based on a random effects analysis of these studies, the summary validity coefficient was $g = 1.13$, $p < 0.001$, 95% CI [0.53, 1.73]. The Q test of homogeneity was used to assess the degree of dispersion of validity coefficients related to training. The Q test value for the four studies was 20.94, $p < 0.001$, and $I^2 = 85.68$.

Analysis of Moderator Variables

The following were nonsignificant moderators in mixed effects (method of moments) meta-regressions: year of publication, number of female participants, and sample size. Participants' mean age was a significant moderator for the training and PSDS association, $b = -0.06$, $z = -3.07$, $p = 0.002$, 95% CI [-0.09, -0.02]. The percentage of participants with doctoral degrees was also a significant moderator, $b = -0.01$, $z = -3.30$, $p = 0.001$, 95% CI [-0.02, -0.00]. Two of four studies reported mean years supervising, and when the homogeneity analysis was run with these two studies, the Q statistic and I^2 were not significant, thus no moderator analysis was run. The subgroup analysis examining the possible role of having a PSDS author as a study author was not applicable to these four studies. Finally, a subgroup analysis examining the role of discipline (majority vs. minority professional psychology) was nonsignificant, both when tau-squared was pooled and when it was not.

Discussion

The purpose of the present study was to summarize the typical score reliability and validity values obtained in research using the PSDS, which is to our knowledge the only theoretically based measure of psychotherapy supervisor development. Based on our analyses, it is clear that, when used appropriately, the PSDS is able to provide reliable and valid data.

The majority of authors of studies included in our meta-analysis relied on reliability coefficients reported in the initial article describing the scale's development (Watkins et al., 1995) to provide evidence of the reliability of their PSDS scores. In order for researchers to use previous studies' reliability estimates, they must ensure that the sample characteristics are comparable (Vacha-Haase, Kogan, & Thompson, 2000). This is because, as Therrien and Hunsley (2012) noted, every sample has unique characteristics that are likely to translate into different scores on the measure and every sample of scores may yield a reliability coefficient that differs from other samples. Watkins and colleagues (1995) initially developed the PSDS on a sample of supervisors with a mean of 16.1 years supervision experience, and 98% of that sample had doctoral degrees. In contrast, of the five studies that did not report their own internal consistency values, the samples had 9.6 years mean supervision experience (based on the four studies that reported this information) and 64% of participants had doctorates (based on the five studies that reported this information). Clearly, the original PSDS sample (Watkins et al., 1995) appears—at least as far as these variables are concerned—to be different in some potentially important ways, especially as the SCM model posits that supervision experience impacts a supervisor's developmental level. Simply put, authors need to

evaluate the reliability of data obtained from their samples in order to be certain that the data have at least an acceptable level of reliability.

Based on data from six studies ($n = 1,163$), the PSDS had a mean internal consistency reliability (split-half, Cronbach's alpha) coefficient of 0.93. According to Hunsley and Mash (2008), internal consistency coefficients of $\geq .90$ are considered to be excellent. Further, because both the lower and upper bound confidence interval of the PSDS also were $\geq .90$, we can conclude that the PSDS showed excellent mean reliability across the samples for which reliability coefficients were reported. However, the scores in this RG also indicated noteworthy variance across samples, which further highlights possible concerns about assuming reliability coefficients derived from one study's sample provide accurate estimates for other studies. The current analysis found that participants' discipline (i.e., whether or not the majority of the sample was professional psychologists) was a significant predictor of variance across PSDS scores. More specifically, there was a higher effect size ($r = 0.94$) when the majority of study participants were not from psychology than when they were ($r = 0.90$). It is interesting (and potentially reassuring for researchers in this area) that a measure originally developed with psychologists and designed explicitly to assess psychotherapy supervisor development may provide reliable data for other supervisors in other professions. Researchers can now make an evidence-informed decision in choosing to use the PSDS, in that the evidence would suggest that the PSDS is a measure that has demonstrated reliability for psychotherapy supervisors, as well as those in related fields of counseling and counseling education.

The current study also summarized available data regarding the convergent validity of the PSDS. We found that the overall effect sizes were substantial for both

training ($g = 1.13$) and experience ($g = 0.40$), based on sample sizes of $N = 448$ and $N = 948$, respectively. These results suggest that the PSDS is able to detect important differences in supervisory training and supervisory experience, though possibly less so for experience. This, in turn, means that key elements underpinning the SCM are accurately reflected within the PSDS. At a practical level, the meta-analytic findings provide some indications that receiving training in supervision and having supervision experience are important for supervisor development, at least as measured by PSDS. However, further research is needed in this area, especially involving controlled studies comparing these groups.

Moderator analyses regarding training and PSDS yielded inverse relations for mean age and for the percentage of participants with doctoral degrees. That is, training accounted for more variance in PSDS scores for younger participants than it did for older participants. Similarly, the higher proportion of participants with doctoral degrees, the less variance of PSDS scores was accounted for by training. There are several ways to interpret these results. If one assumes that older supervisors have been providing supervision longer than have the younger supervisors, these results suggest a waning impact on PSDS scores due to supervision training. Relatedly, it may be that the more training younger supervisors have received, the more similar they are to older supervisors regarding their self-assessed developmental level. As for the moderating effect of doctoral degree, it is possible that those with more extensive and specialized training (as is the case in of doctoral psychology programs) develop professionally across a number of competency domains, including supervision. For these professionals, receiving specific training in supervision may be less important to their supervisory development than it is

for professionals with less extensive educational backgrounds. This does not necessarily diminish the need for all supervisors to receive training but, rather, underscores the importance of this training for a specific set of potential supervisors.

As is the case for all meta-analyses, the main limitation of the current study involves the challenge of identifying and retrieving all of the studies that used the PSDS. To address this limit, we purposefully conducted a very extensive search of electronic databases, a hand search of specific journals, contacted the author of the PSDS, included studies using various samples and designs, and included both published and unpublished research. As a result, 13 studies were located (including four unpublished dissertations), but not all of these could be included in our analyses due to the limited information that was reported in some studies. Even with our extensive search, the analyses are based on a small number of studies. Thus, readers should keep in mind this will affect the generalizability of the findings, especially with respect to the moderator analyses. However, it should also be noted that the overall effect sizes reported are likely more generalizable than results from any one of the individual studies included in the meta-analysis, because the overall effect sizes are based on a much larger sample.

Another limitation of the current study is that the extent of supervisory experience was defined differently across studies. For example, Watkins categorized low, medium, and high experience supervisors as having ≤ 8 years, 8-24 years and ≥ 25 years respectively, whereas Bencivenne's (1999) categories were 1-6 years, 8-13 years, and 15-22 years, and Barnes' experience groups had ≤ 5 years or ≥ 5 years. It is common in many research areas for investigators to categorize continuous data. However, this strategy

reduces the statistical power of analyses and makes it very difficult to compare findings across studies (MacCallum, Zhang, Preacher, & Rucker, 2002; Streiner, 2002).

The current study categorized supervision experience as having provided supervision to a supervisee, post-licensure. Thus, supervision experience gained while participants were students or interns was categorized as training in the current analysis. This distinction is debatable, but aims to capture the fact that education and training in supervision, and indeed in professional psychology in general, includes hands on training. For instance, Lyon et al. (2008) noted that participants had provided a median of 40 hours supervision prior to commencing their internships. However, this was not included as supervisory experience in the current analysis because these participants were psychology interns and thus would have had supervision of their supervision work.

Implications and Recommendations

A key recommendation for researchers stemming from our meta-analyses is the need to report full study data in the methods and results sections of articles, including internal consistency values for all measures used, means and standard deviations for all variables, the correlation matrix of variables used in regression analyses, and sample/group sizes for any subgroup analyses. Three studies identified in our search could not be included in various analyses because they did not report this information. To address the difficulty of varying categorizations of supervision experience and training, when continuous data are available, researchers should avoid arbitrarily categorizing these data in order to use ANOVA statistics. Instead, researchers are strongly encouraged to conduct analyses that use continuous data, such as regression analyses. It would be helpful for future research employing the PSDS to report both the overall score, as well

as the sub-scores for the four factor structure (supervisor competence and effectiveness; commitment to supervision and development of a supervisory identity; self-awareness; and, sincerity in the role of the supervisor) as noted in previous investigations of the factor structure of PSDS (Barnes & Moon, 2006).

The results of our meta-analyses indicate that the PSDS is a psychometrically sound measure of supervisory development. Even though the PSDS is the only measure designed to assess this construct, for those involved in providing training in supervision and for those involved in the study of supervision, these results provide some assurance that, when used appropriately, the PSDS is likely to provide reliable and valid data. However, as we have noted, there is still only a limited amount of research on the PSDS. As the research base grows, it is possible that the meta-analytic estimates we derived may change. Clearly there is a need for more research with the PSDS, and more research on the construct of supervisory development more broadly. For psychologists involved in professional training, for example, it would be particularly helpful to have more research that addressed the sensitivity of the instrument to changes associated with training in supervisory skills.

References

References marked with an asterisk () indicate studies included in the final sample.*

- American Psychological Association. (2006). APA taskforce on the assessment of competence in professional psychology (Final Report). Retrieved from <http://www.apa.org/ed/resources/competency-revised.pdf>
- *Baker, S. B., Exum, H. A., & Tyler, R. E. (2002). The developmental process of clinical supervisors in training: An investigation of the supervisor complexity model. *Counselor Education and Supervision, 42*, 15-30.
- Barker, K. K., & Hunsley, J. D. (2013). The use of theoretical models in psychology supervisor development research from 1994 to 2010: A systematic review. *Canadian Psychology, 54*, 176-185. doi: <http://dx.doi.org/10.1037/a0029694>
- Barnes, K. L. (2002). *Development and initial validation of a measure of counselor supervisor self-efficacy*. (Ph.D. 3045772), Syracuse University, United States -- New York. Retrieved from <https://login.proxy.bib.uottawa.ca/login?url=http://search.proquest.com/docview/305471531?accountid=14701>
- [http://sfx.scholarsportal.info/ottawa?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+%28PQDT%29&atitle=&title=Development+and+initial+validation+of+a+measure+of+counselor+supervisor+self-efficacy&issn=&date=2002-01-01&volume=&issue=&spage=&au=Barnes%2C+Kristin+Lee&isbn=9780493597423&jtitle=&btittle=ProQuest+Dissertations+%26+Theses+\(PQDT\)+database](http://sfx.scholarsportal.info/ottawa?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+%28PQDT%29&atitle=&title=Development+and+initial+validation+of+a+measure+of+counselor+supervisor+self-efficacy&issn=&date=2002-01-01&volume=&issue=&spage=&au=Barnes%2C+Kristin+Lee&isbn=9780493597423&jtitle=&btittle=ProQuest+Dissertations+%26+Theses+(PQDT)+database)

- *Barnes, K. L., & Moon, S. M. (2006). Factor structure of the Psychotherapy Supervisor Development Scale. *Measurement and Evaluation in Counseling and Development, 39*, 130-140.
- *Bencivenne, J. C. (1999). *An investigation of professional supervisor and supervisee development. (marriage and family therapy)*. Bencivenne, Joseph Charles: Seton Hall U, Coll Education And Human Services, US. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc3&AN=1999-95022-325> Available from Ovid Technologies PsycINFO database.
- Borenstein, M. (2009). *Introduction to meta-analysis*. Chichester, U.K.: John Wiley & Sons.
- Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive Meta-Analysis Version 2*. Englewood, NJ: Biostat.
- Crook-Lyon, R. E., Presnell, J., Silva, L., Suyama, M., & Stickney, J. (2011). Emergent supervisors: Comparing counseling center and non-counseling center interns' supervisory training experiences. *Journal of College Counseling, 14*, 34-49.
- *Culbreth, J. B., & Cooper, J. B. (2008). Factors impacting the development of substance abuse counseling supervisors. *Journal of Addictions & Offender Counseling, 29*, 22-35.
- Dewey, D., & Hunsley, J. D. (1990). The effects of marital adjustment and spouse involvement on the behavioral treatment of agoraphobia: A meta-analytic review. *Anxiety Research, 2*, 69-83.

- Ellis, M. V., Ladany, N., Krenzel, M., & Schult, D. (1996). Clinical supervision research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology, 43*, 35-50. doi: 10.1037/0022-0167.43.1.35
- Falender, C. A., Cornish, J. A., Goodyear, R., Hatcher, R., Kaslow, N. J., Leventhal, G., . . . Grus, C. (2004). Defining competencies in psychology supervision: A consensus statement. *Journal of Clinical Psychology, 60*, 771-785. doi: 10.1002/jclp.20013
- Goodyear, R. K., Bunch, K., & Claiborn, C. D. (2005). Current supervision scholarship in psychology: A five year review. *Clinical Supervisor, 24*, 137-147. doi: 10.1300/J001v24n01_07
- Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment*. New York: Routledge.
- Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting "reliability generalization" studies. *Measurement and Evaluation in Counseling and Development, 35*, 113-127.
- Hess, A. K. (1986). Growth in supervision: Stages of supervisee and supervisor development. *Clinical Supervisor, 4*, 51-67.
- Hess, A. K. (1987). Psychotherapy supervision: Stages, Buber, and a theory of relationship. *Professional Psychology: Research and Practice, 18*, 251-259.
- *Hillman, S. L., McPherson, R. H., Swank, P. R., & Watkins, C. E., Jr. (1998). Further validation of the psychotherapy supervisor development scale. *The Clinical Supervisor, 17*, 17 - 32. doi: 10.1300/J001v17n01_02

Hunsley, J. D., & Mash, E. J. (2008). Developing criteria for evidence based assessment:

An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3-16). Oxford: Oxford University Press.

*Lee, H. K., Veach, P. M., & LeRoy, B. S. (2009). An investigation of relationships among genetic counselors' supervision skills and multicultural counseling competence. *Journal of Genetic Counseling, 18*, 287-299. doi:

<http://dx.doi.org/10.1007/s10897-009-9219-y>

*Lyon, R. C., Heppler, A., Leavitt, L., & Fisher, L. (2008). Supervisory training experiences and overall supervisory development in predoctoral interns. *Clinical Supervisor, 27*, 268-284. doi: 10.1080/07325220802490877

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.

Milne, D. L., Reiser, R., Aylott, H., Dunkerley, C., Fitzpatrick, H., & Wharton, S. (2010). The systematic review as an empirical approach to improving CBT supervision. *International Journal of Cognitive Therapy, 3*, 278-294. doi:

10.1521/ijct.2010.3.3.278

Parker, K. C., Hanson, R., & Hunsley, J. D. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin, 103*, 367-373. doi: 10.1037/0033-2909.103.3.367

*Pelling, N. (2008). The relationship of supervisory experience, counseling experience, and training in supervision to supervisory identity development. *International Journal for the Advancement of Counselling, 30*, 235-248.

- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences : a practical guide*. Malden, MA ; Oxford: Blackwell Pub.
- Rodenhauser, P. (1994). Toward a multidimensional model for psychotherapy supervision based on developmental stages. *Journal of Psychotherapy Practice and Research* 3, 1-15.
- Rodenhauser, P. (1997). Psychotherapy supervision: Prerequisites and problems in the process. In C. E. J. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 527-548). New York John Wiley & Sons.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Science*, 5, 127-134.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529-540. doi: 10.1037/0021-9010.62.5.529
- Stoltenberg, C. D., & Delworth, U. (1987). *Supervising counselors and therapists*. San Francisco, CA: Jossey-Bass.
- Stoltenberg, C. D., McNeill, B., & Delworth, U. (1998). *IDM supervision: An integrated developmental model for supervising counselors and therapists*. San Francisco, CA: Jossey-Bass.
- Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry*, 47, 262-266.
- Therrien, Z., & Hunsley, J. (2012). Assessment of anxiety in older adults: a systematic review of commonly used measures. [Review]. *Aging Ment Health*, 16(1), 1-16. doi: 10.1080/13607863.2011.602960

- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*, 6-20. doi: 10.1177/0013164498058001002
- Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample composition and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60*, 509-552.
- *Vidlak, N. W. (2002). *Identifying important factors in supervisor development: An examination of supervisor experience, training, and attributes*. Vidlak, Nicole White: U Nebraska - Lincoln, US. ProQuest Dissertations and Theses database. (727413951)
- Vieceli, A. E. (2007). *A comparison of supervisor development and supervision emphasis of academic and on-site supervisors*. Vieceli, Ada E.: Southern Illinois U Carbondale, US. ProQuest Dissertations and Theses database. (1212791611)
- Watkins, C. E., Jr. (1990). Development of the psychotherapy supervisor. *Psychotherapy: Theory, Research, Practice, Training, 27*, 553-560. doi: 10.1037/0033-3204.27.4.553
- Watkins, C. E., Jr. (1993). Development of the psychotherapy supervisor: Concepts, assumptions, and hypotheses of the supervisor complexity model. *American Journal of Psychotherapy, 47*, 58-74.
- Watkins, C. E., Jr. (1994). The supervision of psychotherapy supervisor trainees. *American Journal of Psychotherapy, 48*, 417-431.
- Watkins, C. E., Jr. (1995). Psychotherapy supervisor development: On musings, models, and metaphor. *Journal of Psychotherapy Practice & Research, 4*, 150-158.

- Watkins, C. E., Jr. (2010). *Development of the Psychotherapy Supervisor: The SCM—Twenty Years Later*. Unpublished manuscript.
- Watkins, C. E., Jr. (2012). Development of the psychotherapy supervisor: Review of and reflections on 30 Years of theory and research. *American Journal of Psychotherapy*, 66, 45-83.
- *Watkins, C. E., Jr., Schneider, L. J., Haynes, J., & Nieberding, R. (1995). Measuring psychotherapy supervisor development: An initial effort at scale development and validation. *The Clinical Supervisor*, 13, 77-90. doi: 10.1300/J001v13n01_06
- Wilson, D. B. (2013). Practical meta-analysis effect size calculator Retrieved June 24, 2013, from <http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>

Table 1

Study details (publication status, sample size, PSDS mean and standard deviation)

Study	Study type and publication status	<i>n</i>	PSDS scores	
			Mean	SD
(Watkins et al., 1995)	Peer reviewed journal	335	108.8	8.3
(Hillman et al., 1998)	Peer reviewed journal	43	103.9	11.2
(Bencivenne, 1999)	Unpublished dissertation	136	110.0	8.9
(Baker et al., 2002)	Peer reviewed journal	19	84.2	14.2
(Barnes, 2002)	Unpublished dissertation	255	107.0	11.9
(Vidlak, 2002)	Unpublished dissertation	99	103.3	8.0
(Culbreth & Cooper, 2008)	Peer reviewed journal	192	-	-
(Lyon et al., 2008)	Peer reviewed journal	233	90.6	12.5
(Pelling, 2008)	Peer reviewed journal	175	108.0	12.2
(Lee et al., 2009)	Peer reviewed journal	122	101.9	10.8
Overall		1649	102	10.9

Table 2

Meta-analysis of internal consistency (Cronbach's alpha and split-half) values of PSDS

Study	<i>n</i>	Statistics for each study			
		Correlation	Lower Limit ^a	Upper Limit ^a	Z-Value
(Watkins et al., 1995)	335	0.90	0.88	0.92	26.83*
		Cronbach			
(Barnes, 2002)	255	0.93	0.91	0.93	26.33*
		Cronbach			
(Lyon et al., 2008)	233	0.91	0.89	0.93	23.17*
		Cronbach			
(Pelling, 2008)	175	0.95	0.93	0.96	22.66*
		Cronbach			
(Lee et al., 2009)	122	0.91	0.87	0.94	16.66*
		Cronbach			
Summary Effect Size – Cronbach's Alpha	1120	0.92*	0.90	0.94	25.70*
(Hillman et al., 1998)	43	0.95	0.91	0.97	11.44*
		Split Half			
Summary Effect Size – Cronbach's Alpha and Split Half	1163	0.93*	0.91	0.94	26.79*

a. 95% confidence interval

* $p < 0.001$

Table 3

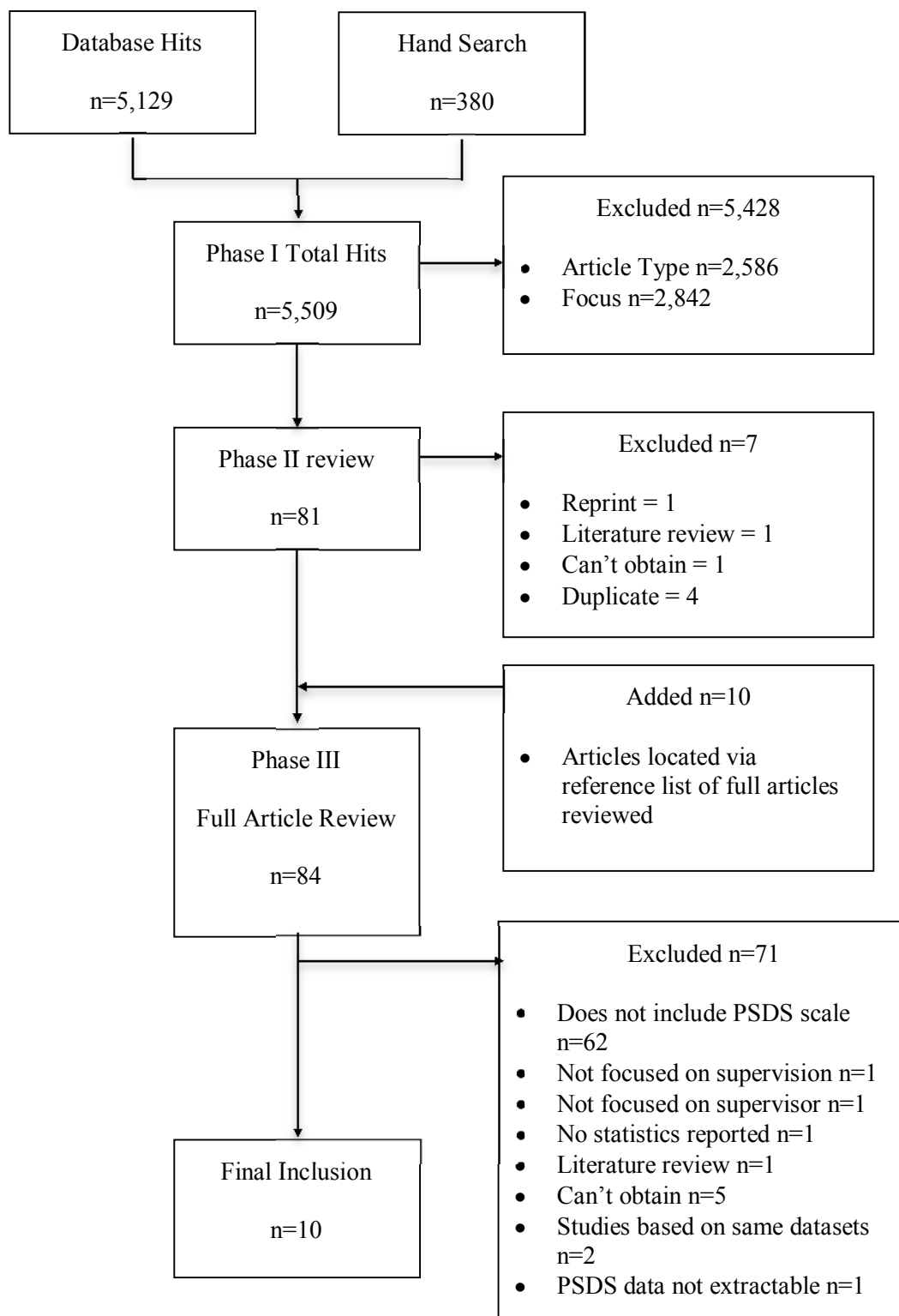
Meta-analysis of validity coefficients for the PSDS

Validity Construct	Study	<i>n</i>		Statistics for each study				
		Less experience/training ^a	More Experience/training ^a	Effect size Hedge's <i>g</i>	Lower Limit ^b	Upper Limit ^b	Z-value	p-value
Experience	(Watkins et al., 1995)	57	262	0.50	0.21	0.78	3.38	0.001
Experience	(Bencivenne, 1999)	41	82	0.06	-0.31	0.42	0.29	0.772
Experience	(Barnes, 2002)	228	57	0.51	0.22	0.80	3.410	0.001
Experience	(Vidlak, 2002)	99	-	0.14	-0.26	0.54	0.71	0.481
Experience	(Lee et al., 2009)	122	-	0.72	0.34	1.10	3.71	0.000
Experience	(Culbreth & Cooper, 2008)	192	-	2.25	1.82	2.68	10.29	0.000
Summary Effect Size – Experience			1140	0.69	0.15	1.23	2.48	0.013
Training	(Bencivenne, 1999)	62	61	0.62	0.27	0.98	3.42	0.001
Training	(Baker et al., 2002)	7	12	2.07	1.04	3.10	3.95	0.000
Training	(Vidlak, 2002)	27	49	0.65	0.20	1.09	2.86	0.004
Training	(Lyon et al., 2008)	230	-	1.54	1.21	1.86	9.19	0.000
Summary Effect Size - Training			448	1.13	0.53	1.73	3.69	0.000

a. If the analysis was correlational, all participants are noted in this table as “less experience/training”

b. 95% confidence interval

Figure 1: Study Flow Diagram



The Impact of Supervisory Experience and Training on Supervisory Outcomes:

A Meta-Analytic Evaluation

Keegan Barker

Abstract

The current meta-analysis was designed to assess the impact of training in clinical supervision and supervisory experience on general supervisory development and self-efficacy in clinical supervisors from various disciplines. Literature searches yielded 23 articles that met inclusion criteria. The effect size on development/self-efficacy from 12 studies that examined supervisory experience was $d = 0.68$, $p < 0.0001$, 95% CI [0.31, 1.06]; from 19 studies that examined the impact of supervisory training on development/self-efficacy, the effect size was $d = 0.69$, $p < 0.0001$, 95% CI [0.39, 1.00]. These two effect sizes did not significantly differ from each other. For both sets of analyses, we examined the variance in scores across studies, and considered sample and study characteristics that were possible moderators of the main results. Based on our findings, we make recommendations for future research, and for trainee and professional development initiatives, including the importance of valuing both training and experience in clinical supervision, and how inclusion of both training and experience fit with a model of competency-based education.

Keywords: professional supervision, practicum supervision, meta-analysis

The Impact of Supervisory Experience and Training on Supervisory Outcomes: A Meta-Analytic Evaluation

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). For professional psychologists in training, clinical supervision is the main teaching method by which they learn to deliver patient care. Participants in the 2012 Association of Psychology Postdoctoral and Internship Centers (APPIC) pre-doctoral internship match reported receiving a median of 312 clinical supervision hours in their doctoral training (APPIC, 2012). Thus, clinical supervision is clearly an integral part of the training process and is increasingly recognized as an important competency for professional psychologists in the United States (American Psychological Association, 2006), Canada (see <http://www.cpa.ca/docs/file/MRA.pdf>), and the United Kingdom (Milne et al., 2010).

Despite its obvious importance, the empirical study of clinical supervision is in its early stages. Discussing the development of supervisory skills, Bernard and Goodyear (Bernard & Goodyear, 1998) concluded that, historically, it has been largely believed that experience as a supervisee and then as a clinician is sufficient to become a supervisor. This belief does appear to be shifting, as reflected in (a) the current emphasis on training requirements in supervision among accredited professional programs (American Psychological Association, 2006; Canadian Psychological Association, 2002) and (b) the growing focus on a competency-based approach to supervision (Falender & Shafranske, 2004).

These recent trends notwithstanding, it is striking that we know little about what constitutes effective clinical supervision, nor do we yet understand much about the impact clinical supervision has on trainee knowledge, attitudes, and skills, or what impact clinical supervision has on client outcomes (Wheeler & Richards, 2007). Bearing this in mind, it is hardly surprising that we also do not know how best to teach trainees to become effective supervisors so that they too can impart effective clinical supervisory practice to their future trainees. This poses a significant challenge for psychology educators, for they are encouraged and increasingly required to train for supervision competency, but there exists negligible evidence in the literature to guide their training decisions (Barker & Hunsley, 2013). As such, educators must currently rely on panel consensus statements, rather than extensive, replicated research findings, to develop training goals and programs (Barker & Hunsley, 2013).

The case for competence-based clinical supervision is compelling, with its emphasis on assisting trainees to translate their acquired knowledge and skills into real world application, and to evaluate trainees using performance outcomes (Falender & Shafranske, 2004). The field is at an important juncture: there is a movement to provide training in clinical supervision, but we do not know the effective ingredients for such programs. As the move to educate trainees for supervisory competence intersects with the historical belief that experience as a counsellor/therapist and as a supervisee is sufficient to become a supervisor (Bernard & Goodyear, 1998, 2014), the relative merits of experience versus training in the development of competent clinical supervisors are still frequently discussed in the research literature. A systematic review of the research literature on supervisor development found that over one-third of the research in this area

focused on the influences of past supervisory experience and training on current supervision practices (Barker & Hunsley, 2013).

A number of issues emerge from the research on the impact of experience and training on supervisory outcomes. First, there are mixed findings with respect to the relation between supervisory experience and supervisory outcomes. Watkins and colleagues (Watkins, Schneider, Haynes, & Nieberding, 1995) found that experience was the single best predictor of supervisory development, explaining 8.2% of the variability in scores on the Psychotherapy Supervisor Development Scale (Watkins et al., 1995). Likewise, Stevens (1998) reported significant associations between experience and increased supervisor self-efficacy when those with five or more years of experience were compared to those with zero to two years of experience (Stevens, Goodyear, & Robertson, 1998). However, a number of studies have failed to find significant relations between years of experience as a supervisor and supervisor development (Lyon, Heppler, Leavitt, & Fisher, 2008; Vidlak, 2002), supervisory working alliance (Rodolfa et al., 1998), or supervisor stance/emphasis during supervision (Stevens et al., 1998). No immediately discernable patterns are apparent that can explain why some studies found significant associations whereas others did not. Thus, a quantitative analysis of this literature is warranted to determine if variation in findings is due to the effect of possible moderators on the study outcomes (e.g., sample characteristics, measures used).

Second, the results of research linking training in supervision and supervisory outcomes are also varied. Several studies have found supervision training to be linked to higher supervisory self-efficacy and supervisor development (Baker, Exum, & Tyler, 2002; Borders, 1996; Johnson & Stewart, 2008; Lyon et al., 2008; McMahon & Simons,

2004; Stevens et al., 1998; Vidlak, 2002), although one study found this link only for those with no prior supervision experience (Haley, 2002). When the literature is examined more closely in order to better understand the variability in the strength of associations reported in the literature, a number of variables emerge as possible moderators: (a) the specific type, or format, of the training (Borders, 1996; Borders & Fong, 1994; Haley, 2002; McMahon & Simons, 2004; Pelling, 2008), (b) the number of training experiences (Lyon et al., 2008; Stevens et al., 1998; Vidlak, 2002), (c) the temporal nature of the training, that is, whether it is brief or sustained (Kavanagh et al., 2008; Vidlak, 2002), and (d) the influence of methodological characteristics of the study, such as variations in how constructs are assessed and sample sizes. It would be beneficial to examine the extent to which these training factors result in differences in supervisory outcomes--that is, to examine their influence as moderators in the relationship between supervisory training and experience on supervisory development and self-efficacy (see Hypotheses, below).

