WHAT IS THE EVIDENCE BASE SUPPORTING BEST PRACTICE SUPERVISION GUIDELINES?
A SYSTEMATIC REVIEW AND MULTILEVEL META-ANALYSIS OF THE ASSOCIATION BETWEEN SUPERVISION AND THERAPIST PERFORMANCE

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

SIENA KELLER TUGENDRAJCH

Dr. Kristin M. Hawley, Dissertation Supervisor

July 2022

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

WHAT IS THE EVIDENCE BASE SUPPORTING BEST PRACTICE SUPERVISION GUIDELINES?

A SYSTEMATIC REVIEW AND MULTILEVEL META-ANALYSIS OF THE ASSOCIATION BETWEEN SUPERVISION AND THERAPIST PERFORMANCE

presented by Siena Keller Tugendrajch,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Professor Kristin M. Hawley

_____

Professor Debora J. Bell

_____

Professor Victoria A. Shaffer

_____

Professor Aaron M. Thompson

DEDICATION

This dissertation is dedicated to my fiancé Andrew for deciding to marry me even after seeing me cry about Excel formulas in the middle of the night and having to leave multiple dates abruptly to see clients in crisis. He keeps moving across the country for my career, he bought me a second monitor for the best, dorkiest birthday present ever so I could finish this dissertation, and I feel lucky every day to be so loved. This dissertation is also dedicated to my family, to my father for trying desperately to understand what "meta data" is, to my mother for always being proud of me, to my brother for always reminding me to keep going, and to my aunt for knowing when to say "enough!" when I need to get off the phone and start working again. This dissertation is also dedicated to my friends who send their love and support from afar as I have bounced around the country to pursue my passions for implementation science and increasing access to high quality mental health services. Finally, this dissertation is dedicated to my cohort, the strong, incredible, resilient women who support one another endlessly, even in a global pandemic where most of us live in different states for this final year of training. There would be no dissertation to read without this love and support that provides the foundation and heartbeat for all that I have done and will do in the future.

ACKNOWLEDGEMENTS

I thank Dr. Kristin M. Hawley for her mentorship, wisdom, compassion, and guidance throughout my time in graduate school, and for her flexibility, patience, and kindness navigating more than two years of our time together during a global pandemic. I also thank her for encouraging me to pursue a research area that was previously unfamiliar to us both, and for supporting me in pursuing non-traditional training experiences that have helped me identify what work is most exciting to me as an aspiring clinical psychologist. I would also like to thank Drs. Debora J. Bell, Victoria A. Shaffer, and Aaron M. Thompson for their time and effort serving as my committee members to enhance this project, and for Drs. Thompson and Shaffer for serving on all my committees as a graduate student and supporting me through these many milestones. Thank you also to my lab-mates and co-authors Jack H. Andrews, Evelyn Cho, Brigid R. Marriot, Rylee S. Park, and Kaitlin M. Sheerin for their many hours of coding and Zoom consensus meetings, in addition to their unwavering support of my program of research. Our lab likes to look at research questions from a very wide angle, and this work would be much harder and much less enjoyable without this amazing team. In this work, I also earned that conducting a meta-analysis actually means conducting tons of meta-analyses, so thank you to Dr. Tyler E. Smith for providing support for the many meta-analyses we ran for the final study. Thank you also to the many undergraduate research assistants I only worked with over Zoom the last two years who dedicated so much time to screening thousands of articles. Finally, thank you to my undergraduate advisor Dr. Casey A. Schofield for welcoming me to the exciting, messy world of clinical psychology research, to my postbaccalaureate mentor Dr. Amy Hughes Lansing for her compassionate mentorship throughout the years, and to my DIS SIG mentor Dr. Brenna B. Maddox

who has supported me well beyond the six months our relationship was designed to last. This dissertation would not have been possible without this immense support, for which I am deeply grateful.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

WHAT IS THE EVIDENCE BASE SUPPORTING BEST PRACTICE SUPERVISION GUIDELINES?

A SYSTEMATIC REVIEW AND MULTILEVEL META-ANALYSIS OF THE ASSOCIATION BETWEEN SUPERVISION AND THERAPIST PERFORMANCE

Siena Keller Tugendrajch

Dr. Kristin M. Hawley, Dissertation Supervisor

ABSTRACT

Clinical supervision is widely considered to be integral to effective clinical training, with best practice guidelines of each mental health discipline proposing common supervision elements for effective supervision. However, it is unclear to what extent these common supervision elements are supported by empirical evidence. The aims of this program of research are to a) describe relevant supervision models including theories and goals of supervision, b) detail common and distinct elements across supervision guidelines from three primary psychotherapy disciplines (i.e., psychology, counselling, & social work), c) review methods and preliminary findings from an initial scoping review of supervision studies, e) conduct a systematic review and meta-analysis evaluating the relationships between commonly proposed supervision elements and relevant outcomes, and f) discuss limitations of existing empirical literature on clinical supervision, and (g) propose future directions for supervision research.

First, we reviewed professional supervision guidelines for psychologists, counselors, and social workers and identified 17 common supervision elements across these best practice supervision guidelines. Next, we conducted a scoping review, screening 1517 published articles, and coding 28 studies from 26 articles. These supervision studies examined the relationship between common supervision elements

and numerous supervisor-, therapist-, and client-level outcomes. Some supervision elements were examined more often than others (e.g., therapist outcomes). In addition, some supervision elements were more consistently associated with positive outcomes (e.g., fostering a collaborative supervisor-supervisee relationship, encouraging therapist self-evaluation).

Next, we conducted a systematic review and meta-analysis to improve understanding of the associations between common supervision elements and meaningful therapist outcomes. We searched for studies that included supervision provided to mental health professionals with pre- and post-supervision measures for at least one relevant therapist outcome (i.e., adherence, competence, self-efficacy). Of 346 full-text articles screened for eligibility, 29 studies from 32 articles met inclusion criteria and contained sufficient data for quantitative analysis. Across these studies, we identified 110 effect sizes from 38 groups receiving supervision with at least one common supervision element, and we also identified 1 group receiving supervision with no common supervision elements, and 7 groups receiving no supervision (including 6 groups that received training). Results indicate that the average pre-post effect sizes for the 38 groups receiving supervision with at least one common supervision element was statistically significant and in the large range, indicating improvement in therapist outcomes following supervision (g = .787). Most of the hypothesized moderators were unrelated to therapist improvements. Contrary to hypothesis, the more common supervision elements incorporated into supervision, the less therapist improvement was achieved pre to post supervision. We also examined the impact of each of the 17 identified common supervision elements separately. None of the common supervision elements were related to greater improvement in therapist outcomes following supervision, when compared with supervision not

containing the given common supervision element. Concerningly, the *lack* of three common supervision elements were each associated with more therapist improvement pre to post supervision: beneficence for the client, technology considerations, and general feedback. Our results suggest that without studying common supervision elements from best practice guidelines systematically (i.e., evaluating the impact of specific common supervision elements on meaningful therapist outcomes), it is challenging to identify which supervision elements will maximize the benefit of supervision. Gaps in existing supervision research and future directions for evaluating supervision are discussed.

Keywords: clinical supervision, supervision guidelines, evidence-based supervision

Introduction and Literature Review

Clinical supervision is a cornerstone of clinician training and a core value of graduate training for mental health professionals. Indeed, all clinical trainees are required to undergo direct clinical supervision prior to graduation and to complete thousands of supervised hours of experience prior to licensure, with strict requirements for who can provide this supervision and what this supervision should entail (e.g., APA, 2015; CACREP, 2016; NASW, 2013). A core assumption underlying this universal requirement is that competent supervision leads to higher quality therapy, which leads to better client or patient outcomes. Supervision, much like licensure, is presumed to provide a gatekeeping function to protect the public from incompetent therapists. Clinical supervision experts have proposed theoretical models of supervision, and professional associations have detailed best practices based on these models, which are intended to guide the conduct of supervision in order to ensure adequate oversight of therapists-in-training and quality of services provided to the public.

Theoretical Models of Supervision

Several supervision models have been developed and proposed for use among supervisors and training programs.

Developmental supervision models. Stage-based developmental models propose that the development of the supervisee proceeds thought sequential phases, with each one building on the last and each one meriting more independence, somewhat akin to human development (e.g., Loganbill, Hardy, & Delworth, 1982). As an example, the Integrated Development Model (Stoltenberg & Delworth, 1987) defines three developmental levels for the supervisee across the following professional activities: intervention skill competence, assessment techniques,

interpersonal assessment, client conceptualization, individual differences, theoretical orientation, treatment goals and plans, and professional ethics. At the first level, the supervisor helps to contain the supervisee's anxiety and role models appropriate behavior. At the second level, the supervisor provides less guidance and encourages more autonomy for the supervisee by transitioning to a more facilitative than didactic emphasis. At the third level, the supervisor works towards ensuring consistency in the supervisee's performance across domains while identifying deficits and helping the supervisee to refine their professional identity (Stoltenberg & Delworth, 1987).

Competency-based supervision models. Competency-based supervision models focus on the supervisee's development of and demonstration of specified therapist competencies. Building on the work of Kaslow and colleagues (2004) emphasizing trainee competencies, the stated goal of competency-based supervision is to integrate relevant knowledge, skills, and behavior into practice (Falender & Shafranske, 2004). As an example, the proposed Supervision Competencies Framework (Falender et al., 2004) emphasizes knowledge (e.g., in the area being supervised, theories of supervision, professional development, evaluation, diversity), skills (e.g., supervision modalities, relationship skills, ability to provide feedback, ability to promote supervisee growth, flexibility, scientific thinking), values (e.g., responsibility, respectful, empowering, commitment to lifelong learning, knowing one's limitations), social context overarching issues (e.g., diversity, ethical and legal issues, knowledge of sociopolitical context, climate where feedback is honest, challenging, and supportive), training of supervision competencies (e.g., coursework in supervision, reception of supervision of supervision), and assessment supervision competencies (e.g., successful completion of supervision course, supervisory experience reflecting diversity, self-assessment and awareness, documentation of

supervisee feedback, assessment of individual and group supervision outcomes). Notably, competency-based models also emphasize the importance of multicultural competence in supervision (Falender, Burnes, & Ellis, 2013).

Treatment orientation-specific supervision models. Many treatment orientations also propose supervision models specifying how a supervisor helps a therapist develop skills and competence that parallel therapist models describing how a therapist helps a client or patient develop skills and competence. While the supervisory relationship between supervisor and supervisee is distinct from a therapeutic relationship between therapist and client, these supervision models suggest a similar process. In cognitive-behavioral therapy (CBT) supervision, the supervisor uses similar CBT practices with their supervisee: 1) modeling what a therapist does during a therapy session, such as setting an agenda, 2) reviewing previous homework, 3) asking for feedback from the client/supervisee, and 4) assigning homework to be completed by the next supervision session (Liese & Beck, 1997). Pretorious (2006) provided a framework for the principles and goals of CBT supervision, the format of CBT supervision meetings, the course and stages of CBT supervision, and the importance of the supervisory relationship in this supervision model.

As an extension of the CBT supervision model, Milne and colleagues (e.g., Milne & Westerman, 2001; Milne et al., 2008; Milne et al., 2014) have worked for two decades to identify empirically supported supervision guidelines for use in the United Kingdom's National Health Service. They conducted several systematic reviews of supervision (Milne et al., 2010; Milne & James, 2000) and developed measures to assess supervisory practices (Supervision and Adherence Guidelines Evaluation [SAGE], 2008; Milne & Reiser, 2008) and supervision fidelity (Fidelity

Framework Checklist; Reiser & Milne, 2014). Building on Pretorius's (2006) previous recommendations specifically for CBT supervision, Milne and Dunkerley (2010) identified four essential supervision guidelines: "Developing the Supervision Contract (including collaborative agenda-setting), Methods of Facilitating Learning (including making supervision an active process with experiential methods, such as reviewing tapes), Evaluation in Supervision (e.g., reviewing one's competence), and the Supervisory Alliance (the relational context)."

Integrated and transtheoretical models. Essentially an orientation-specific supervision model for "integrated" or "transtheoretical" treatment, these models integrate multiple theories and techniques of supervision and focus on what is shared or common across different specific treatment orientations. As an example, Bernard's (1997) transtheoretical discrimination model suggests there are three basic supervisor roles (i.e., therapist, teacher, and consultant) and three foci of supervision (process, conceptualization, and personalization). Competent supervision includes all of these roles and foci. Process or intervention refers to the supervisee's observable skills, such as demonstrating therapeutic techniques in session. Conceptualization refers to the supervisee's cognitive processes, such as recognizing patterns in their clients. Personalization refers to the supervisee's ability to adapt their own personal style and their self-awareness (Beinhart, 2004).

Supervision Guidelines

These existing theoretical models of supervision have influenced the development of professional guidelines for supervision to varying degrees, with the principles of competency-based and developmental models cited most frequently. The national professional organizations representing each of these disciplines have respective guidelines for the practice of supervision: the American Psychological

Association's (APA, 2014) *Guidelines for Clinical Supervision in Health Service Psychology*, the Association for Counselor Education and Supervision's (ACES, 2011) *Best Practices in Clinical Supervision*, and the National Association of Social Workers (NASW, 2013) *Best Practice Standards in Social Work Supervision*. These supervision guidelines are aspirational in nature, informed by supervision theory, and intended to provide a general framework for supervision. They appear to share several core recommended practices (e.g., assessment of supervisee competence, setting expectations for supervision) but each has unique aspects and foci as well (e.g., required supervised supervision is unique to ACES).

In further distinguishing across these guidelines, the APA's (2014) *Guidelines* are heavily influenced by tenets of the meta-theoretical competency-based supervision model (i.e., Falender & Shafranske, 2004, 2007; Fouad et al., 2009), as well as empirical studies of the effects of supervision (e.g., Milne & Reiser, 2012; Watkins Jr., 2011). Within these *Guidelines*, there are seven identified domains: (a) supervision competence (e.g., maintaining supervisor and supervisee competence, continuing education), (b) diversity (e.g., self-awareness of diversity competence, knowledge of the effects of bias, prejudice, and stereotyping), (c) supervisory relationship (e.g., establishing a collaborative relationship, setting expectations for the supervisory relationship), (d) professionalism (e.g., modeling professionalism, providing formative and summative feedback of supervisees' progress towards professionalism), (e) assessment/evaluation/feedback (e.g., monitoring and providing feedback on supervisee performance, direct and clear feedback), (f) professional competence problems (e.g., addressing competence problems directly, gatekeeping), and (g) ethical, legal, and regulatory considerations (e.g., modeling ethical practices, creating a supervision contract).

For counseling, the ACES's (2011) *Best Practices* are designed to provide specific guidance for ACES members on supervision and on supervisor training, based on credentialing and licensure requirements, ethical and legal standards, common beliefs about supervision, conceptual literature, and available evidence for supervision practices (ACES, 1990; Borders, 2014). The ACES (2011) *Best Practices* are also more detailed than APA's (2014) *Guidelines*, especially in describing supervision format. Although the *Best Practices* lack explicit references to published literature, they repeatedly emphasize the need to consider developmentally appropriate supervision processes, similar to developmental models of supervision. They include twelve sections: (a) initiating supervision (e.g., outlining informed consent, discussing expectations,), (b) goal-setting (e.g., developing goals, attending to goals throughout the supervisory relationship), (c) giving feedback (e.g., providing regular, descriptive, constructive, and developmentally appropriate feedback), (d) conducting supervision for individual, group, and triadic formats (e.g., meeting frequently with supervisees, evaluating the effectiveness of supervision), (e) the supervisory relationship (e.g., addressing diversity issues, power differentials, transference, and conflict), (f) diversity and advocacy considerations (e.g., facilitating cultural competence in supervisees, attending to cultural factors in case conceptualization), (g) ethical considerations (e.g., adhering to ethical principles throughout supervisory work, modeling ethical behaviors for supervisees, helping supervisees to protect client welfare), (h) documentation (e.g., establishing accountability through a jointly created supervision contract, written supervision session notes, and documented supervisee evaluations), (i) evaluation (e.g., clearly communicating the evaluation plan with trainees, encouraging supervisee self-evaluation, inviting supervisee feedback), (j) supervision format (e.g., prioritizing

client welfare and supervisee needs over supervisor convenience), (k) the supervisor (e.g., continuing education, knowledge of processes underlying supervision, supervisor qualities), and (l) supervisor preparation: supervisor training and supervision of supervision (e.g., didactic instruction, experiential learning, role-modeling, acknowledging the influence of the supervisory relationship as primary means of learning in supervision). Borders (2014) suggests that the supervisory relationship content area has the greatest empirical support of the twelve sections.

Finally, the NASW's (2013) *Supervision Standards* are described as a general framework and resource for those in the social work supervisory community. They are based on the NASW *Code of Ethics* (2008) and the Association of Social Work Boards *Model Social Work Practice Act* (2011). These *Supervision Standards* suggest that, given the number of available supervision models, it is the supervisor's responsibility to choose a model that best benefits the professional development of their supervisee. The *Supervision Standards* includes five standards: (a) context in supervision (e.g., scope of practice, interdisciplinary supervision, cultural awareness and cross-cultural supervision), (b) conduct of supervision (e.g., competence, role-modeling professional behavior, confidentiality of the supervision process, self-care), (c) legal and regulatory issues (e.g., liability and risk management, documentation), (d) ethical issues (e.g., ethical-decision making, setting appropriate boundaries, self-disclosure), and (e) technology (e.g., distance supervision, awareness of risks and benefits of technologies). The *Supervision Standards* also include additional sections with guidance on terminating the supervisory relationship (e.g., transitioning to a new supervisor, completing documentation by termination) as well as evaluation and outcomes (e.g., tools to measure progress towards identified goals, continuous feedback).

Supervision Goals and Outcomes

Despite the oft-stated importance of clinical supervision, the numerous proposed supervision models and discipline-specific guidelines, historically has been little empirical attention given to the processes, content and outcomes of clinical supervision (Watkins Jr., 2020). The limited research that has been conducted has been criticized for poor methodological rigor, including failure to operationalize supervision techniques and content (Milne, 2007) and failure to define meaningful supervisee or client outcomes or to link supervision practices with these outcomes (Falender, 2014; Holloway & Neufeldt, 1995; Reiser & Milne, 2014) beyond supervisee satisfaction (Bernard & Goodyear, 1998; Watkins Jr, 2014).

In the past decade, however, there has been growth in the formal inquiry of supervision processes, content, and outcomes. Much of this increased attention has emerged from the field of implementation science, out of recognition that supervision may play a critical role in determining the degree to which evidence-based practices (EBPs) are routinely and successfully implemented in mental healthcare (e.g., Carlson et al., 2012; Dorsey et al., 2017). For example, implementation scientists have examined multiple contextual factors that may impact the amount of time allotted for supervision (Choy-Brown & Stanhope, 2018), the use of active learning strategies such as behavioral rehearsal within supervision (Dorsey et al., 2018; Lucid et al., 2018), and the extent to which EBPs are covered in supervision versus administrative issues such as documentation and billing (Accurso, Taylor, & Garland, 2011; Pullman et al., 2018).

When researching supervision practices and their potential effects, it is necessary to clarify the goals of the supervisory relationship. Given the recent emphasis on competency-based models, the fostering of supervisee competence (i.e.,

knowledge, skills, practice) is increasingly a stated goal of supervision (Falender et al., 2004; Milne, 2007). Especially relevant to implementation science, supervisee outcomes may also extend to therapist adherence to and skill in providing EBPs (e.g., Barrett et al., 2019; Bearman et al., 2013; Dorsey et al., 2013; Schoenwald, Sheidow, & Letourneau, 2004). However, although therapist or supervisee outcomes are important and may be the most directly amenable to change via supervision, client outcomes are often seen as the "acid test" of supervision (Milne, 2014). Client outcomes include reduction in symptoms (e.g., Hill & Knox, 2013), but also the client's perspective on the therapeutic alliance or working relationship with the therapist (Weck, Kaufmann, & Witthöft, 2017), client satisfaction with treatment (e.g., Bambling et al., 2006), client attendance, engagement, and drop-out in therapy. Despite the general assumption that supervision practices lead to changed therapist behaviors, which then benefit client functioning, previous reviews of clinical supervision have found little empirical work and even less empirical support for supervision improving client outcomes (e.g., Wheeler & Richards, 2007). While there have been calls to formally monitor client outcomes throughout treatment as it relates to supervision (e.g., Worthen & Lambert, 2007), this does not appear to common practice in the current supervision literature.

Importantly, across these therapist and client outcomes, the majority of supervision studies look at the effects of supervision holistically, without identifying which supervision elements specifically correspond with identified outcomes (Simpson-Southward, Waller, & Hardy, 2017). This lack of specificity in existing research reviews makes it difficult to provide evidence-based guidelines for the process and content of supervision.