The Present Study

In the current study, meta-analysis and meta-regression were used to examine the effects of supervision experience and training on supervisory outcomes and to compare the two overall effect sizes. A meta-analysis is warranted in this particular area for several reasons. First, meta-analysis is an effective tool to assess the consistency of effect sizes across studies, regardless of whether each of the studies' findings in the analysis is significant. It also allows for the application of formulae to partition the variance in order to quantify "true" differences among studies. This then allows us to consider the implications of this variance (such as the influence of covariates/moderators in meta-

regression analyses) and also to distinguish between “real” and “spurious” dispersion (Borenstein, 2009). As noted above in the literature review above, there are various potential moderators (e.g., type/format of training, number of training exposures, nature of training) that may account for variation in study findings (for details, see Hypotheses, below). In addition, study and sample characteristics (age, gender, etc) were examined as possible moderators because these variables can have an impact on study psychometrics and thus impact study findings. Finally, the statistical procedures of meta-analysis and meta-regression can assess the contribution of experience and training in supervision to supervisor outcomes. These analyses could provide valuable information about the relative contributions of experience and training to educators, directors of clinical training, and regulatory bodies. This, in turn, may allow conclusions to be drawn about the extent to which experience and training should be differentially weighted in the development of clinical supervisors, which would provide invaluable guidance to psychology educators in their training efforts.

Hypotheses

Based on the literature reviewed above, it is hypothesized that:

1. Training experiences will have a greater effect on supervisory outcomes than will supervisory experience. This is hypothesized because, overall, studies (noted above) examining training have found a more consistently positive relation between the variables than the mixed picture of the findings regarding experience and supervisory outcomes.
2. Prolonged training will have a larger effect on supervisory outcomes, when compared with concentrated training interventions. This hypothesis is in keeping

- with educational findings in other areas of health care, for example, a meta-analysis of continuing medical education (CME; conferences, courses, rounds, meetings, symposia and lectures), which found that overall, no significant effect of brief educational methods could be detected (Davis et al., 1999)
3. Larger effect sizes will be associated with training that includes an applied component (practicums and/or supervision of supervision; SOS), as compared to classroom based/didactic training. The largest effect sizes will be associated with: combination of didactic-experiential training, and also practicum training in supervision with supervision of supervision (SOS). This hypothesis is in keeping with meta analytic evidence that interactive CME sessions with applied components can effect change in professional practice (Davis et al., 1999)

Method

Literature Search

The search strategy for a recent systematic review on supervisor development by Barker and Hunsley (2013) was used as the basis for the current search strategy. Briefly, this involved developing a comprehensive search strategy for PsycINFO using the subject headings of supervision and development/training. Related search terms associated with the constructs, as defined by PsycINFO's subject headings, were recorded until no new applicable search terms were identified. In addition to the subject headings, applicable keywords were developed. The search strategy was then used as a template for the development of additional searches in the MedLine, CINAHL and ERIC databases, thereby individualizing the search for each database. This general search strategy was used to identify potentially relevant studies (including both published studies and

unpublished dissertations) available as of the end of 2012. No discipline limits were included in the search so that any study in this area could be identified.

The abstracts identified in the search were reviewed in various phases. In Phase I, the first author reviewed abstracts to exclude theoretical papers, book reviews, and letters to the editor. In Phase II, the first author reviewed the retained abstracts to exclude those empirical studies that did not include supervisory training and/or experience on supervisory outcomes. Additionally, a hand search was conducted to augment the electronic search. Key journals' tables of contents 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 *Journal of Counseling Psychology*, *Journal of Clinical Psychology*, *Professional Psychology: Research & Practice*, 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*. Reference lists of included articles and dissertations were searched to identify any relevant publications.

Duplicates and reprints were excluded, as were papers that could not be obtained. Figure 1 is the flow diagram of the study. In summary, 5,901 abstracts were reviewed, with 89 documents meeting our criteria to warrant review of the full document. Upon closer review, two studies (Barnes & Moon, 2006; Crook-Lyon, Presnell, Silva, Suyama, & Stickney, 2011) were based on the same data set as another study (Barnes, 2002; Lyon et al., 2008), and thus the latter studies were retained because they contained more complete information. Of the 89 papers reviewed in full, 66 were excluded for various

reasons (e.g., did not include the targeted supervisory outcome constructs, did not contain necessary statistical information; for full details, see Figure 1).

To be included in the final sample, studies had to be empirical, be focused on clinical supervision of counselling/psychotherapy, and include results related to the impact of training and experience on the outcomes of supervisor development (including supervisory confidence) or efficacy as a supervisor (as rated by self or others). These outcomes were chosen because they were the most frequently occurring outcomes assessed across studies; moreover, they are conceptually related and have been found to be positively correlated (Barnes, 2002; Williams, 2012). If a study contained data related to supervisor development, as measured by the PSDS (Watkins et al., 1995), this was designated the main study outcome, because it was used in the most number of studies, and has demonstrated psychometric properties (Barnes & Moon, 2006; Hillman, McPherson, Swank, & Watkins, 1998; Watkins et al., 1995). Due to the limited number of studies in this area, all study designs were included, as were a wide variety of disciplines related to counselling and psychotherapy (e.g., medicine, nursing, occupational therapy, psychology). Further, both published and unpublished studies were included. The wide inclusion criteria were regarded as appropriate to the study's goal of presenting a general overall picture of the relative contributions of training and experience on supervisory outcomes.

The final sample (indicated in the reference list with an asterisk) was based on 23 studies (including 5 unpublished dissertations), 4 of which were used in the analyses regarding supervisory experience, 11 for the analyses regarding supervisory training, and 8 studies contained information useable in both analyses. Because of the inclusion of

unpublished studies, evaluating the possibility of publication bias affecting results was not necessary. We did not, therefore, conduct analyses to determine the fail safe N or generate funnel plots of effects.

Data Extraction

To describe the included studies and assess possible moderating variables, the following information for each study was coded: year of publication, sample size, participants' mean age, percentage of female participants, percentage of participants with doctoral degrees, percentage of participants with clinical or counseling psychology backgrounds, and percentage of participants who were students. Additionally, information regarding each study's independent/predictor variables was coded: information on participants' experience in providing supervision (mean and standard deviation of years supervising) was extracted, as well as any statistical results pertaining to this experience and supervisory outcomes. For the current meta-analysis, experience was defined as any experience providing supervision, regardless of when this supervision was provided in a supervisor's career path (e.g., providing supervision as a graduate student, or providing supervision post-licensure). However, when supervision was provided with concomitant supervision of supervision (SOS), this was coded as training, because it is considered an experiential training opportunity.

Information pertaining to supervisor training (defined as efforts aimed at improving a participant's knowledge, skills, or attitudes towards supervision) was extracted, and coded as number of hours. In some instances, when there was insufficient information in a study description to generate a quantifiable amount of training, additional information was requested from study authors. If no clarification was received

from study authors, estimates were generated based on the available details (e.g., if the information in the manuscript indicated that training was “2 days” it was estimated this to be a total of 16 hours). When studies reported retrospective data, the mean number of hours training of participants was recorded.

In coding the type of supervision training, the following categories were used: workshop (a lecture, presentation, or classroom based learning over a period of ≤ 2 days/16 hours), didactic course (a lecture, presentation, or classroom based learning over a period of ≥ 2.1 days, 17 hours), practicum (practical training where trainees provide supervision to a supervisee in order to learn about their role as a supervisor), combination (didactic course with a practicum component), combination with SOS (didactic course with a practicum component, and the trainee supervisors met with their own supervisors to review the supervision they provided), SOS (practical training where trainees provide supervision to a supervisee, and trainee supervisors met with their own supervisors to review the supervision they provided). Also coded for each study was information about the outcome variable, that is, the construct assessed, the specific measure used, the number of items in the measure, and the sample’s mean score and standard deviation on the measure.

The majority of variables required “low inference” coding (e.g., mean age, year of publication) in that they involved little or no judgment on the part of the coder (Cooper, 2009). These variables were coded by the first author. The variable with the highest inference was type of supervision training and, therefore, both authors coded this variable for all studies. Cohen’s kappa statistics (Cohen, 1960) was calculated to assess the inter-rater reliability for the type of supervision training. The resulting value of 1.00 is

typically described as indicating excellent reliability (D.V. Cicchetti, 2001; D. V. Cicchetti & Sparrow, 1981) .

When any of the above data were not included in the article, the study author was contacted to obtain this information. Emails requesting information were sent to 23 authors, and 15 responses were received (response rate of 65%). Of these 15 responses, 7 contained requested information, and of these, 5 contained information that was ultimately included in the analyses. Thus, follow up with study authors provided useable information that was incorporated for 22% of the contacts initiated. For the other contacts, the authors did not have the information available or they did not respond to our requests.

Data Analysis

There were three main steps in the analysis. First, the effect sizes for each study were calculated using either the program Comprehensive Meta-Analysis, version 2.0 (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005), or the Campbell Collaboration effect size calculator (Wilson, 2013). Because the two programs can provide calculations for different research designs, this allowed a broad range of research designs to be included.

For correlational statistical results, the following information was entered for each study: r value, sample size, and effect direction. The standard error of r , Fisher's z , and its standard error were calculated for each study. One study (Taylor, Gordon, Grist, & Olding, 2012) reported an omega-squared effect size for its finding. The square root of this value was taken to generate an r value, as suggested by Ferguson (2009). The subsequent computations were carried out using the Fisher's z scale transformed values.

Meta-analyses are not generally performed on the correlation coefficient itself because the variance depends strongly on the correlation (Borenstein, 2009). If the statistical results were not correlational in nature, each study's information (e.g., mean, standard deviation, sample size for intervention and control groups) was used to calculate Cohen's *d*, or standardized mean difference and its associated 95% confidence interval.

In any meta-analysis, the issue of non-independence of results is a concern. As is frequently done in meta-analysis, this was controlled for by allowing each study to contribute only one result (Borenstein, 2009) for the analysis related to training and one result for experience. If a study contained more than one potentially relevant result, a mean effect size across all relevant variables in a study was calculated for that study, thus generating a composite effect size (these are noted in Tables 1 and 2). If a study reported separate results based on training type (didactic, experiential, combination, SOS), separate effect sizes were generated for each, so that training type could be explored in the analysis. A random effects model was used for all summary effect size computations and analyses.

The second step in the analyses was to enter each study's summary data into CMA (Borenstein et al., 2005) in order to calculate overall effect sizes (random effects model) from the 12 studies that examined experience and 19 studies that examined training. In the random effects meta-analysis, each study's mean is weighted, where the weight assigned to each study is the inverse of that study's variance. The summary effect size is then generated from these weighted means. All summary effect sizes are reported as standardized mean difference, or Cohen's *d*. In the third step, the heterogeneity of the overall effect sizes was examined. A *Q* test of homogeneity was used to assess the degree

of dispersion of estimates around the summary effect size. This statistic indicates whether there is sufficient evidence to reject the null hypothesis that all of the studies share a common effect size. The I^2 statistic was also calculated, which indicates the extent to which the observed heterogeneity in estimates is due to true variability across samples. This statistic is expressed in a ratio form, from 0 to 100. As I^2 moves away from 0 it means that some of the variance is nonrandom and can be potentially explained by looking at covariates to explain this diversion. Thus, if Q is significant and a high value for I^2 is obtained, it suggests that there is variance that could be explained by moderators, which we then examined by meta-regression analyses using CMA (random effects model, method of moments). These three steps were carried out for data related to training, and again for data related to experience.

Results

Overview of Studies

The final sample included 23 studies, with a total of 4,930 participants, 63% of whom were female (4 studies did not report participant gender). Mean age of participants was 42.8 years (based on 17 studies). Of the 14 studies that reported information on degrees, 49.8% of participants had doctorates. The participants varied in their professional backgrounds, and included psychologists (Johnson & Stewart, 2000, 2008; Robiner, Saltzman, Hoberman, & Schirvar, 1997; Vidlak, 2002; Watkins et al., 1995), marriage and family therapists (Bencivenne, 1999), doctoral students in counselor education (Baker et al., 2002; Borders, 1996; Borders & Fong, 1994), psychology predoctoral interns (Haley, 2002; Lyon et al., 2008), counseling education/counseling clinical supervisors (Barnes, 2002; Williams, 2012), substance abuse

counselors/addictions counseling supervisors (Culbreth & Cooper, 2008; Laschober, de Tormes Eby, & Sauer, 2012), genetic counseling supervisors (Lee, Veach, & LeRoy, 2009), psychotherapists/mental health professionals (Ronnestad, Orlinsky, Parks, Davis, & The Society for Psychotherapy Research (SPR) Collaborative Research Network, 1997; Stevens et al., 1998), social workers (Gourdine & Baffour, 2004), counsellors (McMahon & Simons, 2004; Taylor et al., 2012), allied health professionals (Kavanagh et al., 2008), and mental health nurses (White & Winstanley, 2010). Mean years of supervision experience was 9.0 across the 12 studies that reported this information.

All of the effect sizes in the experience analyses were based on correlational designs (i.e., retrospective surveys), as were eleven of the studies in the training analyses (Barnes, 2002; Bencivenne, 1999; Haley, 2002; Johnson & Stewart, 2000, 2008; Laschober et al., 2012; Lyon et al., 2008; Robiner et al., 1997; Stevens et al., 1998; Vidlak, 2002; Williams, 2012). A within-subjects, repeated measures design was used in four studies in the training analyses (Borders, 1996; Borders & Fong, 1994; McMahon & Simons, 2004; Taylor et al., 2012). The remaining four studies in the training analyses included a control group. Baker et al. (2002) and Gourdine and Baffour (2004) included an intervention and control group, although participants were not randomized to conditions. Both Kavanagh et al. (2008) and White and Winstanley (2010) used randomized control designs, however in one (White & Winstanley, 2010) the measure of interest for the current meta-analysis was not administered to the control participants. Thus, for the purposes of the current study, the study by White and Winstanley (2010) was considered a within-subjects, repeated measures design.

Experience Analyses

Of the 12 studies reviewed, 7 had used the 18-item Psychotherapy Supervisor Development Scale (Watkins et al., 1995) to assess supervisor development as the main study main outcome. Five studies' outcome was self-rated self-efficacy or confidence as a supervisor, and this was assessed with various measures: Haley (2002) used a 42-item Supervision Self-Efficacy Questionnaire, Ronnestad et al. (1997) used 1-item as part of the Development of Psychotherapists Common Core Questionnaire, Stevens et al. (1998) asked participants 1 question about this construct, Johnson and Stewart (2008) asked supervisors to rate their competence in supervisory roles, and Laschober et al. (2012) asked supervisees to rate their supervisor's efficacy, using a 14-item tool they developed.

Table 1 shows the results of the meta-analysis on supervisory experience and development/efficacy. The summary effect size based on a random effects analysis was $d = 0.68$, $p < 0.0001$, 95% CI [0.31, 1.06]. The Q test value was 253.15, $p < 0.001$, and $I^2 = 95.66$. The significant Q statistic and high I^2 coefficient indicates that there is noteworthy variance across samples, so it was appropriate to examine if any of the coded sample and study characteristics account for this variability.

Analysis of Moderator Variables

A series of meta-regression analyses (mixed effects model, method of moments) was conducted to examine whether coded moderator variables could predict the summary coefficient. The following variables were used as moderators: year of publication (12 studies), total sample size (12 studies), percentage of participants who were female (12 studies), and number of items used to assess the outcome variable (supervisor development, efficacy; 12 studies), mean age of participants (11 studies), the percentage

of participants with a doctoral degree (8 studies), percentage of participants with psychology as discipline (8 studies), percentage of participants who were students (8 studies), and mean years of supervision experience (8 studies). None of the variables was a significant moderator.

Training Analyses

Of the 19 studies reviewed, 6 studies (Baker et al., 2002; Barnes, 2002; Bencivenne, 1999; Lyon et al., 2008; Vidlak, 2002; Williams, 2012) used the 18-item Psychotherapy Supervisor Development Scale (Watkins et al., 1995) to assess supervisor development as the main study main outcome. Thirteen studies' outcome involved self-rated self-efficacy or confidence as a supervisor, and these constructs were assessed with various measures: Borders (1996) and Borders and Fong (1994) used a 36 item Stress Appraisal Scale, Gourdine and Baffour (2004) asked students to rate their supervisors' competency on 4 items, Haley (2002) used a 42-item Supervision Self-Efficacy Questionnaire, Johnson and Stewart (2000) used one item regarding preparedness to supervise, Stevens and colleagues (1998) asked participants one question about this construct, Johnson and Stewart (2008) nine items regarding self perceived competency in three supervisory areas, Kavanagh et al. (2008) developed a 16 item measure of self-efficacy, McMahon and Simons (2004) developed a 38 item Clinical supervision Questionnaire that included items on supervisory confidence, White and Winstanley (2010) used the 36-item Manchester Clinical Supervision Scale to assess supervisor's self rated efficacy in supervision, Newman Taylor and colleagues (2012) developed an 18 item measure of self-rated competence, and Lasochber et al. (2012) asked supervisees to rate their supervisor's efficacy, using a 14 item tool they developed. Information on the

number of items used by Robiner and colleagues (1997) to assess self-rated efficacy was not available.

Table 2 presents the results of the meta-analysis on supervisory training. The summary effect size based on a random effects analysis was $d = 0.69$, $p < 0.0001$, 95% CI [0.39, 1.00]. Of note, three studies' (Gourdine & Baffour, 2004; Kavanagh et al., 2008; Laschober et al., 2012) effect sizes were negative. In the study by Kavanagh et al., (2008) the split condition supervisors showed a drop on self-efficacy over time (these supervisors participated in the training intervention, although their supervisees did not participate until some months later). The authors noted the possibility that training supervisors without providing supervisees with a framework that allows them to benefit from their supervisors' interventions may have had an impact. They recommended that this issue warrants future research, as the issue of "split" conditions has not been investigated in other research to date. In Gourdine and Baffour's (2004) study, the students rated their supervisors on their efficacy in 4 areas of supervision. Overall there was a lack of statistically significant findings in this study, which may be partly attributable to the lack of control regarding other factors (e.g., previous supervision experience). The effect size associated with the study by Laschober et al. (2012) was provided by the study author, and represents the relation between hours of supervisor training that supervisors reported receiving, and their efficacy as rated by their supervisees.

The Q test value for the meta-analysis on training was 156.37, $p < 0.001$, and $I^2 = 88.49$. It was thus appropriate to conduct analyses to examine impact of moderator variables via meta-regressions.

Analysis of Moderator Variables

Meta-regression analyses (mixed effects model, method of moments) were conducted to examine impact of moderator variables: year of publication (19 studies), percentage of participants who were female and percentage of participants with psychology as discipline (15 studies each), number of items used to assess the outcome variable (18 studies), percent participants with doctoral degrees (13 studies), percent participants who were students (16 studies), years of experience supervising (9 studies), length of training initiative (11 studies). None of these was significant. However, sample size was a significant moderator for the association between training and development and self-efficacy, $b = -0.00174$, $z = -1.97$, $p = 0.05$, 95% CI [-0.00347, -0.00001]. Mean age was also a significant moderator, $b = -0.08$, $z = -2.49$, $p = 0.012$, 95% CI [-0.15, -0.018].

To assess the potential impact of the type of training (hypothesis 3), summary effect sizes for each of the training types (see Data Extraction, above) were generated. Significant effect sizes were found for the following training types: SOS $d = 0.39$, $p = 0.03$, 95% CI [0.04, 0.73], combination $d = 0.52$, $p = 0.001$, 95% CI [0.21, 0.83], and combination with SOS $d = 2.65$, $p = 0.003$, 95% CI [0.91, 4.39]. However, the findings regarding heterogeneity between all the training types was not significant, $Q = 7.57$, $p = 0.18$ (see Table 3 for further details).

To assess the potential impact of training being prolonged or concentrated (hypothesis 2), the studies ($N = 8$) that reported the length of the training initiatives were examined by sub group analysis. “Concentrated” studies were those that reported training between seven hours and five days (consecutive), and “prolonged” training were those

initiatives that ran over weeks and months. The summary effect size (random effects analysis) for the five studies that reported “concentrated” training was $d = 0.47$, $p = 0.199$, 95% CI [-0.25, 1.20], and the summary effect size for the three studies that reported “sustained” training was $d = 2.073$, $p = 0.002$, 95% CI [0.79, 3.36]. These two effect sizes were significantly different from each other, $p = 0.034$.

Comparison of Effect Sizes for Training and Experience

To assess whether significant heterogeneity was present between effect sizes based on experience compared to training we examined the associated $Q = 0.000$, $p = 0.97$. Thus, the null hypothesis that the two effect sizes are of the same magnitude cannot be rejected.

Discussion

The purpose of the present study was to summarize the literature examining the impact of supervisory training and experience on supervisor development and efficacy. The results from the meta-analyses indicate that the overall effect sizes were substantial for both training ($d = 0.69$) and experience ($d = 0.68$), and that these effect sizes did not differ significantly from each other. Therefore the first hypothesis that training would have a greater impact than experience was not supported. The current findings are informative, in that previous individual studies have not consistently come to this conclusion, but when the results from across studies are combined, the overall significant, positive effects are clearly evident. Thus, both training and experience are equally important in aiding supervisors in their development and sense of efficacy.

A recent meta-analysis on the PSDS scale (Barker & Hunsley, in press) also reported effect sizes for training ($g = 1.13$) and experience ($g = 0.40$). The studies used in

that meta-analysis were a subset of those included in the current analysis. It is interesting to note a wider gap between in the effect sizes in the previous analysis, compared to those observed in the current study. This may in part be due to the fact that the current study included outcomes based on various measures beyond the PSDS. Some findings were based on as little as one question to assess the outcome construct (the number of items in the outcome construct was not, however, a significant moderator for either the training or experience analysis). When researchers use various measures in their studies, some of which were developed specifically for the study, it may impact study findings due to questions of reliability and validity of the measures employed. That is, if a measure does not produce reliable data for the sample used in the study, or if it is not a valid measure of the construct in question, it may not detect change as a result of experience or training.

If there was indeed more error introduced into the meta-analyses by virtue of possibly unreliable data and questionably valid outcome measures, this may help explain the reduced effect size for training (the previous analysis reported $g = 1.13$, current was $d = 0.69$). It is curious, however, why the effect size associated with experience would increase (the previous analysis reported $g = 0.40$, current was $d = 0.68$) when a wider array of outcome measures was included, beyond the PSDS. One possibility is that the PSDS is more sensitive to changes that result from training, than it is to those from experience. That is, including other outcome measures in the current analysis may have diluted/reduced the overall effect sizes for training. Perhaps for the experience construct, including a wider array of outcome measures beyond the PSDS (as in the current meta-analysis) may be a more sensitive assessment of the impact of experience, and as such larger effect sizes were detected in the current analysis.

Our second hypothesis was supported, in that the results indicated that prolonged training over a period of weeks and months was associated with significant larger effect sizes than found for training initiatives that were concentrated (i.e., occurring over a period of hours or subsequent days). This is in keeping with meta-analytic evidence from the medical field, which has noted that brief educational methods were not associated with significant effects (Davis et al., 1999). It may be that such prolonged learning experiences afford participants either (1) adequate time for supervisor development/self-efficacy to occur and/or (2) that prolonged training gives learners various opportunities to engage in learning/growth oriented that in turn promote supervisor development/self-efficacy. However, as this analysis was based on a limited subset of studies ($N = 8$), these results should be viewed as tentative and therefore require further examination in subsequent studies. Future research could further examine this pattern as well as possible interactions between timing of when such training is offered and the nature of the training activities.

There were no significant differences in effect sizes based on type of training. These results fail to support our third hypotheses (that training that involved an applied component, would be associated with larger effect sizes). When individual effect sizes are examined, combination training ($d = 0.52$), SOS ($d = 0.39$), and combination with SOS ($d = 2.65$) were significant, whereas didactic, practicum and workshop training were not. Given the very small number of studies, more research needs to be done on the different forms of training before any firm conclusions can be drawn. It may be that there exist interaction effects between type of training and stage of career. The development of computer software that permits the development of meta-regression

models with multiple moderators will facilitate the investigation of this relationship.

However, it is important to note that such analyses would likely require more statistical power than is present in the sample of studies in the current meta-analysis.

Moving now to our moderator analyses, an interesting finding was that mean age was a significant (inverse) moderator for training, though not for experience. That is, training accounted for more variance in outcomes for younger participants than it did for older participants. Barker and Hunsley's (in press) meta-analysis found similar results, in that training accounted for more variance in PSDS scores for younger participants (it should be noted however, that there was overlap in the studies that were used in the two meta-analyses). This may suggest a diminishing return of training on supervisor development and self-efficacy as supervisors age. With increasing age, supervisors may also have increased experience, so it is noteworthy the moderator analysis for age and experience was not significant. Unfortunately, the meta-analysis software is not able to enter multiple moderators in a meta-regression, so it is not possible to assess the potential significance of interactions among various variables.

Neither professional discipline nor doctoral degree were significant moderators of these relations, which suggests that both supervision training and supervision experience are important for a wide range of professions and with varying levels of education. It is possible that, although these variables are not individually associated with the supervisory outcomes examined in the meta-analysis, they could be associated in a more complex manner. As previously mentioned, because the meta-analysis software cannot conduct regression analyses with multiple, simultaneous predictor/moderator variables, it

was not possible to assess interaction effects among possible moderators, such as type of training program, for particular disciplines at varying experience levels.

It is worth noting that the overall summary effect sizes demonstrated high levels of between-study variability and that most of the moderators we examined did not contribute significantly in explaining the heterogeneity in the data. The heavy reliance in this literature on correlational designs, retrospective accounts of participants' training and experience, and the range of measures used to assess supervisory outcomes (many of which had limited or no evidence of psychometric adequacy) may well have contributed to the substantial differences authors reported across the individual studies. As only one study included in the training analyses used random assignment of participants to study (Kavanagh et al., 2008), it is obvious that research on the effects of supervisory training is at a very early stage of methodological rigor. Clearly there is a pressing need in this research area for studies using experimental or quasi-experimental designs, for without such studies it is not possible to unambiguously causally link the effect of supervisory training on subsequent supervisor development and self-efficacy.

Possible Limitations of the Study

Unfortunately, not all studies identified in the literature search could be included in the analyses because some did not report basic information on participants (e.g., age, gender, and discipline). Researchers are strongly encouraged to provide descriptive information about their participants, as required by scientific publication standards. Further, authors are strongly encouraged to provide means, standard deviations and sample sizes for all study findings. Presentation of this basic information will allow any future meta-analytic efforts to include as many relevant studies as possible.

As for any meta-analysis, a major challenge in our quantitative review involved the identification and retrieval all of potentially relevant studies. To address this issue as fully as possible, we purposefully conducted a very sensitive search, as evidenced by the number of initial “hits” yielded by our search strategy (i.e., abstracts reviewed), which lends confidence to the search strategy. Nevertheless, it is possible that we overlooked some possibly relevant studies. In particular, we did not search for studies written in a language other than English.

Authors of the studies included in the current analysis used various ways of assessing supervisory experience. Study authors often captured this information in a categorical manner. Categorizing continuous data unnecessarily reduces the statistical power of analyses and makes it very difficult to compare findings across studies (MacCallum, Zhang, Preacher, & Rucker, 2002; Streiner, 2002). Including information on the number of supervisees one has supervised and the number of years of experience, and using these details in subsequent analyses, is likely to provide researchers with the best options for generating simple summaries of supervisory experience.

We included studies that used various measures to assess supervisor development and self-efficacy. This may have impacted the results by introducing possible error due to issues of reliability and validity of the outcome measures (as noted above). Although a possible limit of the current study, including measures beyond the PSDS also builds on a previous meta-analysis (Barker & Hunsley, in press) which reported training and experience results only for the PSDS. Including a variety of studies and measures may more accurately reflect the status of the research in the area, although we recognize that it may introduce measurement error/variability. This points to a wider issue in the field, in

terms of the availability of psychometrically sound measures. Our analyses reflect the results of what is currently in the literature, but caution about the accuracy of the effect sizes is warranted in light of this.

The PSDS and other measures used in the reviewed studies are assessments based on self-report of development and self-efficacy. Meta-analytic and systematic review evidence indicates that, when compared with observational assessments (Davis et al., 2006) and external evaluations (Falchikov & Boud, 1989), both health professionals and students in various professions are limited in their capacity to accurately self-assess. As the supervision research base continues to develop, it will be important that a wide range of outcome measures is used to assess the impact of training and experience on supervision. To move beyond the heavy reliance on self-report instruments, researchers should consider using behavioural measures and/or observations of supervision interventions (Milne, James, Keegan, & Dudley, 2002).

Implications and Recommendations

As noted above, study authors are strongly encouraged to report full study data in their manuscripts, including: means, standard deviations, and sample size for all variables; correlation matrices of study variables; and, sample/group sizes for any subgroup analyses. Researchers are strongly encouraged to report any continuous data as such, in order to maximize the utility of this data for future possible meta-analyses. Additionally, authors are strongly encouraged to report reliability statistics for their individual studies, as every sample has unique characteristics that are likely to translate into different scores on the measure and every sample of scores may yield a reliability coefficient that differs from other samples (Therrien & Hunsley, 2012). Reporting both

years supervising and number of supervisees previously supervised in order to assess experience is recommended. Finally, providing detailed information about any training introduced to participants, such as length (in hours), duration (number of days, and over what time period), and features of the training (didactic, applied, combined).

The results of our meta-analyses indicate that both training and experience in supervision have substantial associations with supervisor development and self-efficacy, for a variety of disciplines and education levels. Training may have diminishing returns as supervisors age, or it may be that the training offered to older trainees needs to be amended for their particular learning needs. Investigating whether supervisor development and self-efficacy translate into meaningful impacts regarding supervisory competence as assessed by actual supervisory behaviours, as well as on supervisee learning and client care, is an important area in need of research. Training clinics may provide a unique venue to assess these research questions, as they include clients, clinicians and supervisors. The creation of research programs that capitalize on this rich opportunity, by assessing variables related to the supervisors' practice (their development, efficacy, knowledge, skills, attitudes), the actual supervision process that occurs between clinician and supervisor, supervisees' learning and development, as well as client outcomes, has the potential to greatly enhance our knowledge about many aspects of clinical supervision. Greater use of experimental and quasi-experimental designs in this research area will allow the field to move beyond noting simple associations among variables. This is a critical issue in our attempts to develop training options designed to ensure supervisor competence.

Based on our findings, for training and professional development programs, it is recommended that both training and experience be included in supervisory initiatives, as each demonstrated significant and positive effects. Further, although based on a limited number of studies, the evidence to date suggests that supervision training be offered over a period of weeks/months. Because, at this time, we do not have experimental evidence to guide us on specifics of optimal training and experience trajectories, we must rely on educational theories and guidelines used in other areas of practice to guide supervisory training initiatives. In this regard, it may be useful to consider the model of competency-based education (CBE), which posits that learning processes should be driven by learners acquiring competence, rather than by having all learners proceed through a set sequence of learning opportunities (i.e., didactic and clinical experiences) for a set period of educational time (Frank et al., 2010). The finding of the current meta-analysis supports the “full version” of CBE described by Hatcher and colleagues (2013, p. 226): “it is not that a student must pass course A, B, and C; it is that the student must demonstrate competencies 1, 2, and 3, which courses A, B, and C, along with other experiences, are designed to promote.” This form of CBE puts students at the center of the learning process, with the learning opportunities designed specifically for them and their competency goals. For some trainees, didactic training may provide much of what is needed to meet their particular learning goals. For others, hands-on practical experience, with SOS, may be what is primarily required. Because the findings of the current meta-analysis suggest that both training and experience may contribute to the supervisory outcomes of development and self-efficacy, it appears that educators and learners must make careful determinations about what fits each learner’s particular needs. In this

context, the ongoing evaluation of competency skills is essential to ensure that learners are benefitting from this process and are developing the necessary knowledge, aptitudes, and skills.