Summary

Despite the longstanding centrality of supervision in the training of all mental health professionals, the availability of numerous supervision models and theories from which to select a supervision approach, and the recent increase in supervision research, it remains unclear whether and to what extent the specific practices recommended in formal supervision guidelines for the three largest mental health disciplines (psychology, counseling, and social work) are actually supported by empirical evidence. Although distinct, there clearly is much overlap in recommended supervision practices across these three sets of guidelines, and that shared content exemplifies the current standards of supervision practice in mental health. Understanding the extent to which the research evidence supports these commonly endorsed supervision practices is needed to inform the theory and practice of supervision within graduate training programs, post-degree licensure procedures, and EBP training and implementation initiatives. A review of the evidence base for these supervision guidelines can also highlight widely recognized, potentially promising, practices that have not yet been empirically evaluated.

Study 1: Common Supervision Elements across Supervision Guidelines

Introduction and Aims

As described in detail above, the supervision guidelines from three of the largest mental health disciplines appeared to have many similarities, with some notable distinctions between the guidelines of these three primary disciplines. The primary goal of this initial study was to determine whether the primary mental health disciplines promoted similar or distinct models of supervision. We sought to identify common supervision elements that are recommended across disciplines (i.e., appear in at least two disciplines' supervision guidelines), as well as which elements are specific to just one discipline's set of guidelines. The supervision elements that could be considered "common supervision elements" across best practice supervision guidelines would then be examined as predictive factors for relevant supervision outcomes (e.g., therapist competence) in existing empirical literature.

Method

Across supervision guidelines or best practice recommendations from psychology (29 guidelines across 7 domains; APA, 2014), counselling (51 guidelines across 12 categories; ACES, 2011), and social work (22 best practices across 5 standards; NASW, 2013), we (another graduate student [K.S.] and I) first operationalized common supervision elements across these major guidelines (e.g., the creation of supervision contracts, fostering a collaborative relationship between supervisor and supervisee/therapist) through a series of consensus meetings. I then compared the two sets of themes to operationalize a final theme for each guideline or recommendation and came to consensus with the other doctoral student. Each element was operationalized based on examples provided in the three sets of guidelines (see Table 1).

Results

In total, we identified 17 common supervision elements as recommended across two or three sets of professional supervision guidelines (see Table 1). If an element was present in at least two of the three guidelines, it was considered a common-supervision element across best practice supervision guidelines. For example, all three sets of guidelines included reference to the supervisor setting expectations or creating a contract, which we referred to "supervisors set expectations/contract," and operationalized examples of this element as "Supervisors identify expected program competencies and performance standards, and assist the supervisee to formulate individual learning goals, Supervisor and supervisee make a written contract, Supervisor engages in informed consent with supervisee at onset of supervision." As an additional example, for the supervision element "self-awareness of supervisor competence," we operationalized examples as "Supervisor pursues continuing education, Supervisor solicits feedback from supervisee, Supervisors only provide supervision within the limits of their own competence." We found few supervision elements that were specific to only one discipline, although some guidelines further specified or highlighted aspects of the common themes we identified. For example, the counselling guidelines suggest that the time allotted for supervision and the format should be determined by supervisee need, rather than convenience or efficiency, which is related to, though more specific than, the common element of setting expectations for supervision format. Of note, we identified more supervision elements specific to social work than for psychology or counselling guidelines, which may reflect unique challenges encountered in social work settings.

As an example, one distinct recommendation from the social work guidelines was to ensure therapist safety (e.g., training the therapist to respond to workplace conflict, threats or assault, and non-violent strategies in responding to crisis situations), whereas the other two sets of guidelines spoke more to beneficence for the client than for the therapist.

Discussion

These initial findings provided qualitative support for the overlap among major supervision guidelines and best practice recommendations. The common supervision elements across guidelines include considerations for client welfare, formal evaluation of supervision processes or supervisee, modeling professionalism, setting expectations or creating a contract for the supervisory relationship, coordinating services with relevant professionals, and using technology to enhance the supervisory process. By operationalizing these elements, we could begin looking for these common supervision elements as predictive factors for meaningful supervision outcomes in a scoping review. The overlap in recommended supervision practices across the three major psychology disciplines suggest some consensual validity to the components of effective supervision, but do not provide direct empirical support for their efficacy. An important future direction was to examine the existing empirical evidence base for these common recommended supervision elements.

Study 2: Scoping Review

Introduction and Aims

After identifying common supervision elements in Study 1, the goal of the scoping review was to examine the extent to which these common supervision elements were explicitly studied in existing literature, how they were studied, and what outcomes were routinely examined. We determined how often studies examined supervision elements individually or combined with other supervision elements. We evaluated the frequency with which studies included control or comparison supervision groups. We also identified studies that evaluated relevant therapist-, and client-level outcomes: supervisory alliance between therapist and supervisor, therapist fidelity, therapist self-efficacy, therapist competence, therapeutic alliance between therapist and client, client perceptions of treatment (e.g., satisfaction), and client progress (e.g., symptom change). In addition to providing a count-based review of supervision outcomes, a primary aim of this preliminary scoping review was to determine the feasibility of a meta-analysis of the association between common supervision elements and therapist- and client-outcomes.

Method

Study selection. Based on Bernard and Goodyear's (1998) definition, we broadly defined clinical supervision as an intervention delivered by a supervisor/more advanced professional to a supervisee/trainee/more junior professional in mental health service. On May 9, 2020, we searched the PubMed database for all published English language studies that examined the relationship between any of these common supervision elements and one or more therapist- or client-level outcomes. Specifically, we searched for [clinical supervision] in the title, abstract or keywords. Additional studies were identified by searching the reference lists of articles returned

in the initial PubMed search results. A total of 1517 published English language articles were identified and screened based on title and abstract. To be included in the review, studies had to a) include a description of supervision format and content, b) include at least one previously identified common supervision element, c) measure at least one supervisor, supervisee (therapist), or client outcome, (d) examine the association between one or more common supervision elements and at least one outcome, and e) the supervision had to relate to mental health, psychological, psychosocial, or counseling services.

Data extraction. We (three doctoral students [J.A., E.C., B.M.], one masters student [R.R.], and I) developed a project codebook (see Appendix A) via iterative cycles of pilot coding and collaborative meetings to define codes capturing all relevant characteristics of the reviewed studies across the following domains: (a) study sample (e.g., post-graduate practitioners), (b) study setting (e.g., graduate training), (c) study design (e.g., between or within subjects), (d) supervision elements, and (e) supervisor, therapist and client outcomes (e.g., supervisor-supervisee or supervisory alliance; therapist competence; therapist-client or therapeutic alliance; client symptom reduction). Because supervision involves a triad (i.e., supervisor, therapist, and client; Inman et al., 2014), supervisor-, therapist-, and client-level outcomes were all included. Four cycles were repeated until all codebook development team members were satisfied that the codebook sufficiently captured all study characteristics necessary to achieve the project aims. Prior to data extraction, the final codebook was also reviewed and approved by all authors, with code definitions further clarified in response to feedback. The authors then applied the final codebook to complete data extraction. Each study was coded by one of the three coders. Routine coding team meetings were held to further clarify coding rules as

needed and ensure consistency of coding across coders. We recorded the count, or how often, each supervision practice was measured across studies, whether the above outcomes were measured and the direction of findings, client population (e.g., child, adult, or mixed), supervision setting (including whether community mental health was specified), study design (cross-sectional or longitudinal, within or between subjects), whether the supervision took place in the context of graduate training, the type of therapy provided (including whether that treatment was an identified evidence-based treatment such as CBT), and whether supervision process or elements were also measured (including whether that measure had been created by the study authors). We report the direction of each observed relationship (e.g., positive, negative, null, mixed [i.e., conflicting results]) between supervision practices and outcomes (see Appendix A for study codebook).

Results

All published studies meeting the above criteria were included: both experimental and observational designs; pre-graduate and post-graduate supervision; all mental health disciplines (i.e., psychological, social work, or counseling services) and all contexts (e.g., graduate training clinics, private practice, mental health agencies, hospitals). Ultimately, 26 articles covering 28 studies were included in the final review. One article (Thew et al., 2019) reported multiple studies, two of which met inclusion criteria. Another article (i.e., Bearman, Schneiderman, & Zoloth, 2017) evaluated two supervision groups that each independently met inclusion criteria and thus were treated as separate studies for the purpose of this review. See Figure 1 for PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart diagram.

Of the 28 included studies, 50.0% (*n* = 14) used within-group designs and 21.4% (*n* = 6) used cross-sectional designs. Seventy-one percent of studies (*n* = 20) evaluated supervision of clinicians implementing EBPs (e.g., motivational interviewing, cognitive-behavioral therapy), and most were conducted at least partially within community mental health settings (64%, *n* = 18). Sample sizes (i.e., therapists, supervisors, or therapists and supervisors together) ranged from a single therapist in a case study design to 429 therapists (*M* = 79.3; *SD* =121.5). About one third of the studies (35.7%, *n* = 10) evaluated therapists in graduate training. About one third of studies included just individual supervision (32.1%, *n* = 9) or group supervision (39.3%, *n* = 11), with fewer including both individual and group supervision (14.3%, *n* = 4). The majority (82.1%, *n* = 23) of studies were conducted in the context of a broader training effort (e.g., supervision course or formal EBP training).

Results from the initial review of supervision studies including common elements are detailed in Table 2. Fifteen of the 17 common supervision elements were included in at least one study, but no element was universally included in all studies, and no single study included all elements (range 1-11; see Table 2 for further detail). The supervision elements included most often were the provision of general feedback to the supervisee (*n* = 20; e.g., feedback provided based on session recordings during supervision meetings), technology considerations (*n* = 13; using novel technologies to conduct supervision), supervisors setting expectations/contract (*n* = 9; e.g., collaborative agenda setting), and a collaborative supervisory relationship (*n* = 9; e.g., supervisor seeking feedback from the therapist). Fewer studies included self-awareness of supervisor competence (*n* = 8), supervisor encourages supervisee evaluation or goal setting (both *n* = 5), or documentation of supervision (*n* = 2). . The

least referenced supervision elements were beneficence, modelling professionalism, providing multicultural supervision, maintaining appropriate relationships, and supporting client advocacy (all $n = 1$).

Most studies included multiple supervision elements ($M = 3.3$, $SD = 5.7$), but few measured any of the included elements independently (25.0%, $n = 7$), generally preventing evaluation of the effect of specific supervision elements on outcomes. Most studies included therapist-report outcome measures (71.4%, $n = 20$), followed by client- (32%, $n = 9$), observer- (28.6%, $n = 8$), and supervisor-report (25.0%, $n = 7$) measures.

We reviewed for supervision outcomes at supervisor-level, therapist-level and client-level. Therapist outcomes were the most commonly evaluated outcomes, including therapist competence ($n = 19$), therapist fidelity ($n = 10$), therapist-supervisor alliance ($n = 7$), and therapist self-efficacy (n=4). More studies evaluated client change in symptoms or functioning ($n = 10$), than client-therapist alliance ($n = 3$) or client perceptions of treatment (e.g., perceived helpfulness, $n = 2$). We also identified 10 studies reporting supervisor-level outcomes (e.g., supervisor adherence to supervision model, supervisor competence, supervisory style).

The associations were variable across outcomes, though several supervision practices were consistently positively linked with outcomes across several domains and multiple studies (see Table 2). All associations between supervision practices and therapist self-efficacy were positive, as were nearly all associations with therapeutic alliance, with the exception of one null finding. Associations for therapist competence and supervisory alliance were predominantly positive, with some null and mixed findings for both outcomes. Findings for therapist fidelity and client progress were much more variable, with somewhat equal positive, null, and mixed findings.

Notably, no studies found negative associated between supervision practices and these outcomes, a potential artifact of publication bias towards positive results. Finally, these studies were published between 1978 and 2020, with no studies from the 1980s and one from the 1990s. More than half were published in 2011 or later, and of those, the majority were published after 2015.

Discussion

Although supervision is widely considered an essential feature of training in psychology, counseling, and social work, what exactly makes for effective supervision remains understudied. While supervision has been examined empirically, the results of this scoping review of published research found that the common supervision elements lacked a clear, consistent empirical evidence base. We identified 28 studies that had investigated the relationship between at least one of these common supervision elements and a relevant supervisor, therapist, or client outcome. Of note, we were able to identify the presence of these common elements in the description of supervision even without explicit reference to published guidelines and in studies published before the guidelines were publicly available. Most employed within-groups observational designs that cannot support causal inferences about effects of the studied supervision practices. Furthermore, almost none of the studies examined supervision elements independently, making it difficult to discern how much any individual element may contribute to the observed effect.

There was substantial variability across the common supervision elements in how often they were included and measured in supervision studies. Some common supervision elements were included in multiple studies (e.g., supervisor provision of formal evaluation and informal feedback), but many were examined in only a few studies (e.g., encouraging supervisee self-evaluation, modeling ethical practices). Five

common supervision elements were explicitly measured by only a single study each (e.g., providing multicultural supervision). Some common supervision elements may be so ubiquitous that study authors simply assume they are standard part of all supervision and thus neglect to describe or measure them. However, without clear operational definitions and measurement of supervision practices included in studies, it is hard for the field to learn and grow. Indeed, very little is known about what common supervision elements are part of routine or standard supervision in any context, much less which elements are consistently associated with supervisor, therapist, and client outcomes.

There are several promising findings to highlight regarding the relationships between supervision elements and outcomes at the supervisor-, therapist-, and client-levels. First, several common supervision elements were associated with positive outcomes whenever examined: namely, fostering a collaborative supervisor-supervisee relationship (or supervisory alliance), documentation of supervision, adherence to established guidelines, encouraging therapist self-evaluation, modelling ethical practices, and coordinating among relevant professionals. These evidence-supported supervision practices are therefore well worth future study and consideration for inclusion in supervision provided during graduate and post-graduate training.

Second, many of the common supervision elements showed strong associations with therapist-level outcomes, including therapist self-efficacy, therapist competence, therapist alliance with their supervisor, and therapist alliance with their client(s). Such therapist-level outcomes may be critical intermediary outcomes that ultimately support or impede client outcomes. Intriguingly, associations between common supervision elements and therapist fidelity, a key intermediate outcome for

EBP training and implementation, were more mixed. It is not clear whether therapist fidelity is a more difficult outcome to achieve, is a more challenging construct to measure, or is actually less responsive to common supervision elements than the other therapist-level outcomes.

Findings from the current review may also have implications for routine supervision in clinical training and practice as well. The consistent positive findings for several supervision practices may give supervisors increased confidence in the value of following these guideline recommendations. Although only examined in five studies, overall adherence to the reviewed guidelines was positively associated with every outcome, suggesting that following these guidelines may well lead to beneficial therapist and client outcomes.

A disheartening finding was that a minority of studies examined effects of supervision practices on client outcomes. As noted above, client outcomes are often considered to be the true test of supervision effectiveness. Across all three sets of guidelines, a consistent theme was that a primary aim of supervision is to benefit the client and to prevent harm to the client. Despite this putative link between supervision and client well-being, we found little empirical examination of this relationship and even less empirical support for this relationship in the published literature. Future research measuring client symptoms and client perspectives is needed to determine the impact of supervision practices on these ultimate outcomes.

While there were several issues with the methodological rigor of the studies in the current review, as has been noted in previous reviews (e.g., Ellis et al., 1996; Kuhne et al., 2019), we did find a growing body of empirical literature examining clinical supervision. We echo previous calls (e.g., Simpson-Southward, Waller, & Hardy, 2017) for researchers to describe in detail the content, format, and processes of

supervision whenever supervision is studied or delivered in the context of a research protocol. We also advocate for future research to incorporate between-subjects designs or randomization, and to examine supervision elements individually, perhaps through dismantling studies, to more convincingly demonstrate causal associations between specific supervision elements and various outcomes.

Our initial count-based scoping review of published literature indicates a growing supervision evidence base; almost twice as many studies were identified in the years 2010-2020 than in the years 2000-2009. Several common supervision elements were included in supervision, and showed positive associations with relevant outcomes. This preliminary review also shows much additional work to be done. More thorough descriptions and operationalizations of supervision practices are needed in these studies to facilitate review and synthesis of the literature. Prospective, randomized studies comparing supervision elements and packages are critical to demonstrate a causal association between supervision and relevant outcomes. Ultimately, findings from the current review (Tugendrajch et al., 2021) suggest that more formal quantitative analysis is warranted to synthesize existing research and better understand the impact of common supervision elements on relevant outcomes.

Study 3: Meta-Analysis

Introduction and Aims

After determining it was possible to examine the identified common supervision elements in previously published studies of clinical supervision, we could move forward in conducting a systematic review and meta-analysis of quantitative supervision findings. The scoping review illustrated that many published studies had evaluated meaningful therapist outcomes following the receipt of supervision containing one or more common supervision elements. While client outcomes may well be the ultimate test of clinical supervision effectiveness, we found few studies actually examining client symptom and functioning change as outcomes. Instead, supervisee (i.e., therapist-level) outcomes were most commonly examined in existing research. This focus to date makes some sense given that supervision is presumed to exert its effect on clients through its effect on supervisees/therapists. As such, for this meta-analysis, we chose to focus on therapist-level outcomes.

Method

Meta-analysis study design and procedures. We followed the PRISMA guidelines for reporting and conducting systematic reviews and meta-analyses (as outlined in Moher et al., 2009). The PRISMA diagram for the meta-analysis is included as Figure 2.

Inclusion/eligibility criteria. We included published and unpublished studies. Similarly to the criteria for the preliminary review, we required studies selected for the meta-analysis to include a) supervision related to mental health, psychological, psychosocial or counseling services, b) with a description of supervision format and content that involves at least one common supervision element, and c) at least one supervisee (therapist) measure of fidelity, competence, or self-efficacy measured pre-

and post- supervision. We continued to define clinical supervision as an intervention delivered by a supervisor/more advanced professional to a supervisee/trainee/more junior professional in mental health service. Given the lack of controlled trials or between-subjects' design in the existing supervision literature identified in the preliminary scoping review, we included studies with both within- and between subjects' designs, as well as case studies. When a study met the majority of these identified criteria, but the therapist variable of interest was not sufficiently described or reported, we contacted the authors to obtain additional information to determine eligibility for analysis.

Search/Data sources and study selection/Search procedure. We broadened our search procedures from what was done for the preliminary scoping review, which included only English-language published studies within PubMed, with "clinical supervision" in the title or abstract. For the present meta-analysis, we searched PubMed with the following terms in October 2020: "((((Psychotherapy/education[MAJR]) OR (Psychotherapy/standards[MAJR])) OR (Teaching/organization and administration[MAJR])) OR (Psychology, Clinical/education[MAJR])) AND (supervisor[Title/Abstract] OR supervisee[Title/Abstract] OR supervision[Title/Abstract] OR supervising[Title/Abstract] OR supervised[Title/Abstract])." We also searched PsycINFO in June 2021 to include both published articles and unpublished theses and dissertations. We searched PsychINFO with a range of title, abstract and key words indicating psychotherapy supervision until we settled upon the following search terms that yielded under 5000 articles. The search terms were "TI (supervision or supervising or supervisory or supervisor or supervisee) AND AB (supervision or supervising or supervisory or supervisor or supervisee) AND TX (competence or

competency or competencies or skills or knowledge or self-efficacy or efficacy or adherence or fidelity).” We included studies that took place in the context of graduate training, therapy training or implementation efforts across multiple treatment settings, and across different disciplines (e.g., counseling, clinical psychology, psychiatric nursing) if the supervision was delivered to professionals providing mental health services (i.e., not lay-people receiving supervision). Although the three supervision best practice guidelines were written in the last decade, we had no restriction on date and screened studies from the beginning of the database as earlier studies may have influenced recommended supervision practices. We also solicited published and unpublished research from several relevant electronic listservs (i.e., Dissemination and Implementation Special Interest Group of the Association of Behavioral and Cognitive Therapies, APA Division 53 the Society of Clinical Child and Adolescent Psychology).

Literature searching was conducted by the lead author with support from university librarians (i.e., Rebecca Graves. for PubMed, Kimberly Moeller for PsycINFO). We identified 768 articles from PubMed and 2911 articles from PsychINFO. The search terms for PubMed and PsycINFO searches are included as Appendix B and the PRISMA diagram (Figure 2) shows the search and screen results. The total number of articles screened without duplicates was 3676. All studies were double-screened by a team consisting of the lead author and several undergraduate research assistants (i.e., KAB, KB, KC, AK, TM, HP, ES, NW) trained in the screening procedures. Of note, the lead author did not screen all studies once the undergraduate assistants were able to screen consistently with the lead author and with one another. All screening discrepancies were discussed via Zoom consensus meetings or written correspondence (i.e., email, Slack).