References

References marked with an () indicate studies included in the final sample*

American Psychological Association. (2006). *APA taskforce on the assessment of competence in professional psychology* (Final Report).

<http://www.apa.org/ed/resources/competency-revised.pdf>

Association of Psychology Postdoctoral and Internship Centers. (2012). *2012 APPIC*

Match: Survey of Internship Applicants. Retrieved April 20, 2014, from

<http://www.appic.org/Match/MatchStatistics/ApplicantSurvey2012Part1.aspx>

*Baker, S. B., Exum, H. A., & Tyler, R. E. (2002). The developmental process of clinical supervisors in training: An investigation of the supervisor complexity model.

Counselor Education and Supervision, 42, 15-30.

Barker, K. K., & Hunsley, J. (in press). Reliability and validity of the Psychotherapy

Supervisor Development Scale: A meta-analytic evaluation. *The Clinical*

Supervisor.

Barker, K. K., & Hunsley, J. D. (2013). The use of theoretical models in psychology supervisor development research from 1994 to 2010: A systematic review.

Canadian Psychology, 54, 176-185. doi: <http://dx.doi.org/10.1037/a0029694>

*Barnes, K. L. (2002). *Development and initial validation of a measure of counselor*

supervisor self-efficacy. (Ph.D. 3045772), Syracuse University, United States --

New York. Retrieved from

[https://login.proxy.bib.uottawa.ca/login?url=http://search.proquest.com/docview/](https://login.proxy.bib.uottawa.ca/login?url=http://search.proquest.com/docview/305471531?accountid=14701)

[305471531?accountid=14701](https://login.proxy.bib.uottawa.ca/login?url=http://search.proquest.com/docview/305471531?accountid=14701)

[2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+%28PQDT%29&atitle=&title=Development+and+initial+validation+of+a+measure+of+counselor+supervisor+self-efficacy&issn=&date=2002-01-01&volume=&issue=&spage=&au=Barnes%2C+Kristin+Lee&isbn=9780493597423&jtitle=&bttitle= ProQuest Dissertations & Theses \(PQDT\) database.](http://sfx.scholarsportal.info/ottawa?url_ver=Z39.88-</p>
</div>
<div data-bbox=)

Barnes, K. L., & Moon, S. M. (2006). Factor structure of the Psychotherapy Supervisor Development Scale. *Measurement and Evaluation in Counseling and Development, 39*, 130-140.

*Bencivenne, J. C. (1999). *An investigation of professional supervisor and supervisee development. (marriage and family therapy)*. Bencivenne, Joseph Charles: Seton Hall U, Coll Education And Human Services, US. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc3&AN=1999-95022-325> Available from Ovid Technologies PsycINFO database.

Bernard, J. M., & Goodyear, R. K. (1998). *Fundamentals of clinical supervision (2nd ed.)*. Needham Heights, MA: Allyn & Bacon.

Bernard, J. M., & Goodyear, R. K. (2014). *Fundamentals of clinical supervision (5th ed.)*. Boston: Pearson.

*Borders, L. D. (1996). Impact of a counseling supervision course on doctoral students' cognitions. *Counselor Education and Supervision, 35*, 204-217.

- *Borders, L. D., & Fong, M. L. (1994). Cognitions of supervisors-in-training: An exploratory study. *Counselor Education and Supervision, 33*, 280-293.
- Borenstein, M. (2009). *Introduction to meta-analysis*. Chichester, U.K.: John Wiley & Sons.
- Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive Meta-Analysis Version 2*. Englewood, NJ: Biostat.
- Canadian Psychological Association. (2002). *Accreditation standards and procedures for doctoral programmes and internships in professional psychology*. Ottawa: Canadian Psychological Association.
- Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology, 23*, 695-700.
- Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- Cooper, H. (2009). Hypotheses and problems in research synthesis. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 19-35). New York: Russell Sage Foundation.
- Crook-Lyon, R. E., Presnell, J., Silva, L., Suyama, M., & Stickney, J. (2011). Emergent supervisors: Comparing counseling center and non-counseling center interns' supervisory training experiences. *Journal of College Counseling, 14*, 34-49.

- *Culbreth, J. B., & Cooper, J. B. (2008). Factors impacting the development of substance abuse counseling supervisors. *Journal of Addictions & Offender Counseling, 29*, 22-35.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of Physician Self-assessment Compared With Observed Measures of Competence: A Systematic Review. *Journal of the American Medical Association, 296*, 1094-1102. doi: 10.1001/jama.296.9.1094
- Davis, D. A., O'Brien, M. A., Freemantle, N., Wolf, F. M., Mazmanian, P., & Taylor-Vaisey, A. (1999). Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *Journal of the American Medical Association, 282*, 867-874. doi: jrv90027 [pii]
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430.
- Falender, C. A., & Shafranske, E. P. (2004). *Clinical supervision: A competency-based approach* (1st ed.). Washington, DC: American Psychological Association.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*, 532-538. doi: 10.1037/a0015808
- Frank, J. R., Snell, L. S., Ten Cate, O., Holmboe, E. S., Carraccio, C., Swing, S. R., . . . Harris, K. A. (2010). Competency-based medical education: Theory to practice. *Medical Teacher, 32*, 638-645. doi: 10.3109/0142159x.2010.501190

- Goodyear, R. K., Bunch, K., & Claiborn, C. D. (2005). Current supervision scholarship in psychology: A five year review. *Clinical Supervisor, 24*, 137-147. doi: 10.1300/J001v24n01_07
- *Gourdine, R. M., & Baffour, T. D. (2004). Maximizing Learning: Evaluating a Competency-Based Training Program for Field Instructors. *The Clinical Supervisor, 23*, 33-53. doi: http://dx.doi.org/10.1300/J001v23n01_03
- *Haley, S. J. (2002). *The influence of supervision training on supervisor self-efficacy among doctoral interns at university counseling centers*. Haley, Sarah Jane: West Virginia U , US. ProQuest Dissertations and Theses database. (725930541)
- Hatcher, R. L., Fouad, N. A., Campbell, L. F., McCutcheon, S. R., Grus, C. L., & Leahy, K. L. (2013). Competency-based education for professional psychology: Moving from concept to practice. *Training and Education in Professional Psychology, 7*, 225.
- Hillman, S. L., McPherson, R. H., Swank, P. R., & Watkins, C. E., Jr. (1998). Further validation of the psychotherapy supervisor development scale. *The Clinical Supervisor, 17*, 17 - 32. doi: 10.1300/J001v17n01_02
- *Johnson, E. A., & Stewart, D. W. (2000). Clinical supervision in Canadian academic and service settings: The importance of education, training, and workplace support for supervisor development. *Canadian Psychology, 41*, 124-130.
- *Johnson, E. A., & Stewart, D. W. (2008). Perceived competence in supervisory roles: A social cognitive analysis. *Training and Education in Professional Psychology, 2*, 229-236. doi: 10.1037/1931-3918.2.4.229

- *Kavanagh, D. J., Spence, S., Sturk, H., Strong, J., Wilson, J., Worrall, L., . . . Skerrett, R. (2008). Outcomes of training in supervision: Randomised controlled trial. *Australian Psychologist, 43*, 96-104. doi: 10.1080/00050060802056534
- *Laschober, T. C., de Tormes Eby, L. T., & Sauer, J. B. (2012). Clinical supervisor and counselor perceptions of clinical supervision in addiction treatment. *Journal of Addictive Diseases, 31*, 382-388. doi:
<http://dx.doi.org/10.1080/10550887.2012.735599>
- *Lee, H. K., Veach, P. M., & LeRoy, B. S. (2009). An investigation of relationships among genetic counselors' supervision skills and multicultural counseling competence. *Journal of Genetic Counseling, 18*, 287-299. doi:
<http://dx.doi.org/10.1007/s10897-009-9219-y>
- *Lyon, R. C., Heppler, A., Leavitt, L., & Fisher, L. (2008). Supervisory training experiences and overall supervisory development in predoctoral interns. *Clinical Supervisor, 27*, 268-284. doi: 10.1080/07325220802490877
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.
- *McMahon, M., & Simons, R. (2004). Supervision training for professional counselors: An exploratory study. *Counselor Education and Supervision, 43*, 301.
- Milne, D. L., James, I., Keegan, D., & Dudley, M. (2002). Teacher's PETS: A new observational measure of experiential training interactions *Clinical Psychology & Psychotherapy, 9*, 187-199.
- Milne, D. L., Reiser, R., Aylott, H., Dunkerley, C., Fitzpatrick, H., & Wharton, S. (2010). The systematic review as an empirical approach to improving CBT supervision.

- International Journal of Cognitive Therapy*, 3, 278-294. doi:
10.1521/ijct.2010.3.3.278
- Pelling, N. (2008). The relationship of supervisory experience, counseling experience, and training in supervision to supervisory identity development. *International Journal for the Advancement of Counselling*, 30, 235-248.
- *Robiner, W. N., Saltzman, S. R., Hoberman, H. M., & Schirvar, J. A. (1997). Psychology supervisors' training, experiences, supervisory evaluation and self-rated competence. *Clinical Supervisor*, 16, 117-144. doi: 10.1300/J001v16n01_07
- Rodolfa, E. R., Haynes, S., Kaplan, D., Chamberlain, M., Goh, M., Marquis, P., & McBride, L. (1998). Supervisory practices of psychologists--Does time since licensure matter? *The Clinical Supervisor*, 17, 177-183. doi:
10.1300/J001v17n02_12
- *Ronnestad, M. H., Orlinsky, D. E., Parks, B. K., Davis, J. D., & The Society for Psychotherapy Research (SPR) Collaborative Research Network. (1997). Supervisors of psychotherapy: Mapping experience level and supervisory confidence. *European Psychologist*, 191-201.
- *Stevens, D. T., Goodyear, R. K., & Robertson, P. (1998). Supervisor development: An exploratory study in changes in stance and emphasis. *Clinical Supervisor*, 16, 73-88.
- Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry*, 47, 262-266.

- *Taylor, K. N., Gordon, K., Grist, S., & Olding, C. (2012). Developing supervisory competence: Preliminary data on the impact of CBT supervision training. *the Cognitive Behaviour Therapist, 5*, 83-92.
- Therrien, Z., & Hunsley, J. D. (2012). Assessment of anxiety in older adults: A systematic review of commonly used measures. *Aging and Mental Health, 16*, 1-16. doi: 10.1080/13607863.2011.602960
- *Vidlak, N. W. (2002). *Identifying important factors in supervisor development: An examination of supervisor experience, training, and attributes*. Vidlak, Nicole White: U Nebraska - Lincoln, US. ProQuest Dissertations and Theses database. (727413951)
- *Watkins, C. E., Jr., Schneider, L. J., Haynes, J., & Nieberding, R. (1995). Measuring psychotherapy supervisor development: An initial effort at scale development and validation. *The Clinical Supervisor, 13*, 77-90. doi: 10.1300/J001v13n01_06
- Wheeler, S., & Richards, K. (2007). The impact of clinical supervision on counsellors and therapists, their practice and their clients. A systematic review of the literature. *Counselling & Psychotherapy Research, 7*, 54-65.
- *White, E., & Winstanley, J. (2010). A randomised controlled trial of clinical supervision: Selected findings from a novel Australian attempt to establish the evidence base for causal relationships with quality of care and patient outcomes, as an informed contribution to mental health nursing practice development. *Journal of Research in Nursing, 15*, 151-167. doi: <http://dx.doi.org/10.1177/1744987109357816>

*Williams, S. L. (2012). Counselor supervisor development and self-efficacy: Factors influencing clinical supervisor competence. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 72, 7722.

Wilson, D. B. (2013). Practical meta-analysis effect size calculator. Retrieved June 24, 2013, from <http://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php>

Table 1

Meta-Analysis of Supervision Experience and Supervisor Development/Self-Efficacy

Study ^a	<i>N</i>	Study statistics			
		<i>d</i>	95% CI	<i>Z</i> Value	<i>p</i> Value
Barnes, 2002	285	0.508	[0.216, 0.800]	3.410	0.001
Bencivenne, 1999	123	0.055	[-0.316, 0.426]	0.290	0.772
Culbreth & Cooper, 2008	192	2.256	[1.826, 2.686]	10.288	0.000
Haley, 2002	145	0.389	[0.059, 0.720]	2.307	0.021
Johnson & Stewart, 2008	155	0.417 ^b	[0.092, 0.742]	2.516	0.012
Laschober et al., 2012	444	0.026	[-0.161, 0.213]	0.273	0.785
Lee et al., 2009	122	0.723	[0.341, 1.105]	3.709	0.000
Rønnestad et al., 1997	1505	1.365	[1.253, 1.477]	23.848	0.000
Stevens et al., 1998	50	1.130	[0.548, 1.712]	3.803	0.000
Vidlak, 2002	99	0.144	[-0.257, 0.545]	0.705	0.481
Watkins et al., 1995	319	0.496	[0.208, 0.784]	3.379	0.001
Williams, 2012	298	0.782	[0.537, 1.027]	6.252	0.000
Summary Effect Size	3,737	0.684	[0.309, 1.058]	3.579	0.000

Note. ^aDetails for each study can be found in Appendix B. ^bThis value represents a composite effect size.

Table 2

Meta-Analysis of Supervision Training and Supervisor Development/Self-Efficacy

Study ^a	N	Study statistics			
		<i>d</i>	95% CI	Z Value	<i>p</i> Value
Baker et al., 2002	19	2.169	[1.092, 3.246]	3.946	0.000
Barnes, 2002	254	0.355 ^b	[0.049, 0.661]	2.274	0.023
Bencivenne, 1999	123	0.626	[0.268, 0.985]	3.424	0.001
Borders & Fong, 1994	9	1.016 ^b	[-0.779, 2.811]	1.110	0.267
Borders, 1996	11	4.300	[1.014, 7.585]	2.565	0.010
Gourdine & Baffour, 2004	60	-0.331 ^b	[-0.853, 0.191]	-1.242	0.214
Haley, 2002	145	0.267 ^b	[-0.078, 0.612]	1.519	0.129
Johnson & Stewart, 2000	114	1.89	[1.336, 2.444]	6.687	0.000
Johnson & Stewart, 2008	155	0.372 ^b	[0.049, 0.696]	2.255	0.024
Kavanagh et al., 2008	46	-0.327	[-0.930, 0.276]	-1.062	0.288
Laschober et al., 2012	435	-0.158	[-0.347, 0.031]	-1.639	0.101
Lyon et al., 2008	230	1.540	[1.211, 1.868]	9.190	0.000
McMahon & Simons, 2004	57	0.848	[0.252, 1.445]	2.788	0.005
Robiner et al., 1997	62	0.561	[0.031, 1.091]	2.074	0.038
Stevens et al., 1998	33	1.093	[0.366, 1.820]	2.947	0.003
Taylor et al., 2012	28	1.805	[0.749, 2.861]	3.350	0.001
Vidlak, 2002	99	0.651	[0.204, 1.097]	2.855	0.004
White & Winstanley, 2010	24	0.769	[-0.147, 1.686]	1.645	0.100
Williams, 2012	298	0.208	[-0.021, 0.438]	1.780	0.075
Summary Effect Size	2,202	0.688	[0.384, 0.992]	4.435	0.000

Note. ^aDetails for each study can be found in Appendix B. ^bThis value represents a composite effect size.