Coding. Coding was conducted by another team, consisting of the lead author and two trained graduate research assistants (J.A. & R.P.), including one who had previously coded for the preliminary scoping review (i.e., J.A.). As with screening, all studies were double-coded (i.e., coded by at least two people) and all coding disagreements were resolved via consensus. The full codebook is included as Appendix C. We coded for study design type (i.e.., between group, within group, cross-sectional, longitudinal), setting (i.e., sample size/description [i.e., race, age, gender, disciplines, highest degree earned for supervisors and/or therapists, supervisor qualifications or training], graduate training, private practice, community mental health, medical/hospital, part of implementation/training effort), client population (i.e., children, adults), treatment (i.e., non/evidence-based practice), supervision format (i.e., individual, group, group size, average length of supervision meetings, duration of supervision meetings, frequency of supervision meetings), supervision content covered (i.e., administrative issues, EBPs, role-playing/active learning strategies, didactic learning), common supervision elements (i.e., presence and description of supervision common elements), and whether the common element had been analyzed independently. For outcomes, we coded reporter (i.e., supervisor, supervisee/therapist, client, observer) and the following therapist outcome measure types: therapist fidelity (i.e., adherence to a specified treatment protocol), therapist self-efficacy or confidence, and therapist competence (i.e., knowledge, skills). We also coded for study/methodological quality variables, country where the study was conducted and whether the study had been published. To examine consistency across coders, we calculated percent agreement between coding pairs for all codes. Average percent agreement for each code is presented alongside the code in Appendix C. Average agreement across all codes was 86.79% (range 55.17-100.00%). Average

percent agreement between J.A. and S.T. was 81.81%, average percent agreement between J.A. and R.P. was 87.37% and average percent agreement between R.P. and S.T. was 90.80%.

Calculating effect sizes. We calculated a standardized mean difference (SMD) for each relevant outcome measure from pre-supervision baseline to post-supervision. To calculate an effect size (Cohen, 1988), we divided the difference between the pre and post scores by the pooled standard deviation, using Hedge's *g* correction for small sample sizes using the *escalc* function in the R metafor package (Viechtbauer, 2010). We requested means and standard deviations from all authors if not already reported in the article. Percent agreement for coding the means and standard deviations is also included in the measure-level codebook in Appendix C. Without that information, the study was not included in the meta-analysis, but we kept track of these studies for our overall discussion of supervision literature. By the time of this write-up, we had received means and standard deviations from 4 out of 30 authors contacted for this information.

Proposed Analyses

Analytical strategy. We used a random effects, multilevel meta-analytic approach. We chose random over fixed effects under the assumption that the results of the meta-analyses should generalize to a larger population and that variability across studies is due to actual differences across supervision study outcomes, rather than exclusively error or noise between studies. Further, traditional meta-analysis assumes independence of effect sizes, which is not appropriate when multiple effect sizes from the same sample or study are included in analyses. The multilevel approach allowed us to address sampling variation in effect size (Level 1), within-study variation (Level 2), and between-study variation (Level 3; Van den Noortgate et

al., 2013). This approach also allowed us to estimate the overall magnitude of pre-post effects of supervision practices on identified therapist outcomes. We also conducted a test of homogeneity to determine whether there were significant differences between study effect sizes. We used a combination of software packages to conduct these analyses (i.e., R, SPSS).

Meta-Analysis Aims and Hypotheses

Aim 1: Evaluate whether and to what extent therapist outcomes improve following supervision incorporating at least one common supervision element.

Hypothesis 1.1: Overall, therapist outcomes improve significantly pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

Hypothesis 1.2: Therapist fidelity improves pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

Hypothesis 1.3: Therapist competence improves pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

Hypothesis 1.4: Therapist self-efficacy improves pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

Aim 2: Evaluate associations between hypothesized moderators and therapist outcomes. Our a priori hypothesized moderators to evaluate were the following: whether the study was part of a formal psychotherapy or supervision training effort where improvements in therapist competence might be expected due to the training (even without supervision), the amount or dose of supervision received, the number of

common supervision elements incorporated into supervision, the inclusion of each common supervision element and the inclusion of other specific teaching or training practices within supervision (i.e., role-play, focusing on fidelity to a given treatment model, reviewing treatment techniques, administrative issues).

Hypothesis 2.1: Receiving training during supervision is associated with greater pre-post improvement in therapist outcomes than supervision without training.

Hypothesis 2.2: Higher supervision dose or amount is associated with greater pre-post improvement in therapist outcomes.

Hypothesis 2.3: Inclusion of more common supervision elements is associated with greater pre-post improvement in therapist outcomes.

Hypothesis 2.4: Inclusion of each common supervision element is associated with greater pre-post improvement in therapist outcomes than supervision without that element included.

Hypothesis 2.5: Including role-play in supervision is associated with more improvements in pre-post therapist outcomes than supervision that does not include role-play.

Hypothesis 2.6: Focusing on fidelity to a specified treatment model in supervision is associated with more improvements in pre-post therapist outcomes than supervision that does not focus on fidelity.

Hypothesis 2.7: Reviewing treatment techniques in supervision is associated with more improvements in pre-post therapist outcomes than supervision that does not explicitly review treatment techniques.

Hypothesis 2.8: Including administrative issues in supervision is associated with less pre-post improvements in therapist outcomes than not including administrative issues.

Results

Included Studies. Of the 3776 records screened, 3330 were excluded by title or abstract, and 346 full text articles were screened for eligibility. Of the full-text articles screened, 287 were excluded for the following reasons: supervision was not relevant to mental health professionals (n = 5), supervision content or format was not described (n =103), described supervision did not include a common supervision element (n = 20), the study did not assess a relevant therapist outcome (n = 141), and there was no pre-post evaluation of a relevant therapist outcome (n = 23). Of the 59 remaining articles that appeared to meet all eligibility criteria, 3 articles contained study data already included from another article (i.e., it was the same study and sample, with no additional therapist outcomes reported), and 2 articles described single case design studies. Another 26 articles were missing the means and standard deviations needed for calculating pre-post effect sizes. The data were unavailable for these 26 studies for the following reasons: no contact information for study first author was found online or in the article (n = 5), no response from first author (n = 11), first author reported that relevant pre-post data had not been collected (n = 4), first author reported the data was no longer available to access (n = 3), first author responded to initial request but then did not provide needed data (n = 5), and the first author had passed away (n = 1). After these attempts to contact study first authors for data needed to calculate effect sizes, the total sample is 29 studies with 46 supervision groups (including 38 groups receiving supervision with at least one common supervision element, 1 group receiving supervision with no common supervision elements, and 7 groups receiving no supervision, see Figure 2 for PRISMA diagram).

Study Characteristics

Table 3 includes relevant descriptive information for the 29 studies with 38 supervision groups with at least one common supervision element that were included in the meta-analysis, 1 group that received supervision with no common supervision elements, and 7 groups that received no supervision during the study period. Across these 7 groups, 6 received various forms of training in lieu of supervision (i.e., workshop training, access to training materials, access to a manual) and one group received no training nor supervision (i.e., the assigned condition was delayed training) during the study period (i.e., Rakovshik et al., 2016). Of the 29 studies, 34.48% (n = 10) included graduate trainees, 62.07% (n = 18) were part of a formal treatment training effort (e.g., workshops, courses, online training), 58.62% (n = 17) were an entirely United States sample (with representation from Europe, Australia, and Asia in the other studies), and 93.10% (n = 27) were published journal articles while 6.90% (n = 2) were dissertations. The studies were all published between 2002 and 2020, with nearly half (n = 14) published between 2001 and 2010 and half (n = 15) published between 2011-2020.

The number of therapists in each study ranged from 4 to 168 (*M* = 38.59, *SD* = 40.55). On average, therapists were 37.78 years old (*SD* = 6.99), 51.81% female (*SD* = 34.45 range = 64.29-100.00%), and 39.04% Caucasian (*SD* = 34.66, range = 28.87-82.00). Of note, 14/29 studies reported therapist age, 24/29 reported gender, and only 11/29 reported racial demographics. Therapist disciplines across studies included counseling (n = 14), psychology (n = 12), social work (n = 7), psychiatry (n = 6), certificate (n = 1), paraprofessionals (n = 1), and other disciplines (n = 9). Therapist highest degree earned included master's (n = 9), bachelor's (n = 8), doctorate (n = 8), associate's (n = 4), high school or equivalent (n = 3), and other degrees (n = 4). Three studies did not report therapist discipline and ten studies reported no degree

information for participating therapists. Of the two studies that reported therapist licensure, on average 76.00% of therapists were licensed (SD = 33.94, range = 52.00-100.00%) and several studies did not report licensure as these samples were primarily graduate trainees who would be ineligible for licensure. Of the 11 studies that reported therapist years of therapy experience, therapists had an average of 5.68 years of experience (SD = 3.39, range 1.00-10.08).

The number of supervisors in each study ranged from 1 to 26 (*M* = 5.32, *SD* = 6.70) and was reported in 22/29 studies. In about half (n = 12) of the 22 studies with available data, 1 or 2 supervisors provided the supervision for one or more supervision groups. Only 4 studies reported supervisor age and across those studies, supervisors were an average of 41.40 years old (*SD* = 5.89). Only 11 studies reported supervisor gender, and across those studies, on average supervisors were 27.49% female (*SD* = 40.85) with a range from 0 to 100%. Of the 5 studies that report supervisor race, supervisors were 15.35% Caucasian (*SD* = 33.35) on average, with a range of 0 to 75%. Across studies, supervisors' disciplines included psychology (n = 11), counseling (n = 3), and other disciplines (e.g., nursing, n = 4). Supervisors' highest degree earned included doctoral degrees (n = 12), masters (n = 3) and other degrees (n = 2). Sixteen studies did not report supervisor discipline and only around half of the 29 studies reported supervisor highest degree earned. Of the 4 studies that reported supervisor licensure, on average supervisors were 99.00% licensed (SD = 2.00, range 96.00-100.00%). Average supervisor years of supervision experience was 14.98 years (SD = 11.54, range = 2.00-36.00). Across all studies, supervision training, qualifications, or credentials prior to beginning the study was required for some supervisors (n = 9), alluded to but not specified (n = 12), not required (n = 5) or not reported (n = 3).

Only 6 studies included child or adolescent clients (including 2 studies that also included adult clients), 20 specified adult clients, and 5 studies did not specify client age. Approximately half (n= 14) of the studies used between group designs; of those, 11/14 randomly assigned therapists to supervision format. The methods of randomization were described in 4 studies. Outcome assessors were blinded to participant's supervision group assignment in 11 studies, supervisors were blinded to study hypotheses in 2 studies, therapists were also blinded to study hypotheses in these same 2 studies. In 5 studies, therapists received different amounts of therapy training, 8 of 17 studies with therapist dropout reported intent-to-treat analyses (12 studies reported no therapist dropout between measurements), and potential researcher allegiance was present (e.g., co-developing the therapy and/or supervision method used in the study) in 20 studies.

The following common supervision elements were included in the 38 supervision groups: general feedback (n = 33), technological considerations (n= 32), formal evaluation (n = 14), self-awareness of supervisor competence (n = 9), supervisor encourages supervisee self-evaluation (n = 9), setting expectations (n = 7), beneficence for the client (n = 4), adherence to established guidelines (n = 4), collaborative supervisory relationship (n = 3), goal setting (n = 3), modeling ethical practices (n = 1), and providing multicultural supervision (n = 1). Most supervision groups contained more than one common supervision element (M = 3.10, SD = 1.37, range 1-7). The following common supervision elements were not explicitly included in any of the identified studies: coordination among relevant professionals, modeling professionalism, maintaining appropriate relationships, documentation of supervision, or supporting client advocacy.

For the 38 supervision groups with at least one common supervision element, supervision dose ranged from 180 to 8645 minutes (M = 1485.04, SD = 1508.12). About one third (n =15) ranged from 0-1000 minutes, another third ranged from 1001-2000 minutes (n = 14), 5 ranged 2001-3000 minutes, 2 ranged 3000-4000 minutes, and a single study was above 8000 minutes. When possible, supervision dose was calculated by multiplying the average number of minutes spent in supervision each week by the average number of weeks supervision was delivered. When supervision meeting time was not given, we estimated that the average number of minutes for supervision meetings was 60 minutes. One study (i.e., Milne & James, 2002) provided so little information about supervision meeting time or duration that we were not able to calculate a supervision dose for this supervision period. Across studies, there was a mix of group supervision (n =13), individual supervision (n =12), and both group and individual supervision provided (n = 4). The most common therapy provided was CBT or cognitive therapy (n = 11), though several studies did not report what kind of therapy was delivered (n = 9). Three studies looked at motivational interviewing, two studies included psychodynamic therapy, two studies referenced other treatments (i.e., psychoeducational groups, family intervention and psychosocial intervention), and one study evaluated CBT and psychodynamic therapy techniques together.

We calculated pre-post effect sizes for all 46 groups (i.e., supervision with a common supervision element, supervision without a common element, no supervision) and all therapist outcome measures that had means and standard deviations reported. This yielded 144 effect sizes total: adherence (n = 63), competence (n = 69), self-efficacy (n = 12). For just the supervision groups with at least one common supervision element (n = 38), the number of effect sizes were as

follows: adherence (n = 45), competence (n = 54), self-efficacy (n = 11), total (N = 110). Within the single supervision group with no common element (n =1), we had 5 effect sizes: adherence (n = 3), competence (n = 2), self-efficacy (n = 0). Finally in the no supervision groups (n = 7), we had 29 effect sizes: adherence (n = 14), competence (n = 14), self-efficacy (n = 1).

Meta-Analyses Results

With support from a consulting statistician (T.S.), we ran univariate models and multilevel models which were nearly identical. In meta-analyses, it is convention to report on the simpler models in this case and so those results will be presented here. As planned, we used Becker's (1998) and Hedge's *g* for standardized mean pre-post difference scores (*g* = Y-X/*s*) using the metafor package in R (Viechtbauer, 2010).

$$g = \frac{\bar{Y} - \bar{X}}{s}, \qquad \delta = \frac{\mu_Y - \mu_X}{\sigma}.$$

The primary meta-analysis dataset contained a total of 110 effect sizes across the 29 studies (54 across 20 studies for competence, 45 across 8 studies for adherence, 11 across 6 studies for self-efficacy) with an average of 3.79 (range 1-14, median = 2) effect sizes per study.

Given that most studies reported more than one effect size and 12 studies contained multiple supervision groups, robust variance estimation (RVE) was used (Tanner-Smith & Tipton, 2014) to account for within study dependence in estimating the meta-analysis models. We used the R package robumeta (Fisher et al., 2017), including Tipton's (2015) correction for small sample bias. To calculate each pooled effect size estimate we estimated a random effects model, and mixed-effects model for the moderator analyses. A more stringent alpha level was used to better control the Type I error rate when the adjusted degrees of freedom for the *t* test statistic was less

than four, as recommended by Tanner-Smith and colleagues (2016). For RVE, I^2 can be used to measure the heterogeneity beyond sample differences and a value of 50-70% would be sufficient heterogeneity to conduct moderator analyses (i.e., moderate to large, Higgins & Thompson, 2003).

Aim 1: Evaluate whether and to what extent therapist outcomes improve following supervision with at least one common supervision element.

Hypothesis 1.1: Overall, therapist outcomes improve significantly pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

We first ran a simple main effects model to determine whether supervision with one or more common supervision elements was associated with pre-post improvements in therapist outcomes on average. Across the 38 supervision groups with at least one common supervision element, the pre-post effect size/pooled random effects point estimate was 0.787 (t = 6.75, p <.001), suggesting that receiving supervision with common supervision elements is associated with therapist improvements. The results also indicated sufficient heterogeneity to conduct additional moderator analyses (I^2 = 80.60%). See table 5 for effect sizes for all supervision conditions (i.e., supervision with at least one common element, supervision with no common element, no supervision).

Hypothesis 1.2: Therapist fidelity improves pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

For therapist adherence, the pre-post effect size/pooled random effects point estimate was 0.509 (t = 2.27, p = 0.0625), suggesting that receiving supervision with common supervision elements was not significantly related to improvements in

therapist adherence. However, the high heterogeneity indicated suggested that additional moderator analyses would be appropriate to conduct (I^2 = 87.69%).

Hypothesis 1.3: Therapist competence improves pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

For therapist competence, the pre-post effect size/pooled random effects point estimate across the 38 groups was 0.879 (t = 6.12, p <0.001), suggesting that receiving common elements supervision was significantly related to improvements in therapist competence. Additionally, the large heterogeneity indicated suggested that additional moderator analyses would be appropriate (I^2 = 78.22%).

Hypothesis 1.4: Therapist self-efficacy improves pre to post supervision, when supervision contains one or more common supervision elements from best practice guidelines.

For therapist self-efficacy, the pre-post effect size/pooled random effects point estimate across the 38 groups was 0.848 (t = 11.5, p <0.001), suggesting that receiving common elements supervision was significantly related to improvements in therapist self-efficacy. However, the small sample size (df = 3.81) suggests that these results should be interpreted with caution and the small heterogeneity suggest that additional moderator analyses may not be appropriate (I^2 = 19.58%).

Although we had no a priori hypothesis about different effects of supervision on the 3 therapist outcome types, we also ran a one-way ANOVA to examine potential differences across the 3 therapist outcome types: fidelity, competence, and self-efficacy. The one-way ANOVA indicated no significant differences among the 3 therapist outcome types for the supervision groups with at least one common supervision element F (2,107) = 0.356, p = 0.701. See Table 6 for overall effect sizes,

subgroup effect sizes and moderator analyses. See Figure 3 for box plot of mean effect sizes across the 3 outcome types.

Aim 2: Evaluate potential associations between identified moderators and therapist outcomes. Our a priori hypothesized moderators to evaluate were whether the study was part of a formal psychotherapy or supervision training effort where improvements in therapist competence would be expected, the amount or dose of supervision received, the inclusion of common elements and other specific practices in supervision. See Table 6 for effect sizes and moderator analyses.

Hypothesis 2.1: Receiving training during supervision is associated with greater pre-post improvement in therapist outcomes than supervision without training.

Contrary to our hypothesis, there was no significant difference in mean effect sizes for supervision groups receiving training during the supervision period versus those that did not receive training during supervision, $t(108) = 0.25$, $p = 0.80$. There were significant pre-post improvements in therapist outcomes for the subsample of supervision groups receiving training during the supervision period ($g = 0.653$, $t = 6.64$, $p < 0.001$) and for the subsample of supervision groups that did not receive training during the supervision period ($g = 0.76$, $t = 3.64$, $p = 0.005$).

Hypothesis 2.2: Higher supervision dose or amount is associated with greater pre-post improvement in therapist outcomes.

Supervision dose was not significantly related to pre-post difference effect sizes across all therapist outcomes, ($F = 2.638$, $t = -1.624$, $\beta = -0.158$, $p = 0.107$).

Hypothesis 2.3: Inclusion of more common supervision elements is associated with greater pre-post improvement in therapist outcomes.

Contrary to our hypothesis, the inclusion of more common supervision elements was related to *smaller* pre-post difference effect sizes across all therapist outcomes, (F = 3.952, t = -1.988, β = -0.188, p = 0.049).

Hypothesis 2.4: Inclusion of each common supervision element is associated with greater pre-post improvement in therapist outcomes than supervision without that element included.

We also examined each common supervision element on its own. Results of these analyses are detailed in Table 7. Contrary to hypothesis, none of the common supervision elements were associated with significantly greater pre-post improvements in therapist outcomes compared to not including the element. In fact, for the following supervision elements, supervision *not* containing this element was associated with greater pre-post improvements in therapist outcomes: 1) beneficence for the client $t(108) = 2.07$, $p = 0.0406$; 6) technology considerations $t(108) = 3.30$, $p = 0.0013$; and 17) general feedback $t(108) = 2.80$, $p = 0.0061$. One common supervision element (i.e., 11) providing multicultural supervision) could not be analyzed as there was only one effect size that corresponded with this element.

Hypothesis 2.5: Including role-play in supervision is associated with more improvements in pre-post therapist outcomes than supervision that does not include role-play.

The inclusion of role-play or active learning techniques in supervision did not moderate the association between supervision and therapist outcomes $t(108) = 0.18$, $p = 0.24$. The subsample of supervision groups including role-play or active learning strategies in supervision showed significant pre-post improvements across therapist outcomes (g = 0.519, t = 3.97, p = 0.002), as did subsample of supervision groups that did not include role play (g = 1.04, t = 3.74, p = 0.005).