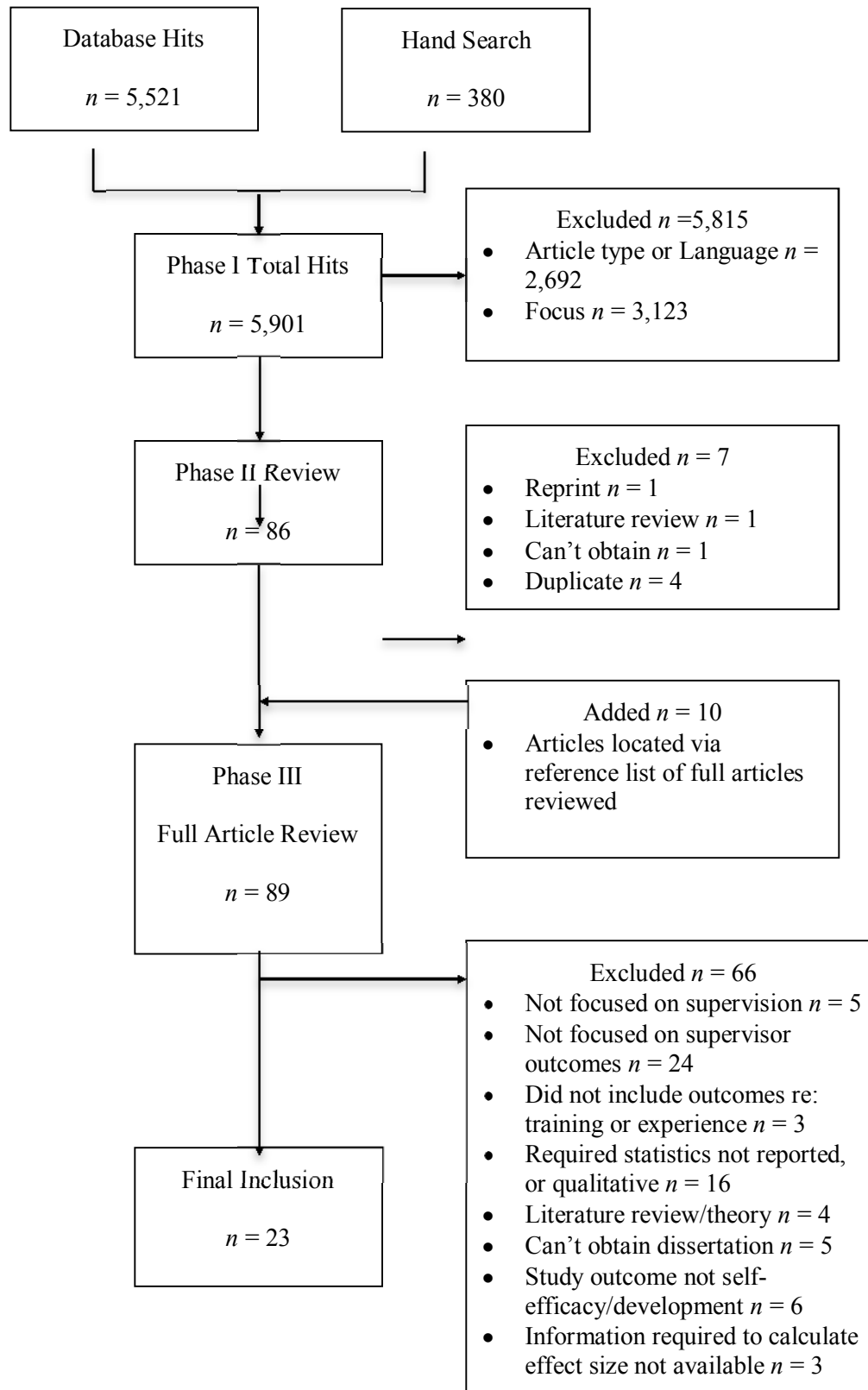
Table 3

Meta-Analysis of Type of Supervision Training and Supervisor Development/Self-Efficacy

Training Type	Number of studies	<i>Statistics for each training type</i>			
		<i>d</i>	95% CI	Z Value	<i>p</i> Value
Workshop	5	0.267	[-0.141, 0.675]	1.283	0.200
Didactic	7	0.329	[-0.027, 0.684]	1.814	0.070
SOS	5	0.387	[0.041, 0.733]	2.190	0.029
Practicum	2	0.497	[-0.169, 1.162]	1.463	0.144
Combination	2	0.522	[0.210, 0.834]	3.281	0.001
Combination with SOS	2	2.645	[0.905, 4.386]	2.979	0.003
Summary effect size	23 ^a	0.421	[0.252, 0.590]	4.892	0.000

Note. ^a Studies could contribute 1 effect size per training type, so some studies may contribute twice to the total.

Figure 1: Study Flow Diagram



General Discussion

The development of competency in clinical supervision is increasingly recognized as an important element of training (American Psychological Association, 2009; Canadian Psychological Association, 2002) and practice (Johnson & Stewart, 2000) in professional psychology. As an emerging area of competency, training students and early career professionals in how to provide clinical supervision is challenging for various reasons. It is difficult to find faculty or supervisors appropriately trained on how to teach supervisory skills (Hadjistavropoulos et al., 2010), supervisors themselves tend not to have received training in clinical supervision (Johnson & Stewart, 2008), and programs are required to provide various training requirements for various competency areas. Further, there is a lack of research on the developmental change process clinical supervisors undergo as they progress toward becoming competent supervisors. This dissertation research was designed to address some of these challenges by quantitatively summarizing to date the research on supervisor development with the use of systematic review and meta-analytic methods. Providing such a summary is useful at this time, for we are at a critical juncture: there are increasing calls to develop and implement programmatic initiatives to assist psychologists to develop competence in clinical supervision, and there exist opportunities to learn from and build upon previous work in the area, with the goal of improving the literature base. Watkins (2012) noted that having a “generation” (p. 48) of work in this area makes it an opportune time for review, so that clear recommendations can be made to guide future inquiry.

Results from the systematic review (Study 1) (Barker & Hunsley, 2013) suggest that research on supervisor development is indeed in its infancy, with an average of 1.6

empirical studies per year between 1994-2010, and only one of the included studies (Kavanagh et al., 2008) involved randomization of participants. Just over half (52%) of the reviewed studies did not use any theoretical model of supervisor development to inform their research design. Of the studies that were informed by theory, applied, or tested it, Watkins' Supervisor Complexity Model (SCM; Watkins, 1990, 1993, 1994) was the most commonly used model, having been used in 11 of the 12 studies that were theory-based. This is likely the case because the SCM is the model that has a measure associated with it: the Psychotherapy Supervisor Development Scale (PSDS; C. E. Watkins, Jr. et al., 1995).

I found in Study 1 that the SCM is the predominant model in the literature on supervisor development. This is likely due to the availability of an associated measure (the PSDS), therefore it became clear that the psychometric properties of the PSDS should be evaluated. In Study 2 (Barker & Hunsley, in press) I conducted a meta-analysis of the psychometric data associated with the PSDS. Results indicated that the majority of studies cited the reliability coefficients reported in the initial article describing the development of the PSDS (C. E. Watkins, Jr. et al., 1995). Despite being a rather common practice, this is not appropriate because the reliability is a function of the measure and the sample on which it is being used, rather than being simply associated with the measure itself. Based on the studies that did report their own reliability data, the PSDS showed excellent (Hunsley & Mash, 2008) mean internal reliability of $r = 0.93$, 95% CI [0.91, 0.94]. None of the following moderators were significant in the RG analysis: year of publication, number of participants who were female, mean age of participants, sample size, percentage of participants with a doctoral degree, and mean

years supervising. Participants' discipline (i.e. whether the majority of the sample was professional psychologists) was a significant predictor of variance across PSDS scores, with higher effect sizes ($r = 0.94$) when the majority of study participants were not from psychology than when they were ($r = 0.90$). It is important to note that both r values are above the cutoff ($\geq .90$) for reliability consistencies considered excellent (Hunsley & Mash, 2008). Thus, the findings of the moderator analyses indicate the PSDS is a reliable measure for supervisors, regardless of gender, age, doctoral degree status and discipline.

The SCM posits that training and experience impact a supervisor's development levels, and the meta-analysis (Study 2) (Barker & Hunsley, in press) found support for this: overall effect sizes from the PSDS were substantial for both training ($g = 1.13$) and experience ($g = 0.40$). It thus appears to be appropriately associated with relevant variables (experience, training) and as such it is reasonable to conclude that there is support for the measure's validity. The findings suggest that receiving training in supervision and having experience providing supervision are important for supervisor development, at least as measured by the PSDS. Moderator analyses regarding training and PSDS yielded inverse relations for mean age and for the percentage of participants with doctoral degrees, suggesting that training accounted for more variance in PSDS scores for younger participants than it did for older participants. Similarly, the higher proportion of participants with doctoral degrees, the less variance of PSDS scores was accounted for by training. The following moderators were not significant for the VG: year of publication, number of female participants, discipline, mean years supervising, and sample size.

The influence of training and experience was further explored in Study 3. This study is an important contribution to the literature because it quantitatively summarizes the research on the impact of training and experience on supervisor development and self-efficacy (the most commonly occurring outcomes in this literature). When both of these outcomes are included, and when a range of disciplines are involved, the meta-analytic findings suggest that both training ($d = 0.69$) and experience ($d = 0.68$) have similar significant impacts on supervisor development and self-efficacy. The finding that these effect sizes were not significantly different from each other may have emerged because the analysis included measures beyond the PSDS. There is currently no evidence that different forms of training have different effects, but the limited number of available studies may impact this.

Similar to Study 2, age was found to be a significant moderator for training, though not for experience. This again suggests a diminishing return of training on supervisor development and self-efficacy as supervisors age. With increasing age, supervisors may also have increased experience, so it is noteworthy that the moderator analysis for experience was not significant. It is possible that there may be interaction effects occurring, but due to limitations in the version of the software used for the analyses, it was not possible to assess for multiple moderators in the meta-regressions.

The following moderators were not significant variables in accounting for the variability in the summary effect sizes for training and for experience: year of publication, mean years supervision experience, number of items used to assess outcome variable, or the percentage of participants who were female, students, had doctoral degrees, identified psychology as their discipline, were students. Supervision training that

was prolonged (weeks/months) had significantly larger effect sizes compared to training that was concentrated (hours/days).

Taken together, the results from these three studies suggest that the SCM and PSDS have largely dominated the research in supervisor development, a finding that is supported by descriptive review by Watkins (2012). In keeping with other measurement research on the PSDS (Barnes & Moon, 2006; Hillman et al., 1998; C. E. Watkins, Jr. et al., 1995), the current results suggest that PSDS appears to be a psychometrically sound measure with the reliability generalization indicating excellent average reliability and the validity generalization indicating the measure's capacity to detect differences among supervisors based on training and experience in supervision. Further, the meta-analytic findings suggest that both experience and training play important roles in their impact on supervisor development and self-efficacy. This is an important finding, as previous research from individual studies indicated 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), and a narrative review noted that training mattered "greatly" (Watkins, 2012, p. 67) whereas experience alone was deemed "not generally sufficient to guarantee supervisor development" (p. 67).

Implications

The finding that just over half of the studies focused on supervisor development did not employ theory in any way may speak to the various theories' utility. The theories generally have not evolved since their conception (Watkins, 2012), in that they remain "general, imprecise and rudimentary" (p. 71). There are calls for their modification, so

that they become more precise and specific, particularly regarding how transitions occur between various levels of development, and the impact of various moderator variables on the growth process (Watkins, 2012). In his review of 30 years of research in the area, Watkins (2012) concluded that none of the theorists have engaged in concentrated or prolonged study of supervisor development. This, as well as the tendency for researchers in the area to engage in “one and done” (p. 73) projects, may account for the lack of refinement of the theories.

Despite this, researchers and educators can be confident that the PSDS is a tool with a history of excellent reliability. This is an important finding, because it is the only measure of its kind. Even though the findings of the reliability generalization found that it has excellent mean reliability, it is strongly recommended that future users of the PSDS evaluate and report reliability data for their own sample, as every sample of scores may yield a reliability coefficient that differ from other samples (Therrien & Hunsley, 2012). However, because the PSDS is a self-report tool, it is limited by users’ capacity to accurately self-assess, and other research indicates that both students and professionals are limited in this capacity (Davis et al., 2006; Falchikov & Boud, 1989).

The result that both training and experience in supervision are associated with supervisor development and self-efficacy may at first glance appear rather underwhelming. In fact, this finding lends support to competency-based education, which espouses that learning should be individualized, rather than by having all learners proceed through a set sequence of learning opportunities for a set period of educational time (Frank et al., 2010). That is, for one trainee experience in clinical supervision may be what is required, whereas for another it is classroom based training, or some

combination of the two. Further to this individualized approach, it does not currently appear that one particular “type” of training has clear advantages to others, in that effect sizes across various training “types” were not significantly different, in terms of supervisor development and self-efficacy, although this may be due to the limited number of studies. This is in contrast to apparent patterns across individual studies, which seemed to indicate that type/format of the training (Borders, 1996; Borders & Fong, 1994; Haley, 2002; McMahon & Simons, 2004; Pelling, 2008), the number of training experiences (Lyon et al., 2008; Stevens et al., 1998; Vidlak, 2002), and whether it is brief or sustained (Kavanagh et al., 2008; Vidlak, 2002) had differential impact on these outcomes. Because most of the research is correlational and retrospective, and because there is no one clear advantage of any one training type, it may be prudent at this point to rely on consensus statements regarding the development of clinical supervision competencies (Falender et al., 2004), guidelines regarding clinical supervision training (Norcross & Halgin, 1997), and professional opinion (Borders, 2010; Falender & Shafranske, 2004) which recommend “combination” approach to training (i.e., both didactic and experiential training). This “combination” approach can be included in competency-based education as the initial starting point for training programs. That is, it can be offered to learners as the initial training modality. From there, learning outcomes can be monitored, and learning plans tailored and individualized to take into account the particular learning needs and styles of individual learners.

Limitations of Reviewed Research

Across the three studies of this dissertation, there emerge limitations to the literature base on supervisor development. First, the research tends not to be guided by

theory. This is problematic because without theory, the ability to formulate clear and concrete hypotheses is limited. Further, the absence of theory results in disparate and singular studies being conducted with no way to tie them together with other studies. This is especially problematic given the dearth of research in the area in general. If numerous investigators, using different methods, measurements, and samples, all conduct research on elements of explicitly articulated theories, meaningful cohesion and synthesis may be possible.

Watkins' SCM is clearly the dominant theory used in this area of research, likely because it is the only theory with an associated measure (the PSDS). Our results indicated that the PSDS appears psychometrically sound. However, the literature is limited in that many studies did not report their own reliability estimates. Further to the issue of reporting, many studies did not report seemingly basic information about study samples, measures used, and results obtained. This again makes comparing and combining results from studies difficult. For example, moderator analyses on the PSDS in the current dissertation were limited due to missing demographic data in many studies.

There is a heavy reliance in the literature on supervisor development on correlational designs, self-report and self-assessment regarding study outcomes, and retrospective accounts of participants' training and experience. It is important to keep this in mind when understanding the results from the meta-analyses, in that causation cannot be determined. It is obvious that research on the impacts of supervisory training is at a very early stage of methodological rigour. Kavanagh and colleagues (2008) and E. White and Winstanley (2010) deserve special credit for implementing randomized, controlled designs in their studies of the impact of training on supervisor development and self-

efficacy. It is interesting to note that one of the studies (Kavanagh et al., 2008) that employed randomization yielded a negative effect size. The authors randomized supervisor trainees to either attend with their supervisees or not. The negative effect for this study emerged because supervisors in the split condition showed a drop on self-efficacy over time. The authors noted the possibility that training supervisors without providing supervisees with a framework that allows them to benefit from their supervisors' interventions may limit the impact of supervisory interventions. However, this study is the only one of its kind to employ a prospective, randomized approach, and also the only study to assess the possible differential impact of whether having supervisory dyads attend training together is beneficial. Clearly, more research into this question is needed.

The retrospective accounts of training and experience in the reviewed research were diversely classified and defined. This makes combining this information for meta-analytic purposes a challenge. As is the case in many research areas, investigators who conducted the research that was included in the meta-analyses often categorized continuous data (e.g., extent of past training and number of years supervision experience). However, this strategy reduces the statistical power of analyses in the original studies, and in the current dissertation makes it difficult to compare and combine findings across studies (MacCallum, Zhang, Preacher, & Rucker, 2002; Streiner, 2002).