Hypothesis 2.6: Focusing on fidelity to a specified treatment model in supervision is associated with more improvements in pre-post therapist outcomes than supervision that does not focus on fidelity.

There were no significant differences in pre-post difference effect sizes for supervision groups that focused on fidelity and for those that did not $t(108) = 0.28$, $p = 0.30$. Focusing on fidelity to a specified treatment model in supervision was associated with significant pre-post improvements across therapist outcomes ($g = 0.651$, $t = 5.01$, $p < 0.001$). However, significant pre-post improvements across outcomes also occurred in the groups that did focus on fidelity during supervision ($g = 0.76$, $t = 3.64$, $p = 0.005$).

Hypothesis 2.7: Reviewing treatment techniques in supervision is associated with more improvements in pre-post therapist outcomes than supervision that does not explicitly review treatment techniques.

There were no significant differences in pre-post effect sizes for supervision groups that focused on treatment techniques and for those that did not $t(108) = 1.49$, $p = 0.14$. The subsample of supervision groups that incorporated review of treatment techniques saw significant pre-post improvements in therapist outcomes ($g = 0.261$, $t = 4.67$, $p = 0.005$), as did the subsample of supervision groups that did not incorporate review of treatment techniques ($g = 0.261$, $t = 4.67$, $p = 0.005$).

Hypothesis 2.8: Including administrative issues in supervision is associated with less pre-post improvements in therapist outcomes than not including administrative issues

Moderator analyses could not be conducted for administrative issues because no study explicitly specified that administrative issues were discussed in supervision (n = 0).

Discussion

In Study 3, we examined 29 supervision studies that included at least one pre and post measurement of a meaningful therapist outcome (i.e., adherence, competence self-efficacy) following the provision of supervision with at least one common supervision element to mental health professionals. We focused on therapist outcomes in this systematic review and meta-analysis for two primary reasons: (1) the primary purpose of supervision is to improve therapist functioning – to yield competent therapists who can then practice independently without ongoing oversight, and (2) therapist-level outcomes were the outcomes most often evaluated and most consistently positively related to supervision in our preliminary review. We included self-report measures (e.g., therapist confidence providing therapy), performance-based measures (e.g., multiple choice knowledge tests to assess competence), and observational measures (e.g., count-based measured of therapist fidelity to a specified treatment protocol).

As in the preliminary scoping review, we were able to identify the presence of most of the common supervision elements we had initially identified, even without specific reference to the published guidelines and even in studies conducted before these guidelines had been published. We found that therapists receiving supervision with at least one common supervision showed significant pre-post improvements when collapsed across all therapist outcomes. We also saw significant pre-post improvements for the specific outcomes, therapist competence and self-efficacy. However, we found no significant pre-post improvement for the specific therapist adherence outcome.

Contrary to our hypothesis, and surprising given the widespread endorsement of these supervision elements, the inclusion of these common supervision elements

does not appear to result in more competent or confident therapists than another other supervision content. We found that receiving more common supervision elements actually predicted lower pre-post improvements, compared to receiving supervision with fewer common elements. Perhaps the common elements, while effective in isolation, can become diluted when combined with too many other elements. However, when we looked at each individual common supervision element, we found that none predicted greater pre-post improvements in therapist outcomes, compared to the pre-post outcomes achieved for supervision not containing each element. Three of the common supervision elements, recommended as best practices for supervision, were actually predictive of less pre-post improvement in therapist outcomes: 1) beneficence, 6) technology considerations, 7) general feedback). Additionally, we found no significant differences in pre-post improvements across therapist outcomes for any of our other hypothesized moderators. Together, these findings unfortunately suggest that the elements that are widely purported to make for better therapists are not clearly supported by available empirical research at this time.

Limitations of the Literature

We encountered several challenges in reviewing the available supervision literature. In the current study, we identified 46 groups to evaluate, including 38 supervision groups receiving at least one common supervision element, one supervision group receiving supervision with no common supervision element, and 7 groups that received no supervision during the study period. Across groups receiving no supervision, the pre-post effect sizes were in the medium range, suggesting that even groups not receiving supervision improved over time across these meaningful therapist outcomes. However, there are important caveats to note with this finding. This subsample is quite small and many of the groups that did not receive supervision

still received training or had access to training materials throughout the study period, so it is somewhat difficult to identify what exactly has contributed to these improvements in this small sample. While we have worked to identify and operationalize as many common supervision elements as possible based on best practice recommendations, none of the studies we identified studied a single supervision element in isolation, which makes teasing apart the influence of individual elements challenging. Many studies also evaluated pre-post outcomes across a single supervision group and did not include a true control comparison group (i.e., a supervision group receiving supervision without a common supervision element), which did not allow us to compare supervision containing at least one common supervision element with supervision not containing one of these elements. Finally, it was challenging to identify supervision groups without a single supervision common element as most supervision descriptions included some reference to providing feedback to supervisees, the final supervision common element we identified, which meant we were not able to meaningfully compare the one "supervision as usual" group we identified to the dozens of supervision groups with at least one common element.

We also encountered challenges comparing measures across studies as few studies used the same measures to assess these therapist outcomes; in other words, most measures were developed for, and used in, just one study or by one research team. Four studies used the Cognitive Therapy Scale to evaluate cognitive therapy adherence, two studies used the Basic Skills Observation (a measure of basic counseling skills such as reflecting, active listening, demonstrating positive regarding), two studies used the motivational interviewing treatment integrity coding system (MITI) to assess MI adherence, and one research group used the same

Counseling Self-Estimate Inventory (a therapist self-efficacy measure) in two separate studies. Several studies used measures that were not clearly psychometrically validated or had not previously been used in research studies. Therapist competence was often evaluated with performance-based tests to assess knowledge and with observational methods to evaluate therapists providing services to actual clients, or with role-play demonstrations, but often with a measure either developed by the research team for the specific study or without clear empirical validation. Not surprisingly, measures of therapist self-efficacy were consistently gathered from therapist self-report, but this outcome was also not measuredly consistently across studies. Given the varied measurement of these therapist outcomes, a future direction would be to identify which measures of therapist adherence, competence, and self-efficacy have been psychometrically validated to use consistently moving forward in studies of therapist outcomes following the receipt of supervision.

There are many potential reasons why we may have had difficulty identifying clear associations between our hypothesized moderators and therapist outcomes. We were very surprised to find that supervision dose was not significantly related to therapist adherence or competence as we assumed more supervision would be beneficial across all outcomes. However, this variable was something we often had to estimate as the information was not always readily available in written the study description. There was also a high degree of variability in the dose of supervision provided (e.g., weekly supervision for graduate trainees for an entire academic year compared with training studies for practicing therapists where a maximum of 2-3 hours of supervision was allotted). Given the variability in how supervision was reported across studies, it is also possible that supervision may have contained additional supervision elements that were not explicitly mentioned or that other

factors contributed to therapist outcomes during the study period (e.g., client population, treatment setting, match between therapist and supervisee, whether supervision was required, etc.). Further, it was disheartening to observe that almost none of our other hypothesized moderators were significantly related to changes in therapist outcomes following supervision. Few of the common supervision elements were associated with significant change in these variables following supervision with these elements included. More concerningly, our findings suggest that the impact of supervision may be more significantly associated with therapist outcomes when certain elements are *not* included (i.e., beneficence for the client, technological considerations, and general feedback). However, as stated above, these common supervision elements were almost never studied individually (e.g., a supervision group that did not include beneficence may have included multiple other common supervision elements) and for many of the studies we evaluated, supervision took place in tandem with training in a specified treatment. As such, it does not feel appropriate to conclude that those three identified common supervision elements above should not be included in supervision as so much more information is needed. Additionally, these three outcomes encompass a wide range of practices; for example, technological considerations could mean listening to audiotapes of sessions before meeting with the therapist, using bug-in-the-ear technology for the supervisor to provide feedback in real time, or watching session videos with the therapist. As such, it is difficult to propose exactly what technological consideration should or should not be included in regular supervision.

We hoped that examining main effects for each common supervision element and moderators would provide some of the first empirical support for including the specific elements that positively predicted outcomes in routine supervision. However,

we were unable to identify a single element of supervision that is clearly supported by current scientific evidence. Of course, ethical standards preclude the delivery of psychotherapy by untrained and unsupervised therapists. As such, we are unable to determine whether any supervision is superior to no supervision, among untrained and unlicensed therapists. However, when we examined different components of supervision that have been consensually developed and espoused as best practice, we found no evidence that these supervision best practices were actually superior to other supervision practices. It is important to note that inclusion of identified supervision best practices, or common supervision elements, is so prevalent we identified only one supervision group in one study that did not include a common supervision element. As such, it is premature to state that the common supervision elements are not superior to other supervision practices; we simply have no compelling data to show they are more worthwhile than other supervision.

In this meta-analysis and review, we are starting to chip away at what practices in supervision may contribute to improvements in therapist outcomes following supervision. Again, within all these common supervision elements, there is room for flexibility and supervisor discretion (which may appeal to supervisors who see supervision as more of an art than a science). For example, formal supervisee evaluation could include completing competence assessments every three months with supervisees, but could also mean reviewing therapy tapes for adherence on a regular basis using a standardized rating scale. Clearly, there is also overlap between many of these elements, such the setting expectations and goal setting elements, which both speak to establishing a long-term supervisor-therapist/supervisee working relationship. We hope that these findings, however muddy, would still be well-

received by practicing supervisors, especially those already incorporating some of these common supervision elements in their current supervision.

Limitations of the Current Study

While Study 3 allowed us to explore therapist outcomes in more quantitative detail, there are some limitations to note. First, our results only include two databases, and a future systematic review could include other databases (e.g., Medline, ERIC). We also did not conduct backwards and forwards literature searching, which may have yielded additional studies to include. Additionally, in our screening we focused on studies that included at least one supervision group incorporating at least one common supervision element that we identified across the three major published guidelines for supervision in psychology, counseling, and social work. Future studies could also include studies that included any supervision group (regardless of content) and that may yield a larger sample of supervision without any common element for comparison. Finally, given the overarching goal of providing to supervision to improve client outcomes, we could have included studies with any therapist or client outcomes, not only studies containing both. While these studies were conducted in multiple settings (e.g., graduate training, a therapy training study, community mental health), with supervisors and therapists from different disciplines across several different countries, we purposefully cast a wide net in the hope that aggregating these findings helps to illustrate the challenges in measuring the impact of supervision across relevant therapist outcomes (which are arguably the most proximal outcomes for supervision to affect).

Conducting a systematic review and meta-analysis in this way also allowed an opportunity to evaluate supervision elements quantitatively and attempt to identify what, if any, elements of supervision could be considered an empirically supported

supervision practice. It also permitted a descriptive evaluation of the methods and the reporting practices often followed within supervision research (e.g., Watkins Jr. et al., 2021), highlighting the need for providing more description of what provided supervision entails. Despite the limited conclusions we are able to draw for including specific supervision common elements in routine supervision, we have welcomed the opportunity to begin to evaluate the empirical basis for these common supervision elements.

OVERALL DISCUSSION

Main Findings

Overall, we found the content of what makes good, or "best practice" supervision to be fairly common across mental health disciplines in psychotherapy. Across counseling, psychology and social work, we identified 17 common elements recommended to be incorporated into supervision in order to achieve optimal therapist outcomes. We also found several studies suggestive of positive relationships between common supervision elements and meaningful outcomes in a preliminary scoping review. However, when we conducted a systematic review and meta-analysis of the association between common supervision elements and therapist outcomes, we found limited empirical support for these common supervision elements derived from expert consensus and outlined in best practice supervision guidelines. Although, therapists showed significant pre-post improvements in at least competence and confidence, there was no indication that the inclusion of common supervision elements in particular led to greater improvements. Indeed, the more of these elements therapists received, the lower their improvement. Three of the common elements in particular yielded worse outcomes when they were included in supervision than when they were not (beneficence toward client; technology considerations, general feedback). These findings are surprising and raise questions for the field of therapy training as a whole.

Of course, these findings are based on limited research. Few studies have examined changes in therapist outcomes following supervision. In the meta-analysis, we identified 29 studies, with 46 supervision groups, that reported therapist outcomes. Even fewer have compared therapist outcomes from different types of supervision; indeed, we chose to examine pre-post effect sizes in our meta-analysis because so few comparative supervision studies have been conducted.

We undertook three studies of supervision elements and meaningful outcomes to better understand the existing evidence base for specific supervision practices. In Study 1, we were able to identify 17 common supervision elements recommended by two or more major disciplines (i.e., counseling, social work, and psychology), suggesting that there is consensus around what supervision best practices may entail for therapists/supervisees providing therapy services. Studies 2 and 3 provided opportunities to begin evaluating to what extent these common supervision elements are explicitly referenced in supervision literature, what kinds of important outcomes are being studied following the receipt of supervision, and whether any of these common supervision elements have a sufficient evidence base to support their use. We hoped that identifying empirical support for specific supervision practices in this way would have broad implications for the training of supervisors and for the delivery and receipt of supervision by therapists in training. However, the lack of definitive support for any common supervision element or proposed moderator beyond the receipt of supervision containing at least one common supervision elements limits the recommendations we are able to make from these findings for supervisors and trainees.

Limitations of the Literature

In conducting these studies, we came across several challenges in extracting the data we hoped to find. First, many supervision studies lacked basic demographic information for participants, making it hard to draw conclusions based on factors such as therapist or supervisor discipline, the setting where therapy was delivered, the type of counseling or intervention delivered, or the client population receiving treatment. Similarly, several studies lacked key information related to the number of supervisors providing supervision, the amount of time spent in supervision, or the modality of

supervision (i.e., group or individual supervision). Additionally, we saw in Studies 2 and 3 that supervision studies rarely include a control group or compare supervision groups in a way that outcomes can be clearly associated with a causal supervision element. With many of these studies, it would be very difficult to replicate the supervision provided based on the published supervision description alone. To that end, many of the elements that we were unable to identify in the 29 meta-analysis any studies seem to be implied in most supervisions, but these elements are rarely explicitly stated and therefore hard to measure. For example, the element focusing on maintaining appropriate relationships is generally expected across therapy service settings. However, the element of coordination among professionals may not be relevant for supervision teams that do not collaborate with professionals from other disciplines (e.g., primary care physicians) and may therefore not be recommended across treatment settings. As stated above, even when we were able to identify supervision elements, they were almost never studied in a way where that element could be causally linked with the measured outcomes. However, one positive finding across Studies 2 and 3 is that the number of supervision studies published in the last 15 years has both grown and also that these studies make more explicit reference to the common supervision elements, suggesting that the published recommendations from multiple disciplines are already incorporated into current supervision practices.

Future Directions

While we have worked from the published guidelines for supervision to operationalized proposed best practice supervision elements, we should mention that other research groups have also worked to identify supervision processes and content. For instance, the Supervision Process Outcomes Coding System (Dorsey et al., 2018) was developed to assess supervision practice elements (e.g., behavioral rehearsal) as

well as content related to the delivery of EBPs. Recently, Choy-Brown and colleagues (2022) developed a framework for understanding what elements of supervision may contribute to EBP implementation and their identified common elements somewhat map on to our identified taxonomy (e.g., agenda setting, eliciting feedback on supervision, fidelity assessment, review of actual practice). This approach highlights an exciting new direction for supervision research focused on supporting the delivery of evidence-based treatment practices using evidence-based supervision practices, as well as continuing to operationalize supervision elements in ways that can be replicated in future studies of supervision.

As mentioned earlier, an important future direction is also to evaluate the effect of supervision practices on client outcomes. A recent meta-analysis (Keum and Wang, 2020) highlighted important implications for client progress outcomes, but did not look at supervision process as granularly as we have in the current study to determine the potential links between supervision elements and these outcomes. To continue advancing our understanding of "evidence-based supervision," promoting studies of supervision supporting EBP implementation while also considering the impact of identified supervision elements on client outcomes will be very important to building this growing evidence base for supervision.

To address the limitations discussed here, ideal next steps would be conducting controlled trials comparing supervision approaches incorporating the elements we found to have the most evidence (i.e., beneficence for the client, technological considerations, and general feedback) similar to treatment dismantling studies. With this approach, it will be critical to clearly describe exactly how supervision was delivered and what, if any, common supervision elements were included. As the evidence base for supervision continues to grow, it may also be worthwhile to broadly

survey supervisors across disciplines on their use of the common elements that we have identified, their perceptions of these supervision common elements, and what supervision element they see as most useful in practice, including any elements that fall outside the recommended guidelines. Eventually, the hope would be these findings would be incorporated into supervisor training in the future to maximize the impact of supervision on therapist and client outcomes.

References

▵ Abbass, A. (2004). Small-group videotape training for psychotherapy skills development. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *28*(2), 151–155. https://doi.org/10.1176/appi.ap.28.2.151

Accurso, E. C., Taylor, R. M., & Garland, A. F. (2011). Evidence-based practices addressed in community-based children's mental health clinical supervision. *Training and Education in Professional Psychology*, *5*(2), 88. https://doi.org/10.1037/e607262010-001

Alfonsson, S., Parling, T., Spannargard, A., Andersson, G., Lundgren, T. (2018). The effects of clinical supervision on supervisees and patients in cognitive behavioral therapy: a systematic review. *Cognitive Behaviour Therapy*, *47*(3), 206–228. https://doi.org/10.1080/16506073.2017.1369559

▵ *Alfonsson, S., Lundgren, T., & Andersson, G. (2020). Clinical supervision in cognitive behavior therapy improves therapists' competence: a single-case experimental pilot study. *Cognitive Behaviour Therapy*, 1–14. https://doi.org/10.1080/16506073.2020.1737571

American Psychological Association (APA). (2015). *SoA: Standards of accreditation for health service psychology*. Retrieved October 23, 2019, from https://www.apa.org/ed/accreditation/about/policies/standards-of-accreditation.pdf

▵ Anderson, T., Crowley, M. E. J., Patterson, C. L., & Heckman, B. D. (2012). The

influence of supervision on manual adherence and therapeutic processes. Journal of Clinical Psychology, 68(9), 972–988. https://doi.org/10.1002/jclp.21879

Association for Counselor Education and Supervision Taskforce on Best Practices in Clinical Supervision. (2011, April). Best practices in clinical supervision. Retrieved May 9, 2020 from https://acesonline.net/wp-content/uploads/2011/10/ACES-Best-Practices-in-clinical-supervision-document-FINAL.pdf

Association for Counselor Education and Supervision. (1990). Standards for counseling supervisors. *Journal of Counseling & Development*, 69, 30–32. https://doi.org/10.1002/j.1556-6676.1990.tb01450.x

Association of Social Work Boards. (2011). Model Social Work Practice Act. Culpepper, VA: Author. Retrieved from www.aswb.org/pdfs/Model_law.pdf

*Bambling, M., King, R., Raue, P., Schweitzer, R., & Lambert, W. (2006). Clinical supervision: Its influence on client-rated working alliance and client symptom reduction in the brief treatment of major depression. *Psychotherapy Research, 16*(3), 317–331. https://doi.org/10.1080/10503300500268524

Barrett, J., Gonsalvez, C. J., & Shires, A. (2019). Evidence-based practice within supervision during psychology practitioner training: A systematic review. *Clinical Psychologist*. https://doi.org/10.1111/cp.12196

Bearman, S. K., Bailin, A., & Sale, R. (2019). Graduate school training in CBT supervision to develop knowledge and competencies. *The Clinical Supervisor*, 1-19. https://doi.org/10.1080/07325223.2019.1663459

*Bearman, S. K., Schneiderman, R. L., & Zoloth, E. (2017). Building an evidence

base for effective supervision practices: An analogue experiment of supervision to increase EBT fidelity. *Administration and Policy in Mental Health*, *44*(2), 293–307. https://doi.org/10.1007/s10488-016-0723-8

*Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., ... & Research Network on Youth Mental Health. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*(6), 518-529. https://doi.org/10.1007/s10488-013-0485-5

*Beckman, M., Forsberg, L., Lindqvist, H., Diez, M., Enö Persson, J., & Ghaderi, A. (2017). The dissemination of motivational interviewing in Swedish county councils: Results of a randomized controlled trial. *PLoS ONE*, *12*(7), 1–16. https://doi.org/10.1371/journal.pone.0181715

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology*, *41*(2), 257-278. https://doi.org/10.1111/j.2044-8317.1988.tb00901.x

Beinart, H. (2004). Models of supervision and the supervisory relationship and their evidence base. In *Supervision and Clinical Psychology* (pp. 51-65). Routledge.

Bernard, J. M. (1997). The discrimination model. In C. E. Watkins, Jr. (Ed.), *Handbook of Psychotherapy Supervision* (p. 310–327).

Bernard, J. M., & Goodyear, R. K. (1998). Fundamentals of clinical supervision. Allyn & Bacon.

Borders, L. D. (2014). Best practices in clinical supervision: Another step in delineating effective supervision practice. *American Journal of Psychotherapy*, *68*(2), 151162.

https://doi.org/10.1176/appi.psychotherapy.2014.68.2.151

▵ Boyes, K. M. (2009). Differential effects of matching theoretical orientation in supervisor/supervisee dyads [ProQuest Information & Learning]. *In Dissertation Abstracts International Section A: Humanities and Social Sciences* (Vol. 69, Issues 11-A). http://proxy.mul.missouri.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=psyh&AN=2009-99090-244&site=ehost-live&scope=site

▵ *Bradshaw, T., Butterworth, A., & Mairs, H. (2007). Does structured clinical supervision during psychosocial intervention education enhance outcome for mental health nurses and the service users they work with? *Journal of Psychiatric and Mental Health Nursing*, *14*(1), 4–12. https://doi.org/10.1111/j.1365-2850.2007.01021.x

▵ *Burlingame, G. M., Earnshaw, D., Ridge, N. W., Matsumo, J., Bulkley, C., Lee, J., & Hwang, A. D. (2007). Psycho-educational group treatment for the severely and persistently mentally ill: how much leader training is necessary? *International Journal of Group Psychotherapy*, *57*(2), 187–218. https://doi.org/10.1521/ijgp.2007.57.2.187

*Carmel, A., Villatte, J. L., Zachary Rosenthal, M., Chalker, S., & Comtois, K. A. (2014). Applying technological approaches to clinical supervision in dialectical behavior therapy: A randomized feasibility trial of the Bug-in-the-Eye (BITE) model. *Cognitive and Behavioral Practice*, *23*(2), 221–229. https://doi.org/10.1016/j.cbpra.2015.08.001

Carlson, L., Rapp, C. A., & Eichler, M. S. (2012). The experts rate: Supervisory

behaviors that impact the implementation of evidence-based practices. *Community Mental Health Journal*, *48*(2), 179-186. https://doi.org/10.1007/s10597-010-9367-4

*Cherniss, C., & Egnatios, E. (1978). Clinical supervision in community mental health. *Social Work*, *23*(3), 219–223. https://doi.org/10.1093/sw/23.3.219

Choy-Brown, M., Baslock, D., Cable, C., Marsalis, S., & Williams, N. J. (2022). In search of the common elements of clinical supervision: A systematic review. *Administration and Policy in Mental Health and Mental Health Services Research*, 1-21. https://doi.org/10.1007/s10488-022-01188-0

Choy-Brown, M., & Stanhope, V. (2018). The availability of supervision in routine mental health care. *Clinical Social Work Journal*, 1-10. https://doi.org/10.1007/s10615-018-0687-0

Cohen, J. (1988). The effect size index: d. *Statistical Power Analysis for the Behavioral Sciences*, *2*(1).

▵ Corbin, M. M. (2012). Personality characteristics effects on supervisory working alliance and counselor trainees' skill development [ProQuest Information & Learning]. In Dissertation Abstracts International: Section B: The Sciences and Engineering (Vol. 72, Issues 7-B). http://proxy.mul.missouri.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=psyh&AN=2012-99020-482&site=ehost-live&scope=site

Council for Accreditation of Counseling and Related Educational Programs (CACREP). (2013). *Section 6: Doctoral Standards.* Retrieved October 23, 2019, from

http://www.cacrep.org/section-6-doctoral-standards-counselor-education-and-supervision

Davidson, K. M., Rankin, M. L., Begley, A., Lloyd, S., Barry, S. J. E. E., McSkimming, P., Bell, L., Allan, C., Osborne, M., Ralston, G., Bienkowski, G., Mellor-Clark, J., & Walker, A. (2017). Assessing patient progress in psychological therapy through feedback in supervision: the MeMOS randomized controlled trial (*Measuring and Monitoring clinical Outcomes in Supervision: MeMOS). *Behavioural and Cognitive Psychotherapy*, *45*(3), 209–224. https://doi.org/10.1017/S1352465817000029

▵ Der Pan, P. J., Deng, L.-Y. F., & Tsai, S.-L. (2008). Evaluating the use of reflective counseling group supervision for military counselors in Taiwan. *Research on Social Work Practice*, 18(4), 346–355. https://doi.org/10.1177/1049731507313981

Dorsey, S., Kerns, S. E., Lucid, L., Pullmann, M. D., Harrison, J. P., Berliner, L., ... & Deblinger, E. (2018). Objective coding of content and techniques in workplace-based supervision of an EBT in public mental health. *Implementation Science*, *13*(1), 19. https://doi.org/10.1186/s13012-017-0708-3

Dorsey, S., Pullmann, M. D., Kerns, S. E., Jungbluth, N., Meza, R., Thompson, K., & Berliner, L. (2017). The juggling act of supervision in community mental health: Implications for supporting evidence-based treatment. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(6), 838-852. https://doi.org/10.1007/s10488-017-0796-z

Ellis, M. V., Ladany, N., Krengel, M., & Schult, D. (1996). Clinical supervision

research from 1981 to 1993: A methodological critique. *Journal of Counseling Psychology*, *43*(1), 35. https://doi.org/10.1037/0022-0167.43.1.35

▵ Eryılmaz, A., & Mutlu, T. (2017). Developing the four-stage supervision model for counselor trainees. *Kuram ve Uygulamada Eğitim Bilimleri/Educational Sciences: Theory & Practice*, 17(2), 597–629. https://doi.org/10.12738/estp.2017.2.2253

Falender, C. A. (2014). Supervision outcomes: Beginning the journey beyond the emperor's new clothes. *Training and Education in Professional Psychology*, *8*(3), 143. https://doi.org/10.1037/tep0000066

Falender, C. A., Burnes, T. R., & Ellis, M. V. (2013). Multicultural clinical supervision and benchmarks: Empirical support informing practice and supervisor training. *The Counseling Psychologist*, *41*(1), 8-27. https://doi.org/10.1177/0011000012438417

Falender, C. A., & Shafranske, E. P. (2004). Clinical supervision: A competency-based approach. Washington, DC: American Psychological Association. https://doi.org/10.1037/10806-000

Falender, C. A., & Shafranske, E. P. (2007). Competence in competency based supervision practice: Construct and application. *Professional Psychology: Research and Practice,* 38(3), 232-240. https://doi.org/10.1037/0735-7028.38.3.232

Falender, C. A., Shafranske, E. P., & Ofek, A. (2014). Competent clinical supervision: Emerging effective practices. *Counselling Psychology Quarterly*, *27*(4), 393-408. https://doi.org/10.1080/09515070.2014.934785

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). Robumeta: Robust variance meta-

regression, version 2.0. Retrieved from https://CRAN.R-project.org/package=robumeta

Fouad, N., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B, Collins, F.L., Jr., & Crossman, R. E. (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. https://doi.org/10.1037/a0015832

*Foxwell, A. A., Kennard, B. D., Rodgers, C., Wolfe, K. L., Cassedy, H. F., & Thomas, A. (2017). Developing a peer mentorship program to increase competence in clinical supervision in clinical psychology doctoral training programs. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry, 41*(6), 828–832. https://doi.org/10.1007/s40596-017-0714-4

▵ Frosch, C. A., Varwani, Z., Mitchell, J., Caraccioli, C., & Willoughby, M. (2018). Impact of reflective supervision on early childhood interventionists' perceptions of self-efficacy, job satisfaction, and job stress. *Infant Mental Health Journal, 39*(4), 385–395. https://doi.org/10.1002/imhj.21718

*Henggeler, S. W., Schoenwald, S. K., Liao, J. G., Letourneau, E. J., & Edwards, D. L. (2002). Transporting efficacious treatments to field settings: the link between supervisory practices and therapist fidelity in MST programs. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53, 31*(2), 155–167. https://doi.org/10.1207/S15374424JCCP3102_02

Hess, A. K. (1986). 4/Growth in supervision: Stages of supervisee and supervisor

development. *The Clinical Supervisor*, *4*(1-2), 51-68.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

Hill, C. E., & Knox, S. (2013). Training and supervision in psychotherapy. *Bergin and Garfield's handbook of psychotherapy and behavior change*, *6*, 775-812.

▵ Hilsenroth, M. J., Kivlighan, D. M. J., & Slavin-Mulford, J. (2015). Structured supervision of graduate clinicians in psychodynamic psychotherapy: alliance and technique. *Journal of Counseling Psychology*, *62*(2), 173–183. https://doi.org/10.1037/cou0000058

Holloway, E. L., & Neufeldt, S. A. (1995). Supervision: Its contributions to treatment efficacy. *Journal of Consulting and Clinical Psychology*, *63*(2), 207. https://doi.org/10.1037/0022-006x.63.2.207

▵ Jones, E., Manassis, K., Arnold, P., Ickowicz, A., Mendlowitz, S., Nowrouzi, B., Wilansky-Traynor, P., Bennett, K., & Schmidt, F. (2015). Translating cognitive behavioral therapy for anxious youth to rural-community settings via tele-psychiatry. *Community Mental Health Journal*, *51*(7), 852–856. https://doi.org/10.1007/s10597-015-9882-4

Kaslow, N. J., Borden, K. A., Collins Jr, 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*(7), 699-712.

Keum, B. T., & Wang, L. (2020). Supervision and psychotherapy process and outcome: A meta-analytic review. *Translational Issues in Psychological Science*. http://dx.doi.org/10.1037/tps0000272

Kühne, F., Maas, J., Wiesenthal, S., & Weck, F. (2019). Empirical research in clinical supervision: a systematic review and suggestions for future studies. *BMC Psychology*, *7*(1), 54. https://doi.org/10.1186/s40359-019-0327-7

▵ Lau, M. A., Dubord, G. M., & Parikh, S. V. (2004). Design and feasibility of a new cognitive-behavioural therapy course using a longitudinal interactive format. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, *49*(10), 696–700. https://doi.org/10.1177/070674370404901008

Loganbill, C., Hardy, E., & Delworth, U. (1982). Supervision: A conceptual model. *The Counseling Psychologist*, *10*(1), 3-42.

▵ Lu, W., Yanos, P. T., Gottlieb, J. D., Duva, S. M., Silverstein, S. M., Xie, H., Rosenberg, S. D., & Mueser, K. T. (2012). Use of fidelity assessments to train clinicians in the CBT for PTSD program for clients with serious mental illness. *Psychiatric Services (Washington, D.C.)*, *63*(8), 785–792. https://doi.org/10.1176/appi.ps.201000458

▵ Manassis, K., Ickowicz, A., Picard, E., Antle, B., McNeill, T., Chahauver, A., Mendlowitz, S., Monga, S., & Adler-Nevo, G. (2009). An innovative child CBT training model for community mental health practitioners in Ontario. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *33*(5), 394–399. https://doi.org/10.1176/appi.ap.33.5.394

▵ *Martino, S., Paris, M. J., K. M., Anez, L., Nich, C., Canning-Ball, M., Hunkele, K., Olmstead, T. A., & Carroll, K. M. (2016). The effectiveness and cost of clinical supervision for motivational interviewing: A randomized controlled trial. *Journal of Substance Abuse Treatment*, *68*, 11–23. https://doi.org/10.1016/j.jsat.2016.04.005

*Meadowcroft, P., Townsend, M. Z., & Maxwell, A. (2018). A sustainable alternative to the gold standard EBP: Validating existing programs. *The Journal of Behavioral Health Services & Research*, *45*(3), 421–439. https://doi.org/10.1007/s11414-018-9599-6

Milne, D. (2007). An empirical definition of clinical supervision. *British Journal of Clinical Psychology*, *46*(4), 437-447. https://doi.org/10.1348/014466507x197415

Milne, D. (2014). Beyond the "acid test": A conceptual review and reformulation of outcome evaluation in clinical supervision. *American Journal of Psychotherapy*, *68*(2), 213-230. https://doi.org/10.1176/appi.psychotherapy.2014.68.2.213

Milne, D., Aylott, H., Fitzpatrick, H., & Ellis, M. V. (2008). How does clinical supervision work? Using a "best evidence synthesis" approach to construct a basic model of supervision. *The Clinical Supervisor*, *27*(2), 170-190. https://doi.org/10.1080/07325220802487915

Milne, D., & Dunkerley, C. (2010). Towards evidence-based clinical supervision: The development and evaluation of four CBT guidelines. *The Cognitive Behaviour Therapist*, *3*(2), 43-57. https://doi.org/10.1017/S1754470X10000048

Milne, D., & James, I. (2000). A systematic review of effective cognitive-behavioural supervision. *British Journal of Clinical Psychology*, *39*(2), 111-127. https://doi.org/10.1348/014466500163149

▵ Milne, D. L., & James, I. A. (2002). The observed impact of training on competence in clinical supervision. *The British Journal of Clinical Psychology*, *41*(Pt 1), 55–72. https://doi.org/10.1348/014466502163796

Milne, D., & Reiser, R. P. (2012). A rationale for evidence-based clinical supervision.

Journal of Contemporary Psychotherapy, 42(3), 139-149. https://doi.org/10.1007/s10879-011-9199-8

*Milne, D. L., Reiser, R. P., & Cliffe, T. (2013). An N = 1 evaluation of enhanced CBT supervision. *Behavioural and Cognitive Psychotherapy*, *41*(2), 210–220. https://doi.org/10.1017/S1352465812000434

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Milne, D., & Westerman, C. (2001). Evidence-based clinical supervision: Rationale and illustration. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *8*(6), 444-457. https://doi.org/10.1002/cpp.297

▵ Nakamura, B. J., Selbo-Bruns, A., Okamura, K., Chang, J., Slavin, L., & Shimabukuro, S. (2014). Developing a systematic evaluation approach for training programs within a train-the-trainer model for youth cognitive behavior therapy. *Behaviour Research and Therapy*, *53*, 10–19. https://doi.org/10.1016/j.brat.2013.12.001

National Association of Social Workers. (2008). Code of ethics of the National Association of Social Workers. Retrieved October 23, 2019, from https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English

National Association of Social Workers (NASW). (2013). *Best practice standards in social work supervision.* Retrieved October 23, 2019, from http://www.socialworkers.org/Practice/Practice-Standards-Guidelines

Pretorius, W. M. (2006). Cognitive behavioural therapy supervision: recommended practice. *Behavioural and Cognitive Psychotherapy*, *34*(4), 413-420. https://doi.org/10.1017/S1352465806002876

* ▵ Rakovshik, S. G., McManus, F., Vazquez-Montes, M., Muse, K., & Ougrin, D. (2016). "Is supervision necessary? Examining the effects of Internet-based CBT training with and without supervision": Correction to Rakovshik et al. (2016). *Journal of Consulting and Clinical Psychology*, *84*(12), 1093. https://doi.org/10.1037/ccp0000169

* ▵ Reese, R. J., Aldarondo, F., Anderson, C. R., Lee, S.-J., Miller, T. W., & Burton, D. (2009). Telehealth in clinical supervision: a comparison of supervision formats. *Journal of Telemedicine and Telecare*, *15*(7), 356–361. https://doi.org/10.1258/jtt.2009.090401

▵ Reese, R. J., Usher, E. L., Bowman, D. C., Norsworthy, L. A., Halstead, J. L., Rowlands, S. R., & Chisholm, R. R. (2009). Using client feedback in psychotherapy training: An analysis of its influence on supervision and counselor self-efficacy. *Training and Education in Professional Psychology*, *3*(3), 157–168. https://doi.org/10.1037/a0015673

Reiser, R. P., & Milne, D. (2012). Supervising cognitive-behavioral psychotherapy: Pressing needs, impressing possibilities. *Journal of Contemporary Psychotherapy*, 42, 161-171. https://doi.org/10.1007/s10879-011- 9200-6

Reiser, R. P., & Milne, D. L. (2014). A systematic review and reformulation of outcome evaluation in clinical supervision: Applying the fidelity framework. *Training and Education in Professional Psychology*, *8*(3), 149. https://doi.org/10.1037/tep0000031

*Resko, S. M., Walton, M. A., Chermack, S. T., Blow, F. C., & Cunningham, R. M.

(2012). Therapist competence and treatment adherence for a brief intervention addressing alcohol and violence among adolescents. *Journal of Substance Abuse Treatment*, *42*(4), 429–437. https://doi.org/10.1016/j.jsat.2011.09.006

*Rothrauff-Laschober, T. C., Eby, L. T. de T., & Sauer, J. B. (2013). Effective clinical supervision in substance disorder treatment programs and counselor job performance. *Journal of Mental Health Counseling*, *35*(1), 76–94. https://doi.org/10.17744/mehc.35.1.50n6w37328qp8611

▵ Ruzek, J. I., Rosen, R. C., Marceau, L., Larson, M. J., Garvert, D. W., Smith, L., & Stoddard, A. (2012). Online self-administered training for post-traumatic stress disorder treatment providers: design and methods for a randomized, prospective intervention study. *Implementation Science: IS*, *7*, 43. https://doi.org/10.1186/1748-5908-7-43

▵ Safran, J., Muran, J. C., Demaria, A., Boutwell, C., Eubanks-Carter, C., & Winston, A. (2014). Investigating the impact of alliance-focused training on interpersonal process and therapists' capacity for experiential reflection. Psychotherapy Research : Journal of the Society for Psychotherapy Research, 24(3), 269–285. https://doi.org/10.1080/10503307.2013.874054

*Schmidt, G. L., & Bonjean, M. J. (1995). Family therapy education for psychiatry residents: A pilot study of efficacy. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, *19*(2), 74–80. https://doi.org/10.1007/BF03341534

*Schoenwald, S. K., Sheidow, A. J., & Chapman, J. E. (2009). Clinical supervision in treatment transport: Effects on adherence and outcomes. *Journal of Consulting and Clinical Psychology*, *77*(3), 410. https://doi.org/10.1037/a0013788

Schoenwald, S. K., Sheidow, A. J., & Letourneau, E. J. (2004). Toward effective quality assurance in evidence-based practice: Links between expert consultation, therapist fidelity, and child outcomes. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 94-104. https://doi.org/10.1207/S15374424JCCP3301_10

▵ Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: a dissemination trial of three strategies of training clinicians in cognitive-behavioral therapy. *Journal of Consulting and Clinical Psychology, 73*(1), 106–115. https://doi.org/10.1037/0022-006X.73.1.106

Simpson-Southward, C., Waller, G., & Hardy, G. E. (2017). How do we know what makes for "best practice" in clinical supervision for psychological therapists? A content analysis of supervisory models and approaches. *Clinical Psychology & Psychotherapy*, *24*(6), 1228-1245. https://doi.org/10.1002/cpp.2084

▵ Smith, J. L., Amrhein, P. C., Brooks, A. C., Carpenter, K. M., Levin, D., Schreiber, E. A., Travaglini, L. A., & Nunes, E. V. (2007). Providing live supervision via teleconferencing improves acquisition of motivational interviewing skills after workshop attendance. *The American Journal of Drug and Alcohol Abuse, 33*(1), 163–168. https://doi.org/10.1080/00952990601091150

*Son, E., & Ellis, M. V. (2013). A cross-cultural comparison of clinical supervision in South Korea and the United States. *Psychotherapy (Chicago, Ill.)*, *50*(2), 189–205. https://doi.org/10.1037/a0033115

Stoltenberg, C. D., & Delworth, U. (1987). *Supervising counselors and therapists: A*

developmental approach. Jossey-Bass.

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. Research Synthesis Methods, 5(1), 13–30. https://doi.org/10.1002/jrsm.1091

*Thew, G. R., Powell, C. L., Kwok, A. P., Lissillour Chan, M. H., Wild, J., Warnock-Parkes, E., Leung, P. W., & Clark, D. M. (2019). Internet-based cognitive therapy for social anxiety disorder in Hong Kong: Therapist training and dissemination case series. *JMIR Formative Research*, *3*(2), e13446. https://doi.org/10.2196/13446

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. Psychological Methods, 20(3), 375–393. Available from: https://doi.org/10.1037/met0000011

Tugendrajch, S. K., Sheerin, K. M., Andrews, J. H., Reimers, R., Marriott, B. R., Cho, E., & Hawley, K. M. (2021). What is the evidence for supervision best practices?. *The Clinical Supervisor*, *40*(1), 68-87. https://doi.org/10.1080/07325223.2021.1887785

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*(2), 576-594. https://doi.org/10.3758/s13428-012-0261-6

Viechtbauer W (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, **36**(3), 1–48. https://doi.org/10.18637/jss.v036.i03.