Limitations of the Current Research

As is the case for any review-based research (e.g., systematic review, meta-analysis), the search strategy is akin to a sampling strategy in original research, and the articles to the study sample. As such, the sensitivity and specificity of the search strategy

are important considerations when assessing the representativeness of the sample of included studies. All of the studies in the current dissertation used a very sensitive search strategy, as evidenced by the high number of initial “hits.” This decision was made in order to maximize the possibility that studies would be included in the review, given the limited number of studies in the area.

However, this also means that for the two meta-analyses, research participants from disciplines beyond psychology were included. This was a practical decision, made to include the greatest number of studies in the meta-analyses. As such, a variety of disciplines were included, with varying educational requirements and experiences. The findings indicate that doctoral degree was a significant moderator for training and PSDS scores, such that when there were a higher proportion of participants with doctoral degrees, less variance in PSDS scores was accounted for by training. This suggests a differential association of training with PSDS scores for those with more extensive and specialized training (as is the case in of doctoral psychology programs). Thus future researchers would do well to consider the variable of years of education and training (e.g., degree level), rather than discipline *per se*, as possible moderator variables.

There were a number of moderator analyses run in the meta-analyses. This may have inflated the number of Type I errors. This leads to a consideration of what may be more important to guard against in the context of these particular studies, in this particular area of research: Type I or Type II error. Given the nature of the literature, erring on the side of being “generative” in testing hypotheses appears appropriate. That is, given the fact that research is limited and that the main purposes of the current project were to generate a summary of what has been done in the field and to make

recommendations for future work, generating ideas via the results of moderator analyses can help guide potential follow up questions for future research.

Recommendations for Future Research

Very broadly, future studies in this area would benefit from explicit articulation of the theoretical underpinnings that guide the research questions and hypotheses. Furthermore, study authors are strongly encouraged to report complete information about:

1. Participants (gender, mean age, discipline, highest degree attained)
2. Training interventions (length in hours; duration in number of days, and over what time period; and whether the training was didactic, applied, and/or combined)
3. Experience providing supervision may be best assessed by inquiring about both the number of supervisees one has supervised and the number of years of experience providing supervision. Related, research in to how best to conceptualize and understand “low”, “medium” and “high” experience is needed, possibly with the use of multiple moderators in meta-regression analyses (see below).
4. Measures used (calculating and reporting reliability coefficients based on the specific sample of the study)
5. Results (means and standard deviations for all continuous variables, the correlation matrix of continuous variables, sample/group sizes for any subgroup analyses). When continuous data are available, researchers should avoid arbitrarily categorizing these data in order to use ANOVA statistics. Instead, researchers are strongly encouraged to conduct analyses that use continuous data,

such as correlation and regression analyses. This will maximize the utility of this data for future possible meta-analyses. Further, and in keeping with trends and requirements in social science and medicine, authors are encouraged to report effect sizes and confidence intervals for their results, which will also aid in future meta-analyses.

The emergence of computer programs that have the capacity assess for the impact of multiple moderator variables is a promising development. It is possible such meta-regression models that contain more than one variable may help better explain the observed heterogeneity in the studies' findings. However, future researchers will need to be mindful of whether there exists sufficient statistical power for such interactive models.

For future research employing the PSDS, it is suggested that authors report both the overall score, as well as the sub-scores for the four factor structure (supervisor competence and effectiveness; commitment to supervision and development of a supervisory identity; self-awareness; and, sincerity in the role of the supervisor) as noted in previous investigations of the factor structure of PSDS (Barnes & Moon, 2006). The findings from the systematic review and meta-analyses indicate that the SCM and its associated measure, the PSDS, are by far the most often used, and that the PSDS has a history of excellent reliability and some indications of validity. This is good news for the field. However, future researchers would do well engage in theory and measure development, including the use of outcome measures that move beyond self-assessment. The fact that one theory and its measure dominate a limited research base can be understood to mean that it has been rigorously tested and that it has demonstrated superiority. It can also be understood to reflect the limited research and that the research

that is done, uses this theory and measure. As has been noted previously, (Borders, 1989; Ellis, 1991), it is recommended that researchers empirically evaluate the assumptions and premises of existing theoretical models of supervisor development, and use this research to expand and develop the model(s). 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.

Understanding the process through which supervisors go as they gain competence in supervision is a research area ripe for energetic researchers to emerge and stake their claim in the domain by developing systematic programs of inquiry. In terms of specific future research foci, to begin, research linking theoretical models of supervisor development (e.g., SCM) to competency frameworks (Hatcher et al., 2013; Kaslow et al., 2009; Rodolfa et al., 2013) could provide invaluable guidance to the field. In such a conceptualization, the model could providing an overall “road map” about supervisors’ development, with the competency framework providing behavioural “road signs” that indicate development is occurring. At this time, the SCM is the most often researched model and further developing and building on or expanding this model by making links to behaviourally oriented competency framework would be useful. Indeed, Gosselin and colleagues (Gosselin, Barker, & Kogan, 2014) have taken initial steps toward this, by combining a definition of supervisor development with a supervision competency framework, and then applying a best evidence synthesis to assess its utility.

The emergent model that blends theory of supervisor development with a competencies framework could then be investigated via a validation study. Ideally, these

validation efforts would involve data from various sources including: supervisors' own self-efficacy/development, the extent to which their supervisees are meeting training goals, and the extent to which associated clients are improving. Linking supervisory behaviour with supervisee and client outcomes is the ultimate goal, because the purpose of supervision is to ensure the integrity of clinical services and to develop service provision competence in the supervisee (Falender & Shafranske, 2004). Training clinics in professional psychology are in an ideal position to build such research programs, for they often possess all the necessary "ingredients": clients, trainees providing services, and supervising psychologists. In many jurisdictions (e.g., Ontario), the requirement that any supervision delivered by a trainee (i.e., psychology graduate student or intern) must be supervised by a registered psychologist also provides opportunity to study supervision of supervision and its impact on the learning and development of supervisors in training, their supervisees, and their clients.

It is clear that increasing the use of experimental and quasi-experimental study designs would be an important step forward in developing the literature and evidence base in this area. As the number of such studies increases, updates of the systematic review and meta-analyses in this dissertation will be warranted in order to ensure that results from the emerging literature base are accurately summarized, integrated and built upon. The current studies thus provide a platform from which to build the evidence base, in terms of a 'baseline' assessment of the literature to date. The search strategy and coding manuals used in this dissertation are important contributions that may assist future researchers in summarizing the literature in this area. It is my hope that the above

suggestions will aid in professional psychology's quest to provide evidence-based training aimed at developing trainees' supervision competency.

References

- American Psychological Association. (1994). *Graduate study in psychology*. Washington, D.C.
- American Psychological Association. (2006). *APA taskforce on the assessment of competence in professional psychology* (Final Report).
<http://www.apa.org/ed/resources/competency-revised.pdf>
- American Psychological Association. (2009). *Guidelines and principles for accreditation of programs in professional psychology*.
<http://www.apa.org/ed/accreditation/about/policies/guiding-principles.pdf>
- Association of Psychology Postdoctoral and Internship Centers. (2012). *2012 APPIC Match: Survey of Internship Applicants*. Retrieved April 20, 2014, from
<http://www.appic.org/Match/MatchStatistics/ApplicantSurvey2012Part1.aspx>
- Baker, S. B., Exum, H. A., & Tyler, R. E. (2002). The developmental process of clinical supervisors in training: An investigation of the supervisor complexity model. *Counselor Education and Supervision, 42*, 15-30.
- Baltes, P. B. (1983). Life-span developmental psychology: Observations on history and theory revisited. In R. M. Lerner (Ed.), *Developmental psychology: Historical and philosophical perspectives* (pp. 79-111). Hillsdale, NJ: Erlbaum.
- Barker, K. K., & Hunsley, J. (in press). Reliability and validity of the Psychotherapy Supervisor Development Scale: A meta-analytic evaluation. *The Clinical Supervisor*.

- Barker, K. K., & Hunsley, J. D. (2013). The use of theoretical models in psychology supervisor development research from 1994 to 2010: A systematic review. *Canadian Psychology, 54*, 176-185. doi: <http://dx.doi.org/10.1037/a0029694>
- Barnes, K. L., & Moon, S. M. (2006). Factor structure of the Psychotherapy Supervisor Development Scale. *Measurement and Evaluation in Counseling and Development, 39*, 130-140.
- Bencivenne, J. C. (1999). *An investigation of professional supervisor and supervisee development. (marriage and family therapy)*. Bencivenne, Joseph Charles: Seton Hall U, Coll Education And Human Services, US. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=psyc3&AN=1999-95022-325> Available from Ovid Technologies PsycINFO database.
- Bernard, J. M., & Goodyear, R. K. (1992). *Fundamentals of clinical supervision*. Boston: Allyn and Bacon.
- Bernard, J. M., & Goodyear, R. K. (1998). *Fundamentals of clinical supervision (2nd ed.)*. Needham Heights, MA: Allyn & Bacon.
- Bernard, J. M., & Goodyear, R. K. (2004). *Fundamentals of clinical supervision (3rd ed.)*. Boston: Pearson Education.
- Bernard, J. M., & Goodyear, R. K. (2014). *Fundamentals of clinical supervision (5th ed.)*. Boston: Pearson.
- Borders, L. D. (1989). A pragmatic agenda for developmental supervision research. *Counselor Education and Supervision, 29*, 16-24.

- Borders, L. D. (1994). *Supervision: Exploring the Effective Components*. ERIC/CASS Counseling Digest Series (pp. Report: ED372339. 372153p).
- Borders, L. D. (1996). Impact of a counseling supervision course on doctoral students' cognitions. *Counselor Education and Supervision, 35*, 204-217.
- Borders, L. D. (2010). Principles of best practices for clinical supervision training programs. In J. R. Culbreth & L. L. Brown (Eds.), *State of the art in clinical supervision* (pp. 127-150). New York: Routledge.
- Borders, L. D., & Fong, M. L. (1994). Cognitions of supervisors-in-training: An exploratory study. *Counselor Education and Supervision, 33*, 280-293.
- Borenstein, M. (2009). *Introduction to meta-analysis*. Chichester, U.K.: John Wiley & Sons.
- Canadian Psychological Association. (2002). *Accreditation standards and procedures for doctoral programmes and internships in professional psychology*. Ottawa: Canadian Psychological Association.
- Canadian Psychological Association. (2011). *Accreditation standards and procedures for doctoral programmes and internships in professional psychology* (Fifth revision). Ottawa: Canadian Psychological Association.
- Caruso, M., Dooley, K., & Dabit, J. (1995). *An evaluation of a clinical supervision course format*. Paper presented at the Annual Conference of the Midsouth Educational Research Association.
- Chagnon, J., & Russell, R. K. (1995). Assessment of Supervisee Developmental Level and Supervision Environment across Supervisor Experience. *Journal of Counseling & Development, 73*, 553-558.

- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of Physician Self-assessment Compared With Observed Measures of Competence: A Systematic Review. *Journal of the American Medical Association, 296*, 1094-1102. doi: 10.1001/jama.296.9.1094
- Davis, D. A., O'Brien, M. A., Freemantle, N., Wolf, F. M., Mazmanian, P., & Taylor-Vaisey, A. (1999). Impact of formal continuing medical education: do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *Journal of the American Medical Association, 282*, 867-874. doi: jrv90027 [pii]
- Ellis, M. V. (1991). Research in clinical supervision: Revitalizing a scientific agenda. *Counselor Education and Supervision, 30*, 238-251.
- Ellis, M. V., & Ladany, N. (1997). Inferences concerning supervisees and clients in clinical supervision: An integrative review *Handbook of psychotherapy supervision* (pp. 447-507). Hoboken, NJ: John Wiley & Sons Inc.
- Ellis, M. V., Ladany, N., Kregel, M., & Schult, D. (1996). Clinical supervision research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology, 43*, 35-50. doi: 10.1037/0022-0167.43.1.35
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430.
- Falender, C. A., Cornish, J. A., Goodyear, R., Hatcher, R., Kaslow, N. J., Leventhal, G., . . . Grus, C. (2004). Defining competencies in psychology supervision: A consensus statement. *Journal of Clinical Psychology, 60*, 771-785. doi: 10.1002/jclp.20013

- Falender, C. A., & Shafranske, E. P. (2004). *Clinical supervision: A competency-based approach* (1st ed.). Washington, DC: American Psychological Association.
- Falender, C. A., & Shafranske, E. P. (2012). The importance of competency-based clinical supervision and training in the twenty-first century: Why bother? *Journal of Contemporary Psychotherapy*, *42*, 129-137. doi: <http://dx.doi.org/10.1007/s10879-011-9198-9>
- Fergusson, D., Glass, K. C., Hutton, B., & Shapiro, S. (2005). Randomized controlled trials of aprotinin in cardiac surgery: Could clinical equipoise have stopped the bleeding? *Clinical Trials*, *2*, 218-229; discussion 229-232.
- Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B., . . . Rodolfa, E. R. (2009). Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology*, *3*(4, Suppl), S5-S26. doi: <http://dx.doi.org/10.1037/a0015832>
- Frank, J. R., Snell, L. S., Ten Cate, O., Holmboe, E. S., Carraccio, C., Swing, S. R., . . . Harris, K. A. (2010). Competency-based medical education: Theory to practice. *Medical Teacher*, *32*, 638-645. doi: 10.3109/0142159x.2010.501190
- Gandolfo, R. L., & Brown, R. (1987). Psychology intern ratings of actual and ideal supervision of psychotherapy. *Journal of Training & Practice in Professional Psychology*, *1*, 15-28.
- Gosselin, J., Barker, K. K., & Kogan, C. (2014). Setting the stage for an evidence-based model of psychotherapy supervisor development *Canadian Psychology*, *submitted*.

- Hadjistavropoulos, H., Kehler, M., & Hadjistavropoulos, T. (2010). Training graduate students to be clinical supervisors: A survey of Canadian professional psychology programmes. *Canadian Psychology/Psychologie canadienne*, *51*, 206-212.
- Haley, S. J. (2002). *The influence of supervision training on supervisor self-efficacy among doctoral interns at university counseling centers*. Haley, Sarah Jane: West Virginia U , US. ProQuest Dissertations and Theses database. (725930541)
- Hatcher, R. L., Fouad, N. A., Grus, C. L., Campbell, L. F., McCutcheon, S. R., & Leahy, K. L. (2013). Competency benchmarks: Practical steps toward a culture of competence. *Training and Education in Professional Psychology*, *7*, 84.
- Hess, A. K. (1986). Growth in supervision: Stages of supervisee and supervisor development. *Clinical Supervisor*, *4*, 51-67.
- Hess, A. K. (1987). Psychotherapy supervision: Stages, Buber, and a theory of relationship. *Professional Psychology: Research and Practice*, *18*, 251-259.
- Hess, K. A., & Hess, A. K. (1983). Psychotherapy supervision: A survey of internship training practices. *Professional Psychology: Research and Practice*, *14*, 504-513.
- Hillman, S. L., McPherson, R. H., Swank, P. R., & Watkins, C. E., Jr. (1998). Further validation of the psychotherapy supervisor development scale. *The Clinical Supervisor*, *17*, 17 - 32. doi: 10.1300/J001v17n01_02
- Holloway, E. L. (1987). Developmental models of supervision: Is it development? *Professional Psychology: Research and Practice*, *18*, 209-216. doi: 10.1037/0735-7028.18.3.209

- Hunsley, J. D., & Mash, E. J. (2008). Developing criteria for evidence based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3-16). Oxford: Oxford University Press.
- Johnson, E. A., & Stewart, D. W. (2000). Clinical supervision in Canadian academic and service settings: The importance of education, training, and workplace support for supervisor development. *Canadian Psychology, 41*, 124-130.
- Johnson, E. A., & Stewart, D. W. (2008). Perceived competence in supervisory roles: A social cognitive analysis. *Training and Education in Professional Psychology, 2*, 229-236. doi: 10.1037/1931-3918.2.4.229
- Kaslow, N. J., Borden, K. A., Collins, F. L., Forrest, L., Illfelder-Kaye, J., Nelson, P. D., . . . Willmuth, M. E. (2004). Competencies conference: Future directions in education and credentialing in professional psychology. *Journal of Clinical Psychology, 60*, 699-712.
- Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency Assessment Toolkit for professional psychology. *Training and Education in Professional Psychology, 3*(4, Suppl), S27-S45. doi: 10.1037/a0015833
- Kavanagh, D. J., Spence, S., Sturk, H., Strong, J., Wilson, J., Worrall, L., . . . Skerrett, R. (2008). Outcomes of training in supervision: Randomised controlled trial. *Australian Psychologist, 43*, 96-104. doi: 10.1080/00050060802056534
- Leddick, G. R., & Dye, H. (1987). Effective supervision as portrayed by trainee expectations and preferences. *Counselor Education and Supervision, 27*, 139-154.