Watkins Jr, C. E. (2011). Does psychotherapy supervision contribute to patient

outcomes? Considering thirty years of research. *The Clinical Supervisor*, 30(2), 235-256. https://doi.org/10.1080/07325223.2011.619417

Watkins Jr, C. E. (2014). The supervisory alliance: A half century of theory, practice, and research in critical perspective. *American Journal of Psychotherapy*, *68*(1), 19-55. https://doi.org/10.1176/appi.psychotherapy.2014.68.1.19

Watkins Jr, C. E. (2020). What do clinical supervision research reviews tell us? Surveying the last 25 years. *Counselling and Psychotherapy Research*, *20*(2), 190-208. https://doi.org/10.1002/capr.12287

Weck, F., Kaufmann, Y. M., & Witthöft, M. (2017). Topics and techniques in clinical supervision in psychotherapy training. *The Cognitive Behaviour Therapist*, *10*. https://doi.org/10.1017/S1754470X17000046

▵ *Weck, F., Jakob, M., Neng, J. M. B., Höfling, V., Grikscheit, F., & Bohus, M. (2016). The effects of bug-in-the-eye supervision on therapeutic alliance and therapist competence in cognitive-behavioural therapy: A randomized controlled trial. *Clinical Psychology and Psychotherapy*, *23*(5), 386–396. https://doi.org/10.1002/cpp.1968

Watkins Jr, C. E., Vîşcu, L. I., & Cadariu, I. E. (2021). Psychotherapy supervision research: On roadblocks, remedies, and recommendations. *European Journal of Psychotherapy & Counselling*, *23*(1), 8-25. https://doi.org/10.1080/13642537.2021.1881139

▵ Weingardt, K. R., Cucciare, M. A., Bellotti, C., & Lai, W. P. (2009). A randomized trial comparing two models of web-based training in cognitive-behavioral therapy for substance abuse counselors. Journal of Substance Abuse Treatment, 37(3), 219–227. https://doi.org/10.1016/j.jsat.2009.01.002

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 and Psychotherapy Research*, *7*(1), 54-65. https://doi.org/10.1080/14733140601185274

*Woo, S. M., Hepner, K. A., Gilbert, E. A., Osilla, K. C., Hunter, S. B., Munoz, R. F., & Watkins, K. E. (2013). Training addiction counselors to implement an evidence-based intervention: Strategies for increasing organizational and provider acceptance. *Cognitive and Behavioral Practice*, *20*(2), 232–244. https://doi.org/10.1016/j.cbpra.2012.03.004

Worthen, V. E., & Lambert, M. J. (2007). Outcome oriented supervision: Advantages of adding systematic client tracking to supportive consultations. *Counselling and Psychotherapy Research*, *7*(1), 48-53. https://doi.org/10.1080/14733140601140873

▵ *Xavier, K., Shepherd, L., & Goldstein, D. (2007). Clinical supervision and education via videoconference: A feasibility project. *Journal of Telemedicine and Telecare*, *13*(4), 206–209. https://doi.org/10.1258/135763307780907996

*denotes article inclusion in Study 2 review

▵ denotes article inclusion in Study 3 meta-analysis

Table 1.
Study 1: Common Elements across Supervision Guidelines

Guideline elements	Examples and operationalization	APA[a] (2014)	ACES[b] (2011)	NASW[c] (2013)
Beneficence/supervisor responsibility for the client	• Supervisor assigns supervisee developmentally appropriate cases • Supervisor increases supervision time for challenging cases • Supervisor sees supervisee's clients as their responsibility	X	X	X
Coordination among relevant professionals	• If appropriate, supervision group includes other relevant professionals (e.g., primary care providers, psychiatrists)	X	X	X
Modeling professionalism	• Stated adherence to licensure or accreditation standards (e.g., referencing APA or NASW standards for licensure or accreditation in description of supervision format or processes)	X	X	X
Supervisor assesses supervision/supervisee (formal/ ongoing)	• Formal evaluation of supervisee performance (separate from feedback provided in supervision meetings) • Completion of quantitative measures (e.g., competence, adherence, etc.) • Use of live observation and video/audio review techniques in formal observation	X	X	X
Supervisors set expectations/contract	• Supervisors identify expected program competencies and performance standards, and assist the supervisee to formulate individual learning goals • Supervisor and supervisee make a written contract (e.g., in the case of social work, this would be a legal contract) • Supervisor engages in informed consent with supervisee at onset of supervision	X	X	X

Technology considerations	• Supervisor uses technology to enhance the supervision process (e.g., watching videos) • Supervisors engage in learning (e.g., reading guidelines) to learn about the use of technology for supervision • Supervisors discuss the use of technology with clients with their supervisees (e.g., use of internet searches, social media)	X	X	X
Self-awareness of supervisor competence	• Supervisor pursues continuing education • Supervisor solicits feedback from supervisee • Supervisors only provide supervision within the limits of their own competence • Supervisor self-evaluation of competence • Emphasis on supervisor continuing education	X	X	X
Adherence to established guidelines	• Reference to existing guidelines in description of supervision format (e.g., APA, NASW, ACES guidelines) • Does NOT include reference to psychotherapy models or theoretical orientations	O	X	X
Maintaining appropriate relationships	• Supervisor does not engage in inappropriate relationships with supervisee (e.g., a romantic relationship)	O	X	X
Modeling ethical practices	• Supervision format references managing conflicts with professionalism	X	O	X
Providing multicultural supervision	• Supervisor infuses multicultural considerations regarding the supervisee into supervision • Supervisor builds their own multicultural awareness through various means (e.g., trainings)	X	O	X
Documentation of supervision	• Supervisory progress notes	X	O	X
Collaborative supervisory relationship	• Supervisor and supervisee creating contract together	X	X	O

	• Supervisee provides feedback regarding supervisor or supervision format			
Supporting client advocacy	• Supervisor models advocacy for human rights and intervention with institutions and systems	X	X	O
Goal setting	• Creating a supervision contract with identified, developmentally appropriate goals • Referencing identified goals during supervision meetings	X	X	O
Supervisor encourages supervisee self-evaluation	• Supervisor solicits ongoing feedback from supervisee on their own competence and incorporates that feedback into supervisee evaluation	X	X	O
Supervisor provides feedback to supervisee (general)	• Any review of taped/audio-recorded sessions • Any discussion of supervisee performance	X	X	O

Note. [a]American Psychological Association, [b]Association for Counselor Education and Supervision, [c]National Association of Social Workers

Table 2.
Study 2: Outcomes Associated with Supervision Elements from Preliminary Review

Supervision guideline common elements	Number of Studies	Therapist fidelity	Therapist self-efficacy	Therapist competence	Supervisory alliance	Therapeutic alliance	Client perceptions of treatment	Client progress
17. Supervisor provides feedback to supervisee (general)	20	5 (1), 2 (0), 1(1/0)	2 (1)	12 (1), 1 (0)	6 (1), 1 (0)	1 (1), 1 (0)	1(0)	4 (0), 1 (1/0) 1 (1/-1)
4. Supervisor assesses supervision/supervisee (formal/ongoing)	16	5 (1), 1 (0), 1 (1/0)	1 (1)	10 (1)	5 (1), 1 (0)	3 (1)	2(1)	1 (1), 3 (0), 2 (1/-1)
6. Supervisor uses technology to enhance supervision	13	2 (1)	2 (1)	9 (1)	3 (1) 1 (0)	1 (1)	n/a	2 (1), 2 (0)
5. Supervisors set expectations/contract	9	1 (1), 1 (0),	1 (1)	6 (1), 1 (0)	3 (1)	n/a	n/a	1 (1/-1)
13. Collaborative supervisory relationship	9	1 (1)	2 (1)	6 (1)	4 (1)	1 (1)	n/a	1(0)
7. Self-awareness of supervisor competence	8	3 (1), 1 (0),	1 (1)	5 (1)	1 (1)	n/a	n/a	2 (1), 1 (0), 1 (1/-1)
16. Supervisor encourages supervisee self-evaluation	5	n/a	1 (1)	4 (1)	3 (1)	1 (1)	n/a	1 (1) 1 (0)
15. Goal setting	5	1 (1), 1 (1/0)	1 (1)	3 (1)	2 (1)	n/a	n/a	1 (1/-1)
12. Documentation of supervision	2	1 (1)	n/a	1 (1)	n/a	n/a	n/a	n/a
1. Beneficence/supervisor responsibility for the client	1	1 (1/0)	n/a	n/a	n/a	n/a	n/a	1 (1/-1)
2. Coordination among relevant professionals	1	n/a	1 (1)	1 (1)	1 (1)	n/a	n/a	n/a
8. Adherence to established guidelines	1	n/a	1 (1)	1 (1)	1 (1)	n/a	n/a	n/a
10. Modeling ethical practices	1	n/a	1 (1)	1 (1)	1 (1)	n/a	n/a	n/a
11. Providing multicultural supervision	1	n/a	1 (1)	1 (1)	1 (1)	n/a	n/a	n/a

14. Supporting client advocacy	1	1 (1/0)	n/a	n/a	n/a	n/a	n/a	1 (1/-1)
3. Modeling professionalism	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a
9. Maintaining appropriate relationships	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Note. 1 = positive association, 0 = null association, 1/0 = positive and null associations, 1/-1 = mixed or conflicting associations; numbers before parentheses are # of studies with this finding.

Table 3.
Study 3: Meta-Analysis Characteristics of Studies and Supervision Groups

									Supervision				
Author Year	Supervision Group Type(s)	Group #/ Abbreviations	Country	Supervisee N	Supervisee Discipline(s)	Supervisor Discipline(s)	Client Population	Treatment Provided	Setting	Modality	Frequency	Length/ Duration	Approximate Dose (Minutes)
Abbass 2004	S-CE	1	CAN	26	PSYR	NR	NR	STDP	GRT	IS	Weekly	90-180 min/ 24 weeks	3240
Alfonsson et al 2020	S-CE	1	SWE	6	PSYC, COUN	NR	Adults	CBT	PC, PSYR-C	GS	Weekly	50-60 min/ 5-8 weeks	357.5
Anderson et al 2012	S-CE	1	US	16	PSYC, PSYR	NR	Adults	TLDP	PP	IS	Monthly-Weekly	54 min/ 60-72 weeks	2754
Boyes 2009	S-CE, S-CE	2 (MTO, UTO)	US	27	COUN	COUN	Adults	NS	GRT	IS	Weekly	60 min/ 40 weeks	2400 per group
Bradshaw et al 2007	S-CE, S-CE	2 (CG, EG)	UK	23	NURS	NURS	NR	FIPIPS	GRT	GS	Biweekly	60-90 min/ 36 weeks	1350 per group
Burlingame et al 2007	S-CE, NS	2 (CS, NS)	US	11	NURS	NURS	Adult	PEG	PSYR-C	GS	Weekly	Unclear, est. 60 min/ 12 weeks	720 for S-CE, n/a for NS
Corbin 2012	S-CE, S-CE	2 (EM, CNM)	US	15	COUN	COUN	NR	NS	GRT	IS	Weekly	60 min/ 12 weeks	720 per group
Der Pan et al 2018	S-CE	1	TW	21	COUN	NR	Adults	NS	MC	GS	Monthly	120 min/ 104 weeks	3120
Eryılmaz et al 2017	S-CE	1	TUR	17	COUN	NR	Adults	NS	GRT	IS, GS	Weekly	60 min/ 10 weeks	600
Frosch et al 2018	S-CE	1	US	33	NR	NR	Children	NS	EIP	GS	2-3 weeks	40 min/ 36 weeks	1440
Hilsenroth et al 2015	S-CE	1	US	24	PSYC	PSYC	Adults	PDT, CBT	GRT	IS, GS	At least weekly	90 min IS, 120 min GS/	1260

												6 weeks	
Jones et al 2015	S-CE	1	CAN	53	COUN, SW	NR	Children	CBT	MS	GS	Weekly	Est. 60 min/ 20 weeks	1200
Lau et al 2004	S-CE	1	CAN	14	PSYC, PSYR, SW, PARA	NR	Adults	CBT	MS	GS	Weekly	120 min/ 10 weeks	1200
Lu et al 2012	S-CE	1	US	25	PSYC, COUN, SW	PSYC	Adults	CBT	MS	GS	Weekly	Est. 60 min/ 24 weeks	1440
Manassis et al 2009	S-CE	1	CAN	22	COUN, SW	NR	Children	CBT	MS	GS	Weekly	90 min/ 20 weeks	1800
Martino et al 2016	S-CE, S-NCE	2 (MIASTEP, SAU)	US	66	NR	NR	Adults	MI	MS	IS	Variable	35 min/ 247 sessions across ~48 weeks	8645 for S-CE, 100 for S-NCE
Milne James 2002	S-CE	1	UK	6	PSYR, NURS, CERT	PSYC	Unclear	NS	MS	IS	Unclear	Unclear/ 4 distinct phases	Cannot be estimated
Nakamura et al 2014	S-CE	1	US	4	COUN, SW	PSYC	Children	CBT	PMH	IS	Weekly	60 min/ 9 weeks	540
Weck et al 2016	S-CE, S-CE	2	GER	23	PSYC	PSYC	Adults	CBT	GRT	IS	Weekly	50-60 min (BITE) 50 min (DVB)/ 26 weeks	1495 for BITE, 1300 for DVB
Rakovshik et al 2016	S-CE, NS, NS	3 (IBT-S, IBT-CW, DT)	RUS, UKR	61	PSYC, PSYR	NR	Adults	CBT	MS	IS	Weekly	30 min/ 13 weeks	390 for IBT-S, 0 for IBT-NS
Reese et al 2009A	S-CE	1	US	9	COUN	PSYC	Adults, Children	NS	GRT	GS	Weekly	150 min/ 12 weeks	1800
Reese et al 2009 B	S-CE, S-CE	2 (CF, NF)	US	28	PSYC, COUN	PSYC, COUN	Adults, Children	NS	GRT	IS, GS	Weekly for IS, GS	Est. 90 min/ 32 weeks	2880 for both groups
Ruzek et al 2012	S-CE, NS	2 (WTS, WT)	US	168	PSYC, PSYR, SW	PSYC	Adults	CBT	MS	GS	Weekly	45-60 min/ 6 weeks	315 for WTS, 0 for WT

Safran et al 2014	S-CE, S-CE, S-CE	AFT16, AFT8, AFTall	US	22	PSYC	PSYC	Adults	CT	GRT	GS	Weekly	90 min/ 14 weeks, 22 weeks, or 8-16 weeks	1260 for AFT16, 1980 for AFT8, 1080 for AFTall
Sholomskas et al 2005	S-CE, NS, NS	3 (MWTS, MW, MO)	US	78	NR	NR	NR	CBT	MS	IS	Variable	Up to 3 60 min meetings/ 12 weeks	180 for S-CE, 0 for both NS
Smith et al 2007	S-CE		US	12	PSYC, COUN	PSYC	Adults	MI	MS	IS	Weekly	90 min/ 5 weeks	450
Smith et al 2012	S-CE, S-CE, NS	3 (Tape, TCS, WS)	US	97	COUN	COUND	Adults	MI	MS	IS	Weekly	60 min/ 5 weeks	300 for both S-CE, 0 for NS
Weingardt et al 2009	S-CE, S-CE	2 (HF, L)	US	147	COUN	NR	Adults	CBT	MS	GS	Weekly	60 mins/ 4 weeks	240 for both
Xavier et al 2007	S-CE		AUSNZ	18	PSYC, SW	NR	Adults	NS	MS	IS, GS	Monthly	120 mins/ 4 modules + up to 4 30 min IS calls	600

Note. Supervision Group Type: S-CE = supervision with common element included, S-NCE = supervision without common element, NS = no supervision,
Country: CAN = Canada, SWE = Sweden, US = United States, AUSNZ = Australia/New Zealand, UK = United Kingdom, TW = Taiwan, TUR = Turkey, GER = Germany, RUS = Russia, UKR = Ukraine
Discipline: PSYR = Psychiatry, PSYC = psychology, COUN = counseling, NURS = nursing, CERT = certificate, OCC = occupational therapy, MC = military counseling, EIP = early intervention/prevention, SW = social work, PARA = paraprofessionals,
Therapy: STDP = short-term dynamic psychotherapy, CBT = cognitive behavioral therapy, TLDP = time-limited dynamic psychotherapy, FIPIPS = family intervention and psychological intervention with psychotic symptoms, PEG = psycho-educational group, PDT-psychodynamic therapy, NTP = no therapy provided, MI = motivational interviewing, CT = cognitive therapy
Setting: GRT = graduate or resident training, PC = primary care, PSYR-C psychiatric care, PP = private practice, MS = multiple settings,
Modality: GS = group supervision, IS = individual supervision
General: NS = not specified, NR = not reported

Table 4.
Study 3: Meta-Analysis Common Supervision Elements and Outcomes

Author Year	Supervision Group Type(s)	Group Description (s)	Treatment Training Provided	Supervision Common Element(S)	Outcome (s) Measured & Number of Outcomes	Measure Name(s)	Pre-Post Effect Size(s) (Standard Deviations)
Abbass 2004	S-CE		Unclear	5, 6, 13, 16, 17	TC-1	Knowledge of Short-Term Dynamic Psychotherapy	2.16 (0.13)
Alfonsson et al 2020	S-CE		Yes-CBT	6,17	TC-1	Cognitive Therapy Scale—Revised	0.20 (0.17)
Anderson et al 2012	S-CE		Yes-TLDP	6, 17	TC-2	Vanderbilt Therapeutic Strategies Scale (Interviewing Style & Specific Strategies Subscales)	0.30 (0.07); 0.49 (0.07)
Boyes 2009	S-CE, S-CE	MTO	No	6, 8, 17 for both S-CEs	TC-2 (1 per group)	Basic Skills Observation	0.66 (0.10)
	S-CE, S-CE	UTO	No	6, 8, 17 for both S-CEs	TC-2 (1 per group)	Basic Skills Observation	0.43 (0.07)
Bradshaw et al 2007	S-CE, S-CE	EG	Yes-FIPIPS	4, 5, 6, 7, 17 for EG	TC-10 (5 per group)	Multiple-choice questions 1-5 knowledge about case management, knowledge about psychological interventions for psychotic symptoms, knowledge about family intervention, attitudes about schizophrenia and schizophrenia family work, general knowledge about schizophrenia and PSI	1.00 (0.14); 1.73 (0.23); 0.24 (0.09); 0.28 (0.09); 0.69 (0.11)
	S-CE, S-CE	CG	Yes-FIPIPS	4, 6, 17 for CG	TC-10 (5 per group)	Multiple-choice questions 1-5 knowledge about case management, knowledge about psychological interventions for psychotic symptoms, knowledge about family intervention, attitudes about schizophrenia and schizophrenia family work, general knowledge about schizophrenia and PSI	1.15 (0.14); 0.55 (0.10); 0.29 (0.09); 0.18 (0.08); 0.20 (0.09)

Burlingame et al 2007	S-CE	CS	Yes-PEG	17	TC-4 (2 per group)	Psycho–Educational Group Questionnaire for Nurses (PEGQ–N), Symptom Management Questionnaire for Nurses; SMQ–N	2.51 (0.83); 1.26 (0.36)
	NS	NS	Yes-PEG	n/a	TC-4 (2 per group)	Psycho–Educational Group Questionnaire for Nurses (PEGQ–N), Symptom Management Questionnaire for Nurses; SMQ–N	0.25 (0.21); 0.69 (0.25)
Corbin 2012	S-CE	EM	No	6, 8, 17 for both S-CEs	TC-2 (1 per group)	Basic Skills Observation	1.49 (0.26)
	S-CE	CNM	No	6, 8, 17 for both S-CEs	TC-2 (1 per group)	Basic Skills Observation	3.18 (0.86)
Der Pan et al 2018	S-CE		No	5, 10, 16, 17	TC-7	The Counseling Competence Inventory (CCI) – 7 subscales Professional Behavior subscale, Process skill-depth interaction, Process skill-message interaction, Process skill-basic communication, Conceptualization Skill, Personalization Scale, Personal Traits	0.74 (0.06); 1.40 (0.09); 0.95 (0.07); 0.96 (0.07); 1.52 (0.10); 0.77 (0.06); 0.96 (0.07)
Eryılmaz et al 2017	S-CE		No	1, 4, 6, 16, 17	TC-1	Counselor competencies evaluation form - competency level	2.82 (0.29)
Frosch et al 2018	S-CE		No	5, 17	TSE-1	Reflective Super- vision Self-Efficacy Scales for Supervisees	1.01 (0.05)
Hilsenroth et al 2015	S-CE		Yes-PDT	6, 17	TA-2	Comparative Psychotherapy Process Scale—Cognitive-Behavioral, Comparative Psychotherapy Process Scale psychodynamic-interpersonal,	-0.18 (0.04); 0.28 (0.04)
Jones et al 2015	S-CE		Yes-CBT	6, 11	TC-1	The Child CBT Multiple Choice test (CQ)	0.52 (0.02)
Lau et al 2004	S-CE		Yes-CBT	6, 16, 17	TC-1	Cognitive Therapy Scale	0.65 (0.09)

Lu et al 2012	S-CE		Yes-CBT	1, 4, 6, 7, 17	TA-14	CBT for PTSD Fidelity Scale-14 subscales, agenda setting, overview of program, pacing and efficient use of time, manual adherence, crisis plan, breathing retraining, use of educational materials, informative psychoeducation, homework review, assign homework, teaching effectiveness, interpersonal effectiveness, reduction of client distress, overall session quality	0.44 (0.04); -0.18 (0.04); 0.14 (0.04); -0.17 (0.04); 0.72 (0.05); 0.00 (0.04); -0.02 (0.04); 0.01 (0.04); 0.11 (0.04); 0.16 (0.04); -0.17 (0.04); 0.11 (0.04); -0.31 (0.04); 0.00 (0.04)
Manassis et al 2009	S-CE		Yes-CBT	1, 6	TC-1	CBT knowledge multiple choice test	0.90 (0.06)
Martino et al 2016	S-CE	MIASTEP	Yes-MI	4, 6, 7 for MIASTEP	TA-6 (3 per group) TC-4 (2 per group)	Independent Tape Rater Scale: Fundamental MI Adherence, Advanced MI Adherence, MI-Inconsistent Adherence, Fundamental MI Competence, Advanced MI Competence	-0.14 (0.04); 0.24 (0.04); -0.56

							(0.05); 0.30 (0.04); 0.66 (0.05)
	S-NCE	SAU	Yes-MI	n/a for SAU	TA-6 (3 per group) TC-4 (2 per group)	Independent Tape Rater Scale: Fundamental MI Adherence, Advanced MI Adherence, MI-Inconsistent Adherence, Fundamental MI Competence, Advanced MI Competence	-0.27 (0.04); 0.13 (0.04); 0.20 (0.04); -0.11 (0.04); -0.21 (0.04)
Milne James 2002	S-CE		No	6, 7, 15, 17	TC-5	Process Evaluation of Teaching and Supervision (PETS)- 5 subscales: reflection, conceptualization, planning, experimenting, experiencing	-1.19 (0.47); 1.00 (0.25); -0.66 (0.20); 5.06 (2.30); 0.22 (0.17)
Nakamura et al 2014	S-CE		Yes-CBT	4, 7	TC-4	Practice Element Train the Trainer - Therapist Version of the Therapy Process Observational Coding System for Child Psychotherapy (PETT-S TPOCS) | Problem-Solving (Observed), Problem Solving (Self Report), Exposure (Observed) Exposure (Self-Report)	10.36 (13.67); 1.15 (0.41); 3.58 (1.85); 0.52 (0.28)
Weck et al 2016	S-CE	BITE	Yes-CBT	4, 6, 17 for BITE	TC-2 (1 per group)	Cognitive Therapy Scale	-0.38 (0.10)

	S-CE	DVB	Yes-CBT	6, 17 for DVB	TC-2 (1 per group)	Cognitive Therapy Scale	-0.28 (0.09)
Rakovshik et al 2016	S-CE	IBT-S	Yes-CBT	6, 13, 17 for IBT-S	TC-3 (1 per group)	Cognitive Therapy Scale	1.31 (0.10)
	NS	IBT-CW	Yes-CBT	n/a for IBT-CW	TC-3 (1 per group)	Cognitive Therapy Scale	0.50 (0.02)
	NS	DT	No	n/a for DT	TC-3 (1 per group)	Cognitive Therapy Scale	0.06 (0.07)
Reese et al 2009A	S-CE		No	6, 17	TSE-1	Counselling Self-Estimate Inventory (COSE)	0.42 (0.12)
Reese et al 2009B	S-CE	CF	No	1, 4, 6, 17 for CF,	TSE-2 (1 per group)	Counselling Self-Estimate Inventory (COSE)	1.43 (0.18)
	S-CE	NF	No	6, 17 for NF	TSE-2 (1 per group)	Counselling Self-Estimate Inventory (COSE)	1.35 (0.11)
Ruzek et al 2012	S-CE	WTS	Yes-CBT	4, 5, 6, 7, 15, 16, 17, for WTS,	TA-4 (2 per group) TC-2 (1 per group) TCSE-2 (1 per group)	Standardized Patient Skills Assessment - Motivational Enhancement, Self-reported skills implementation, CBT Knowledge, Perceived self-efficacy	1.04 (0.04); 0.31 (0.02); 0.88 (0.03); 0.71 (0.03)
	NS	WT	Yes-CBT	n/a for WT	TA-4 (2 per group) TC-2 (1 per group) TCSE-2 (1 per group)	Standardized Patient Skills Assessment - Motivational Enhancement, Self-reported skills implementation, CBT Knowledge, Perceived self-efficacy	0.42 (0.02); 0.50 (0.02); 0.66 (0.03); 0.61 (0.03)
Safran et al 2014	S-CE	AFT16	No	16, 17 for AFT16	TC-6 (2 per group)	The Experiencing Scale – Mode and Peak	0.23 (0.10); 0.62 (0.12)
	S-CE	AFT8	No	16, 17 for AFT8	TC-6 (2 per group)	The Experiencing Scale – Mode and Peak	1.12

							(0.12); 0.53 (0.09)
	S-CE	AFTall	No	6, 16, 17 for AFTall	TC-6 (2 per group)	The Experiencing Scale – Mode and Peak	0.69 (0.05); 0.59 (0.05)
Sholomskas et al 2005	S-CE, NS	MWTS	Yes-CBT	4, 6, 17 for MWTS, n/a for MW	TA-6 (3 per group) TC-8 (4 per group)	Yale Adherence Competence Scale (YACS) Role Play 1 (functional analysis): Adherence, Skill, Role Play 2 (coping with craving): Adherence, Skill, Role Play 3 (Seemingly irrelevant decisions): Adherence, Skill CBT Knowledge Test	2.22 (0.13); 2.16 (0.12); 1.51 (0.08); 1.37 (0.07); 1.27 (0.07); 1.36 (0.07); 0.87 (0.05)
	NS	MW	Yes-CBT	n/a for MW	TA-2 (3 per group) TC-6 (3 per group)	Yale Adherence Competence Scale (YACS) Role Play 1 (functional analysis): Adherence, Skill, Role Play 2 (coping with craving): Adherence, Skill, Role Play 3 (Seemingly irrelevant decisions): Adherence, Skill, CBT Knowledge Test	1.52 (0.09); 1.61 (0.10); 0.86 (0.06); 0.74 (0.05); 0.52 (0.07); 0.48 (0.05); 0.67 (0.05)

	NS	MO	Yes-CBT	n/a for MO	TA-2 (3 per group) TC-6 (3 per group)	Yale Adherence Competence Scale (YACS) Role Play 1 (functional analysis): Adherence, Skill, Role Play 2 (coping with craving): Adherence, Skill, Role Play 3 (Seemingly irrelevant decisions): Adherence, Skill, CBT Knowledge Test	1.94 (0.12); 1.61 (0.09); 1.29. (0.07); 1.33 (0.08); 0.60 (0.05); 0.61 (0.05); 0.34 (0.04)
Smith et al 2007	S-CE		Yes-MI	4, 6, 17	TA-3	# MI Adherent Statements, # MI NonAdherent Statements, # of Total Reflections	0.90 (0.12); 0.00 (0.08); 0.68 (0.10)
Smith et al 2012	S-CE	TCS	Yes-MI	4, 6, 17	TA-21 (7 per group)	MI spirit, empathy, # of MI-non-adherent statements, reflection to question ration, % of complex reflections, % of open-ended questions, % of MI-adherent behavior	0.84, (0.05); 0.60 (0.05); -0.63. (0.05); 1.38 (0.08); 0.12 (0.04); 0.49 (0.04); 0.99 (0.07)

	S-CE	Tape	Yes-MI	4, 6, 17	TA-21 (7 per group)	MI spirit, empathy, # of MI-non-adherent statements, reflection to question ration, % of complex reflections, % of open-ended questions, % of MI-adherent behavior	0.61, (0.04); 0.49 (0.04); -0.51. (0.04); 0.53 (0.04); 0.27 (0.04); 0.82 (0.05); 0.85 (0.06)
	NS	WS	Yes-MI	n/a for NS	TA-21 (7 per group)	MI spirit, empathy, # of MI-non-adherent statements, reflection to question ration, % of complex reflections, % of open-ended questions, % of MI-adherent behavior	0.74, (0.04); 0.36 (0.04); 0.04. (0.03); 0.30 (0.04); 0.38 (0.04); -0.01 (0.03); 0.52 (0.06)
Weingardt et al 2009	S-CE	HF	Yes-CBT	6, 7 for HF	TC-2 (1 per group) TSE-2 (1 per group)	CBT Knowledge Test, Provider Efficacy Questionnaire (PEQ) – Total	0.67 (0.02); 0.74 (0.02)
	S-CE	LF	Yes-CBT	6, 7 for CF	TC-2 (1 per group) TSE-2 (1 per group)	CBT Knowledge Test, Provider Efficacy Questionnaire (PEQ) – Total	0.57 (0.02); 0.96

							(0.03)
Xavier et al 2007	S-CE		Yes-NS	5, 6	TSE-4	confidence in addressing issues related to death and dying with people with cancer, confidence in assessing and treating pain in people with cancer, confidence in helping people with cancer deal with uncertainty/anxiety about the future, confidence- assessing and treating sexuality and body image in people with cancer	0.72 (0.07); 0.45 (0.06); 1.15 (0.09); 1.13 (0.09)

Note. S-CE = supervision with common element included, S-NCE = supervision without common element,
TC = therapist competence, TA = therapist adherence, TSE = therapist self-efficacy
Common Elements: 1- beneficence for the client, 4-formal evaluation, 5-setting expectations, 6- technological considerations, 7-self-awareness of supervisor competence, 8-adherence to established guidelines, 10- modeling ethical practices, 11- providing multicultural supervision, 13- collaborative supervisory relationship, 15- goal setting, 16-supervisor encourages supervisee self-evaluation, 17-general feedback

Table 5.

Study 3: Meta-Analysis Main Effects Results for Supervision Groups

Supervision with at least 1 CSE	Outcome	*n*	*k*	δ:	*SE*	*t*	*df*	*p*	95% CI	I^2
	All Outcomes	110	29	0.787	0.117	6.75	25.8	<0.001	0.58-1.03	80.60
	Adherence	45	8	0.509	0.225	2.27	6.18	0.0625	-0.036-1.06	87.69
	Competence	54	10	0.879	0.143	6.12	17.6	<0.001	0.577-1.19	78.22
	Self-Efficacy	6	6	0.848	0.074	11.5	3.81	0.000425	0.639-1.06	19.58

Note: CSE=common supervision element; n number of effect size estimates, k number of studies, δ: random effects pooled standardized mean pre-post difference effect size. Total *n* for meta-analysis = 110, total *k* = 29.

Table 6.
Study 3: Meta-Analysis Moderator Results

Moderator				Subgroup analysis							Moderator test	
		n	*k*	ES (δ)	*SE*	*t*	*df*	*P*	95% CI	I2	Test statistic	*p*
Outcome	All Outcomes	110	29	0.787	0.117	6.75	25.8	<0.001	0.58-1.03	80.60	$F(2,107) = 0.356$	0.701
	Adherence	45	8	0.509	0.225	2.27	6.18	0.0625	-0.036-1.06	87.69		
	Competence	54	10	0.879	0.143	6.12	17.6	<0.001	0.577-1.19	78.22		
	Self-Efficacy	6	6	0.848	0.074	11.5	3.81	0.000425	0.639-1.06	19.58		
Supervision Dose	All Outcomes	105	28	n/a	n/a	n/a	n/a	n/a	n/a	n/a	$F = 2.638$, $t = -1.624$, $\beta = -0.158$	0.107
Number of Common Elements	All Outcomes	110	29	n/a	n/a	n/a	n/a	n/a	n/a	n/a	$F = 3.952$, $t = -1.988$, $\beta = -0.188$	0.049
Training		98	17	0.653	0.982	6.64	14.3	<0.001	0.442-0.863	74.13	$t(108) = 0.2501$	0.8030
	No training	34	13	0.76	0.218	3.49	11.4	0.00478	0.283-1.24	81.75		
Role-Play		57	13	0.519	0.131	3.97	10.7	0.00232	0.231-0.808	78.46	$t(108) = 0.1816$	0.2400
	No role-play	28	10	1.04	0.278	3.74	8.74	0.00483	0.41-1.68	85.23		
Fidelity Focus		84	20	0.651	0.13	5.01	17.6	<0.001	0.377-0.925	80.99	$t(108) = 0.285$	0.3000
	No fidelity focus	25	9	0.261	1.01	3.88	7.68	0.00506	0.406-1.62	82.71		
Treatment Techniques		78	20	0.681	0.146	4.67	17.5	0.000206	0.374-0.987	81.68	$t(108) = 1.4908$	0.1389
	No Treatment Techniques	7	3	0.749	0.442	1.69	1.98	0.234	-1.17-2.67	89.00		
Admin Issues		0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

	No admin issues	0	0	n/a	n/a	n/a	n/a	n/a	n/a	n/a		

Note: A = adherence, C = competence, SE = self-efficacy, n = number of effect size estimates, k = number of studies, δ: random effects pooled standardized mean pre-post difference effect size. Total *n* for meta-analysis = 110, total *k* = 29.

Element/ Moderator				Subgroup Analysis							Moderator Test	
		n	*k*	δ:	*SE*	*t*	*df*	*p*	95%	I^2	Test statistic	*p*
CE1		5	1	1.03	0.43	2.4	3.94	0.0755	-0.17-2.23	88.69	*t*(108) = 2.0729	0.0406
	No CE1	79	25	0.771	0.116	6.65	21.8	0.00000115	0.53-1.01			
CE4		64	11	0.771	0.256	3.01	9.15	0.0145	0.192 -1.35	87.16	*t*(108) = 0.3865	0.6999
	No CE4	46	20	0.781	0.127	6.17	17.6	0.00000895	0.515-1.05	75.48		
CE5		22	6	1.07	0.191	5.58	4.9	0.00272	0.572-1.56	74.60	*t*(108) = 0.5511	0.5827
	No CE5	88	24	0.698	0.128	5.44	20.8	0.0000216	0.432-0.965	79.71		
CE6		92	25	0.734	0.124	5.89	22.9	0.00000531	0.476-0.991	81.20	*t*(108) = 3.2991	0.0013
	No CE6	18	5	1.07	0.178	6.02	2.76	0.0116	0.476-1.67	63.22		
CE7		41	7	0.548	0.163	3.36	4.75	0.0218	0.122-0.975	84.11	*t*(108) = 0.4209	0.6746
	No CE7	69	23	0.848	0.135	6.28	21.1	0.0000031	0.567-1.13	78.85		
CE8		4	2	1.33	0.886	1.51	1	0.373	-9.93-12.6	82.60	*t*(108) = 1.0078	0.3158
	No CE8	106	27	0.769	0.12	6.44	24.1	0.00000116	0.523-1.02	81.12		
CE10		7	1	1.04	< 0.001	26120388519786088	1	0	1.04-1.04	83.49	*t*(108) = 0.4980	0.6195
	No CE10	103	28	0.778	0.121	6.43	24.8	0.00000101	0.529-1.03	80.88		
CE11		n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
	No CE11	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a		
CE13		2	2	1.72	0.426	4.03	1	0.155	-3.69-7.13	68.88	*t*(108) = 1.0381	0.3016
	No CE13	108	27	0.709	0.108	6.55	23.6	0.000000976	0.486-0.933	78.59		
CE15		9	1	0.737	0.00547	135	1	0.00473	0.667-0.806	83.11	*t*(108) = 0.1775	0.8594
	No CE15	101	27	0.791	0.124	6.37	24.2	0.00000134	0.535-1.05	80.53		

CE16		20	6	1.24	0.331	3.76	4.9	0.0136	0.389-2.1	83.62	$t(108) = 0.6625$	0.5091
	No CE16	90	23	0.663	0.112	5.9	19.8	0.00000933	0.429-0.898	78.49		
CE17		96	24	0.799	0.143	5.6	22	0.0000124	0.503-1.09	82.72	$t(108) = 2.7992$	0.0061
	No CE17	5	1	0.753	0.969	7.77	2.93	0.00479	0.441-1.07	65.56		

Table 7.

Study 3: Meta-Analysis Common Elements Results

Common Elements: 1- beneficence for the client, 4-formal evaluation, 5-setting expectations, 6- technological considerations, 7-self-awareness of supervisor competence, 8-adherence to established guidelines, 10- modeling ethical practices, 11- providing multicultural supervision, 13- collaborative supervisory relationship, 15- goal setting, 16-supervisor encourages supervisee self-evaluation, 17-general feedback

Figure 1. Study 2: PRISMA Diagram for Initial Review of Common Supervision Elements