- Lyon, R. C., Heppler, A., Leavitt, L., & Fisher, L. (2008). Supervisory training experiences and overall supervisory development in predoctoral interns. *Clinical Supervisor, 27*, 268-284. doi: 10.1080/07325220802490877
- MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19-40.
- McMahon, M., & Simons, R. (2004). Supervision training for professional counselors: An exploratory study. *Counselor Education and Supervision, 43*, 301.
- Milne, D. L. (1989). Personal issues in clinical supervision. *Irish Journal of Psychology, 10*, 353-367.
- Milne, D. L. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology, 46*, 437-447.
- Milne, D. L. (2009). *Evidence-based clinical supervision: Principles and practice*. Malden, MA ; Chichester, West Sussex: BPS Blackwell.
- Milne, D. L., James, I., Keegan, D., & Dudley, M. (2002). Teacher's PETS: A new observational measure of experiential training interactions *Clinical Psychology & Psychotherapy, 9*, 187 - 199
- Milne, D. L., & James, I. A. (2002). The observed impact of training on competence in clinical supervision. *British Journal of Clinical Psychology, 41*(1), 55-72. doi: 10.1348/014466502163796
- Milne, D. L., Reiser, R., Aylott, H., Dunkerley, C., Fitzpatrick, H., & Wharton, S. (2010). The systematic review as an empirical approach to improving CBT supervision. *International Journal of Cognitive Therapy, 3*, 278-294. doi: 10.1521/ijct.2010.3.3.278

- Nelson, P. D. (2007). Striving for competence in the assessment of competence: Psychology's professional education and credentialing journey of public accountability. *Training and Education in Professional Psychology, 1*(1), 3.
- Norcross, J. C., & Halgin, R. P. (1997). Integrative approaches to psychotherapy supervision. In C. E. Watkins Jr (Ed.), *Handbook of psychotherapy supervision*. Chichester: Wiley.
- Padesky, C. A. (1996). Developing cognitive therapist competency: Teaching and supervision models. In P. M. Salkovskis (Ed.), *Frontiers of cognitive therapy* (pp. 226-292). London: Guilford Press.
- Patton, S. M. (1987). *The training and development of counselors and psychotherapists: Towards a comprehensive developmental model*. Patton, Susan M : U Maryland.
- Pelling, N. (2008). The relationship of supervisory experience, counseling experience, and training in supervision to supervisory identity development. *International Journal for the Advancement of Counselling, 30*, 235-248.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences : a practical guide*. Malden, MA ; Oxford: Blackwell Pub.
- Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis, MN: University of Minnesota.
- Rodenhauser, P. (1994). Toward a multidimensional model for psychotherapy supervision based on developmental stages. *Journal of Psychotherapy Practice and Research, 3*, 1-15.

- Rodenhauser, P. (1997). Psychotherapy supervision: Prerequisites and problems in the process. In C. E. J. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 527-548). New York John Wiley & Sons.
- Rodolfa, E. R., Greenberg, S., Hunsley, J., Smith-Zoeller, M., Cox, D., Sammons, M., . . . Spivak, H. (2013). A competency model for the practice of psychology. *Training and Education in Professional Psychology, 7*, 71.
- Rodolfa, E. R., Haynes, S., Kaplan, D., Chamberlain, M., Goh, M., Marquis, P., & McBride, L. (1998). Supervisory practices of psychologists--Does time since licensure matter? *The Clinical Supervisor, 17*, 177-183. doi: 10.1300/J001v17n02_12
- Russell, R. K., & Petrie, T. A. (1994). Issues in training effective supervisors. *Applied & Preventive Psychology, 3*, 27-42. doi: 10.1016/S0962-1849%2805%2980106-4
- Schwartz, E. J. (1989). *Supervisors of psychology interns: Selected aspects of their experiences*. Schwartz, Ellen J : New York U, US.
- Scott, K. J., Ingram, K. M., Vitanza, S. A., & Smith, N. G. (2000). Training in supervision: A survey of current practices. *The Counseling Psychologist, 28*, 403-422. doi: 10.1177/0011000000283007
- Stevens, D. T., Goodyear, R. K., & Robertson, P. (1998). Supervisor development: An exploratory study in changes in stance and emphasis. *Clinical Supervisor, 16*, 73-88.
- Stoltenberg, C. D. (1981). Approaching supervision from a developmental perspective: The counselor complexity model. *Journal of Counseling Psychology, 28*, 59-65.

- Stoltenberg, C. D., & Delworth, U. (1987). *Supervising counselors and therapists*. San Francisco, CA: Jossey-Bass.
- Stoltenberg, C. D., McNeill, B., & Delworth, U. (1998). *IDM supervision: An integrated developmental model for supervising counselors and therapists*. San Francisco, CA: Jossey-Bass.
- Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry, 47*, 262-266.
- Taylor, K. N., Gordon, K., Grist, S., & Olding, C. (2012). Developing supervisory competence: Preliminary data on the impact of CBT supervision training. *the Cognitive Behaviour Therapist, 5*, 83-92.
- Therrien, Z., & Hunsley, J. D. (2012). Assessment of anxiety in older adults: A systematic review of commonly used measures. *Aging and Mental Health, 16*, 1-16. doi: 10.1080/13607863.2011.602960
- Tobin, D., & McCurdy, K. (2006). Adlerian-Focused Supervision for Countertransference Work with Counselors-in Training. *62*(154-167).
- Vidlak, N. W. (2002). *Identifying important factors in supervisor development: An examination of supervisor experience, training, and attributes*. Vidlak, Nicole White: U Nebraska - Lincoln, US. ProQuest Dissertations and Theses database. (727413951)
- Vieceli, A. E. (2007). *A comparison of supervisor development and supervision emphasis of academic and on-site supervisors*. Vieceli, Ada E.: Southern Illinois U Carbondale, US. ProQuest Dissertations and Theses database. (1212791611)

- Watkins, C., Schneider, L. J., Haynes, J., & Nieberding, R. (1995). Measuring psychotherapy supervisor development: An initial effort at scale development and validation. *The Clinical Supervisor, 13*, 77-90. doi: 10.1300/J001v13n01_06
- Watkins, C. E., Jr. (1990). Development of the psychotherapy supervisor. *Psychotherapy: Theory, Research, Practice, Training, 27*, 553-560. doi: 10.1037/0033-3204.27.4.553
- Watkins, C. E., Jr. (1993). Development of the psychotherapy supervisor: Concepts, assumptions, and hypotheses of the supervisor complexity model. *American Journal of Psychotherapy, 47*, 58-74.
- Watkins, C. E., Jr. (1994). The supervision of psychotherapy supervisor trainees. *American Journal of Psychotherapy, 48*, 417-431.
- Watkins, C. E., Jr. (1995a). Psychotherapy supervisor and supervisee: Developmental models and research nine years later. *Clinical Psychology Review, 15*, 647-680. doi: 10.1016/0272-7358%2895%2900038-Q
- Watkins, C. E., Jr. (1995b). Psychotherapy supervisor development: On musings, models, and metaphor. *Journal of Psychotherapy Practice & Research, 4*, 150-158.
- Watkins, C. E., Jr. (2012). Development of the psychotherapy supervisor: Review of and reflections on 30 Years of theory and research. *American Journal of Psychotherapy, 66*, 45-83.
- Watkins, C. E., Jr., Schneider, L. J., Haynes, J., & Nieberding, R. (1995). Measuring psychotherapy supervisor development: An initial effort at scale development and validation. *The Clinical Supervisor, 13*, 77-90. doi: 10.1300/J001v13n01_06

- White, E., & Winstanley, J. (2010). A randomised controlled trial of clinical supervision: Selected findings from a novel Australian attempt to establish the evidence base for causal relationships with quality of care and patient outcomes, as an informed contribution to mental health nursing practice development. *Journal of Research in Nursing, 15*, 151-167. doi: <http://dx.doi.org/10.1177/1744987109357816>
- White, K. E. (1998). *Becoming a supervisor: An intensive study of the early development of clinical supervisors*. University of Manitoba, Canada. ProQuest Dissertations and Theses database. (736787841)
- Williams, S. L. (2012). Counselor supervisor development and self-efficacy: Factors influencing clinical supervisor competence. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 72*, 7722.
- Winstanley, J. (2000). Manchester Clinical Supervision Scale. *Nursing Standard, 14*, 31-22.
- Worthington, E. L. J. (1984). Empirical investigation of counsellors as they gain experience. *Journal of Counseling Psychology, 31*, 63-75.
- Worthington, E. L. J. (1987). Changes in Supervision as Counselors and Supervisors Gain Experience: A review. *Professional Psychology: Research and Practice, 18*, 189-208.

Appendix A:

The Psychotherapy Supervisor Development Scale (C. Watkins, L. J. Schneider, J. Haynes, & R. Nieberding, 1995)

Please answer the following items using the scale below:

1	2	3	4	5	6	7
Never	Rarely	Sometimes	Half the time	Often	Most of the time	Always

- If you believe that an item is not at all true of you then you would use the “NEVER” descriptor.
- If you believe just the opposite then use the “ALWAYS” descriptor.
- If you believe somewhere in between about an item, then use the descriptor (e.g., SOMETIMES or HALF THE TIME) that would best convey your thoughts.

- ___ 1. If asked, “Do you really feel like a psychotherapy /counseling supervisor?”, I could honestly answer ‘yes’”.
- ___ 2. I believe I have a good awareness about myself as supervisor, the impact that I have on supervisees, and how I affect the supervisory situation as a whole.”
- ___ 3. I believe I have a good knowledge of and understanding about the supervision process itself.
- ___ 4. If asked, “Can you give a good assessment of yourself as supervisor?”, could easily answer ‘yes’.”
- ___ 5. Becoming a supervisor is an ongoing process that requires much time and energy, but I see myself as well on my way to getting there.
- ___ 6. I have a realistic awareness about my limitations and weaknesses as a supervisor.

- ___ 7. As a supervisor, I structure the supervision experience effectively.
- ___ 8. I must say that, when I perform my supervisory responsibilities I often think of myself as an imposter.
- ___ 9. I believe I am able to increasingly foster a sense of self-sufficiency in my supervision.
- ___ 10. When needed, I am able to be appropriately assertive and confrontive with my supervisees.
- ___ 11. Right now, I feel ill-at-ease and somewhat confused with the supervisor role.
- ___ 12. Sometimes I believe I'm just playing at being a supervisor.
- ___ 13. I have a realistic awareness about my limitations and weaknesses as a supervisor.
- ___ 14. I believe I am generally effective in dealing with transference/countertransference issues in supervision.
- ___ 15. Becoming and being a supervisor demands a commitment (i.e., to keep working at developing oneself as supervisor) that I believe I have made.
- ___ 16. I consider the supervision that I provide to be helpful to my supervisees.
- ___ 17. I just don't consider myself that identified with the supervisor role.
- ___ 18. I consider supervision to be a very important role that I perform.

Appendix B: Details regarding studies included in Study 3

Characteristics of Study Samples and Design

Study	Specified Analysis	N ^a	Sample		Study design
			Characteristics	Characteristics	
Baker et al., 2002	Training	19	Doctoral level participants in a supervision practicum		Prospective: Intervention (currently enrolled in practicum with SOS) vs. control
Barnes, 2002	Experience Training	285 254	Council for Accreditation of Counselling & Related Educational Programs (CACREP) counsellor education programs conducting supervision or enrolled supervision course		Retrospective survey, correlational
Bencivenne, 1999	Experience Training	123	American Association Marital & Family Therapy Approved Supervisors		Retrospective survey, correlational
Borders & Fong, 1994	Training	9	Doctoral level students; CACREP approved program Counsellor education; enrolled in required supervision training		Prospective: Within-subject, repeated-measures design. SOS.
Borders, 1996	Training	11	Doctoral level students; CACREP approved program Counsellor education; enrolled in required supervision course		Prospective: Within-subject, repeated-measures design. Intervention: Combination with SOS.
Culbreth & Cooper, 2008	Experience	192	Substance abuse counselling clinical supervisors (registered or certified w/ Alcohol and Other Drug Abuse certification board		Retrospective survey, correlational
Gourdine & Baffour, 2004	Training	60	Master's of Social Work supervisors/field instructors and their students		Prospective: Intervention (workshop) provided to supervisors. Random sample of students of instructors who had (intervention) and had not (control) received intervention

Study	Specified Analysis	N ^a	Sample		Study design
			Characteristics		
Haley, 2002	Experience Training	145	Pre-doctoral interns; APA accredited university counselling sites		Retrospective survey, correlational
Johnson & Stewart, 2000	Training	114	Supervisors from 45 Canadian university or clinical settings		Retrospective survey, correlational
Johnson & Stewart, 2008	Experience Training	155	Psychologists providing supervision at academic & internship programs		Retrospective survey, correlational
Kavanagh et al., 2008	Training	46	Supervision dyads: allied health practitioners in Mental Health		Prospective: random assignment to intervention or control. Intervention: Workshop
Laschober et al., 2012	Experience Training	444	Addictions counsellors and their clinical supervisors (dyads)		Retrospective survey, correlational
Lee et al., 2009	Experience Training	122	Genetic counselling supervisors		Retrospective survey, correlational
Lyon et al., 2008	Training	230	Pre-doctoral interns at APA accredited sites		Retrospective survey, correlational
McMahon & Simons, 2004	Training	57	Australian counsellors		Prospective: Within-subject, repeated-measures design. Intervention: Didactic training
Robiner et al., 1997	Training	62	Supervisors APA accredited clinical psychology internships		Retrospective survey, correlational
Rønnestad et al., 1997	Experience Training	1505	Psychotherapists from 12+ countries		Retrospective survey, correlational
Stevens, Goodyear, & Robertson, 1998	Experience Training	50	Practicing psychotherapists		Retrospective survey, correlational
Taylor et al., 2012	Training	28	Psychological wellbeing practitioners (United Kingdom)		Prospective: Within-subject, repeated-measures design. Intervention: Didactic training
Vidlak, 2002	Experience Training	99	Pre-doctoral intern supervisors at APA internship sites		Retrospective survey, correlational

Study	Specified Analysis	N ^a	Sample		Study design
			N ^a	Characteristics	
Watkins et al., 1995	Experience	319	Members of APA Division of Psychotherapy		Retrospective survey, correlational
White & Winstanley, 2010	Training	24	Mental health nurses		Prospective: Random assignment to intervention or control; but outcome of interest not administered to control group (thus study treated as Prospective, Within-subject, repeated-measures design. Intervention: Didactic training)
Williams, 2012	Experience Training	298	American Counselling Association & Counsellor Education and Supervision clinical supervisors		Retrospective survey, correlational

Note. ^aSample sizes reflect the number of participants included in respective analysis