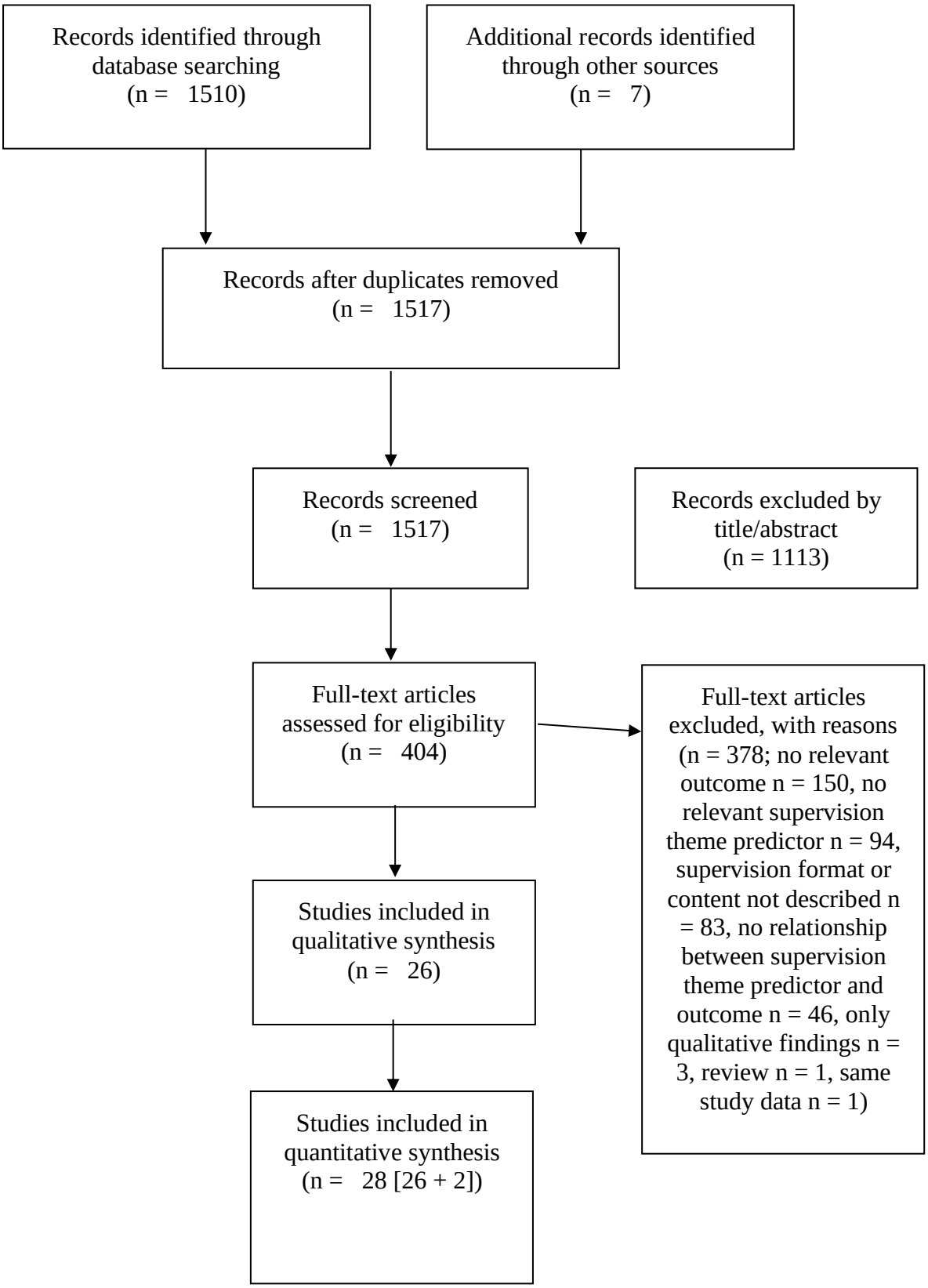


Figure 2. Study 3: PRISMA Diagram for Meta-Analysis

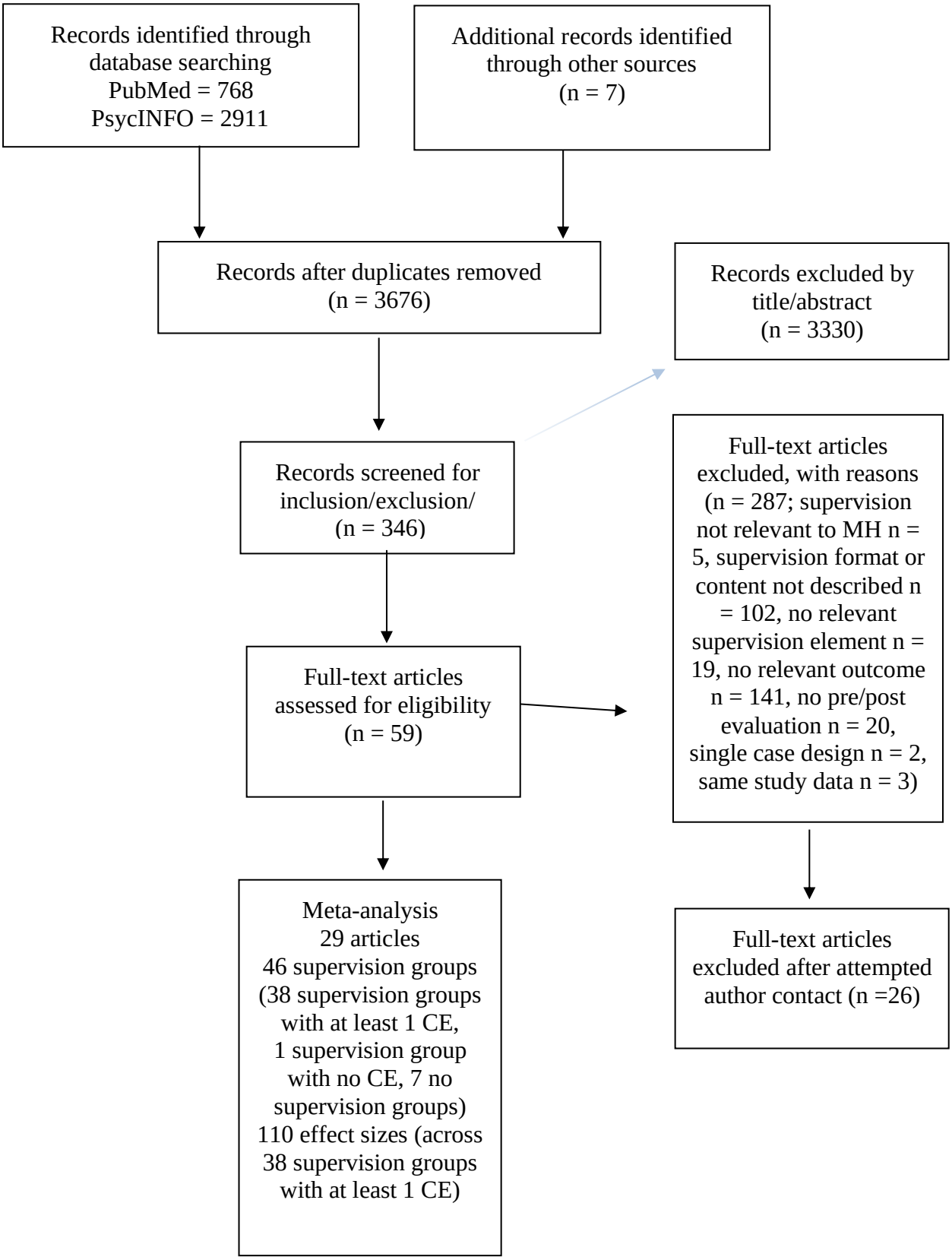
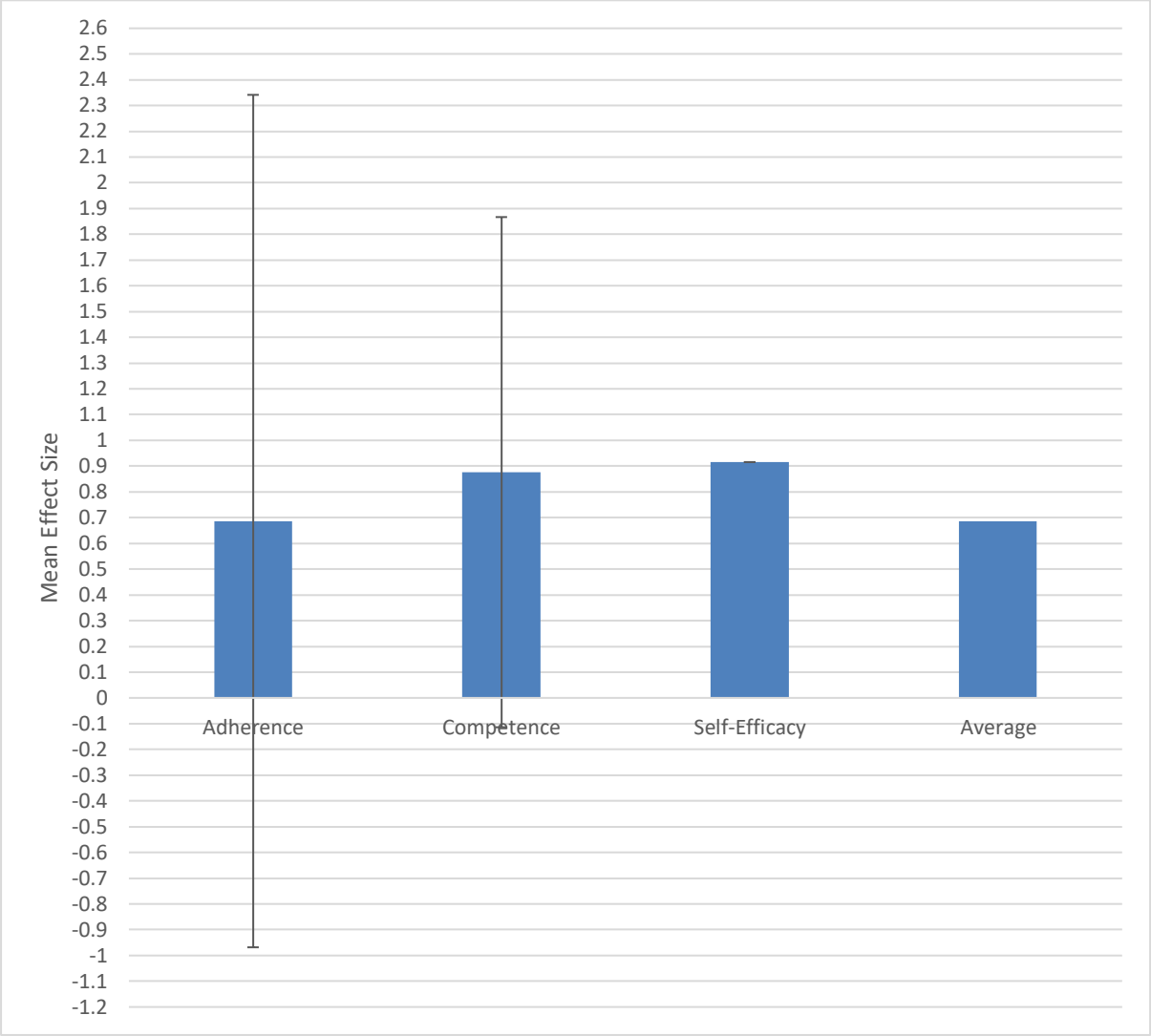


Figure 3. Study 3: Mean Effect Sizes Across Primary Outcomes



Note: Average = across adherence, competence, and self-efficacy effect sizes.

Appendix A: Study 2: Codebook Variables

Article/study type: empirical, review, proposed guidelines, between group, within-group, cross-sectional, longitudinal, other design
Setting: sample size/description (including supervisor training or qualification), graduate training, private practice, community mental health, medical/hospital, other setting
Client: population-children/adults, problem type defined, problem description
Treatment: evidence-based practice/non-evidence-based practice, description
Supervision: individual/group format, group size, length, duration, individual/group/combined frequency, in-person/phone/video conference meeting format, video/audio/live observation review, orientation, part of training/implementation effort, administrative issues covered, evidence-based practices covered, role-playing/active learning strategies used
Common elements: presence of 17 common supervision elements (described in detail in Table 1)
Outcomes: therapist/supervisee-reported, supervisor-reported, client-reported, and/or observer reported 1) therapist fidelity, 2) therapist self-efficacy, 3) therapist competence, 4) supervisory alliance, 5) therapeutic alliance, 6) client perceptions of treatment, 7) client progress; descriptions/direction of findings

Appendix B: Study 3: Meta-Analysis Search Criteria

Database	Date Searched	Search Terms Used	Number of Citations
PubMed	10/20/20	((((Psychotherapy/education[MAJR]) OR (Psychotherapy/standards[MAJR])) OR (Teaching/organization and administration[MAJR])) OR (Psychology, Clinical/education[MAJR])) AND (supervisor[Title/Abstract] OR supervisee[Title/Abstract] OR supervision[Title/Abstract] OR supervising[Title/Abstract] OR supervised[Title/Abstract])	768
PsychINFO	6/23/21	TI (supervision or supervising or supervisory or supervisor or supervisee) AND AB (supervision or supervising or supervisory or supervisor or supervisee) AND TX (competence or competency or competencies or skills or knowledge or self-efficacy or efficacy or adherence or fidelity)	2911

Appendix C: Meta-Analysis Codebook

Code	Key	Description	Average Percent Agreement
Entire Study Level			
Article Type: Published?	1= yes, 0 = no	published in a journal or book	100.00%
Article Type: Year of Publication/Study	year published	if unpublished, date of dissertation/thesis publication	100.00%
Setting: Entirely United States Sample?	1= yes, 0 = no, 9999 if not reported	0/no if sample includes participants from US and outside US, can assume based on author affiliations if not explicitly stated	100.00%
Setting: Trainees	1= yes, 0 = no	includes bachelors, masters, or doctoral students, medical residents or interns, any trainee/student participants	96.55%
Therapist/Supervisee: N	# of therapists/supervisees	baseline N	96.55%
Therapist/Supervisee: Mean Age	mean age in years, 9999 if not reported	number, can give range if average not given	82.76%
Therapist/Supervisee: Gender	% female, 9999 if not reported		75.86%
Therapist/Supervisee: Race	% White, 9999 if not reported		96.55%
Therapist/Supervisee: Discipline-Psychology	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	93.10%
Therapist/Supervisee: Discipline-Counseling	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	100.00%
Therapist/Supervisee: Discipline-Social Work	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	93.10%

Therapist/Supervisee: Discipline-Psychiatry	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	93.10%
Therapist/Supervisee: Discipline-Paraprofessional	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	93.10%
Therapist/Supervisee: Discipline-Certificate	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	96.55%
Therapist/Supervisee: Discipline-Unknown	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	93.10%
Therapist/Supervisee: Discipline-Other	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported, 8888 if 100% of participants fall into other discipline categories	79.31%
Therapist/Supervisee: Highest Degree Earned-Unknown	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	58.62%
Therapist/Supervisee: Highest Degree Earned-No Degree/Less than HS	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	82.76%
Therapist/Supervisee: Highest Degree Earned-HS/Equivalent	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	82.76%
Therapist/Supervisee: Highest Degree Earned-Bachelor's	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	79.31%
Therapist/Supervisee: Highest Degree Earned-Associates	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	68.97%
Therapist/Supervisee: Highest Degree Earned-Non-Masters Certificate	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	75.86%

Therapist/Supervisee: Highest Degree Earned-Masters	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	79.31%
Therapist/Supervisee: Highest Degree Earned-Doctorate	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	62.07%
Therapist/Supervisee: Highest Degree Earned-Other	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported, 8888 if 100% of participants fall into other discipline categories	55.17%
Therapist/Supervisee: Licensure	% licensed, 8888 for students	9999 if not reported	96.55%
Therapist/Supervisee: Years of Experience Mean	number of years as reported, 9999 if not reported		79.31%
Supervisor: N	# of supervisors	baseline N	89.66%
Supervisor: Mean Age	mean age in years, 9999 if not reported	can give range if average not given	96.55%
Supervisor: Gender	% female, 9999 if not reported		86.21%
Supervisor: Race	% White, 9999 if not reported		93.10%
Supervisor: Discipline-Psychology	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Discipline-Counseling	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Discipline-Social Work	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Discipline-Psychiatry	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Discipline-Paraprofessional	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Discipline-Certificate	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Discipline-Unknown	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%

Supervisor: Discipline-Other	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported, 8888 if 100% of participants fall into other discipline categories	79.31%
Supervisor: Highest Degree Earned-Unknown	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	79.31%
Supervisor: Highest Degree Earned-No Degree/Less than HS	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Highest Degree Earned-HS/Equivalent	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Highest Degree Earned- Bachelor's	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	86.21%
Supervisor: Highest Degree Earned-Associates	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Highest Degree Earned-Non-Masters Certificate	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	82.76%
Supervisor: Highest Degree Earned-Masters	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	89.66%
Supervisor: Highest Degree Earned-Doctorate	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported	79.31%
Supervisor: Highest Degree Earned-Other	1= yes, 0 = no	1 if any percent endorsed, 9999 if not reported, 8888 if 100% of participants fall into other discipline categories	72.41%
Supervisor: Licensure	1 = yes, 0 = no, 8888 for students	9999 if not reported	89.66%
Supervisor: Years of Experience Mean	number of years, 9999 if not reported		96.55%

Supervisor: Supervision Training, Qualifications, or Credentials Required Prior to Beginning Study?	1= yes, 0 = no, 2 = not specified, 9999 = unknown	yes if supervisors were required to take a course, obtain a certificate, complete required training in SUPERVISION (NOT treatment) to become supervisors **prior to beginning the study**, 2 if some kind of supervision training or previous supervision experience seems implied/expected but not explicitly defined. Must be separate from supervisor training provided as part of the study supervision	65.52%
Client: Population- Children/Adolescents	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	if vague/unclear, default to 1 for adults and 9999 for youth	75.86%
Client: Population- Adults	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	if vague/unclear, default to 1 for adults and 9999 for youth	82.76%
Study Type: Between Group Design?	1= yes, 0 = no	1/yes if more than one type of supervision is provided to different groups of therapists (with different therapists in each group OR the same therapists at another time as a distinct cohort/supervision group), 1/yes if active supervision group AND a control/wait list/no supervision group, 0/no if different treatment is provided across groups but supervision appears the same	100.00%

Study Type: Between Group Design-Comparison Type	1 = two or more active supervision conditions containing at least one supervision common element, 2 = one or more active supervision conditions compared with SAU (with no supervision common element), 3 = one or more active supervision conditions compared with no supervision, 4 = one or more active supervision compared with SAU AND no supervision, 8888 = N/A	8888 for single group, within group design	100.00%
Study Type: Longitudinal Study?	1= yes, 0 = no	data is collected at least 2 time points for any study group	93.10%
Study Type: Study Duration	entire study time period in weeks from baseline to last assessment collected	8888 if cross-sectional study, 1 month = 4.3 weeks	62.50%
Risk of Bias: Groups Randomly Assigned-Supervision Format?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	8888 for single group, within group design; 1/yes if supervisors and/or therapists are randomly assigned to supervision format/type; if the same group of therapists receives multiple kinds of supervision, code as between group study	82.76%
Risk of Bias: Groups Randomly Assigned-Treatment?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	8888 for single group, within group design; 1/yes if clients are randomly assigned to treatment groups, 0 no if there are multiple supervision groups but clients are not randomly assigned to the supervision groups	86.21%

Risk of Bias: Methods of Randomization Sequence Generation Described?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	8888 for single group, within group design or if no random assignment; 1/yes if description of randomization process is described (e.g., coin flip, random number generator), 0 if there is random assignment but the process is not described	89.66%
Risk of Bias: Blinding of Supervisors to Supervision Condition and/or Study Hypothesis?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	8888 for single group, within group design, 0 when condition/hypothesis could not be hidden from participants	89.66%
Risk of Bias: Blinding of Supervisees/Therapists to Supervision Condition and/or Study Hypothesis?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	8888 for single group, within group design, 0 when condition/hypothesis could not be hidden from participants	86.21%
Risk of Bias: Measure Bias-Potential Change in Therapist or Supervisor Behavior Due to Observation/Measurement?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	1/yes if therapists/supervisees or supervisors seem pressured to overperform/look better as a study participant, 0 if therapists or supervisors are unaware of study measurement process, 8888 if using knowledge or performance-based measures	82.76%
Risk of Bias: Training Dosage Differences Between Groups?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	1/yes if supervision groups complete different amounts of training in the same treatment or different treatments, 8888 if single group design and/or if no training in treatment is provided, separate from supervision	86.21%
Risk of Bias: Treatment Dosage Differences Between Groups?	1= yes, 0 = no, 9999 = unknown, 8888 = N/A	1/yes if average treatment length, session length, etc. differs substantially between supervision groups	89.66%

Risk of Bias: Blinding of Outcome Assessors?	1= yes, 0 = no, 9999 = unknown	1 if therapists/supervisees are blinded/unaware of study hypothesis, 0 when condition/hypothesis could not be hidden from participants 9999 if unclear, 8888 if self-reported therapist measure	79.31%
Risk of Bias: Intent-to-Treat Analyses?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	1/yes if analyses were conducted to address loss of data due to attrition; 8888 if no missing data reported	79.31%
Risk of Bias: Researcher Allegiance Potentially Present?	1= yes, 0 = no, 8888 = N/A, 9999 = unknown	1/yes if study author was involved in specified treatment or supervision method development, 8888 if not using specified supervision or treatment method (which may or may not have developed for the current study)	72.41%
Supervision Group Level			
Supervision Group: Was Supervision Provided to This Group By a Supervisor?	1= yes, 0 = no	0/no if group is waitlisted to receive supervision later	97.56%
Supervision Group: Did Therapists Receive Standardized Feedback or Complete Progress Monitoring Assessment?	1= yes, 0 = no	can be 1/yes even if supervision was not provided, feedback could be on client outcomes or therapist performance, includes standardized behavioral rehearsal or role play assessments, 9999 if unclear, but default to 0 if not reported	82.93%
Supervision Group-Treatment: Treatment Provided by Therapists/Supervisees?	1= yes, 0 = no, 9999 = unknown	treatment provided to real clients (or to clients therapists think are real), NOT just practicing skills with confederates/role plays	97.56%
Supervision Group-Treatment: Manualized Treatment/Protocol?	1= yes, 0 = no, 9999 = unknown, 8888 if N/A	1/yes if the treatment is named, if a citation is given, a treatment manual is explicitly used	87.80%

Supervision: Individual	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided	95.12%
Supervision: Group	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided	95.12%
Supervision: Group Size	# of supervisees per supervision group (# if stated, 8888 if N/A, 9999 if unknown)	8888 if supervision not provided, 8888 if only individual supervision provided	75.61%
Supervision: Length -> Individual Length	average supervision meeting times in minutes	8888 if supervision not provided; 9999 if unknown, can put range if stated that way, 0 if type of supervision not provided	82.93%
Supervision: Group Length	average supervision meeting times in minutes	8888 if supervision not provided; 9999 if unknown, can put range if stated that way, 0 if type of supervision not provided	85.37%
Supervision: Total Individual and/or Group Length	average supervision meeting times in minutes	8888 if supervision not provided; 9999 if unknown, can put range if stated that way, 0 if type of supervision not provided	72.00%
Supervision: Duration	supervision time period in weeks	8888 if supervision not provided, 1 month = 4 weeks, if average and rage are provided, use the average	70.59%
Supervision: Individual Frequency	1 = 1x/week, 2 = 2x/week, etc. 8888 if N/A, 9999 if unknown)	8888 if supervision not provided; 9999 if unknown, can put range if stated that way, 0 if type of supervision not provided	85.37%
Supervision: Group Frequency	1 = 1x/week, 2 = 2x/week, etc., 8888 if N/A, 9999 if unknown)	8888 if supervision not provided; 9999 if unknown, can put range if stated that way, 0 if type of supervision not provided	75.61%
Supervision: Total Individual and/or Group Frequency	1 = 1x/week, 2 = 2x/week, etc. 8888 if N/A, 9999 if unknown)	sum of individual and group frequency	78.05%

Supervision: Supervisor-Supervisee Intentionally Matched Reported?	1= yes, 0 = no, 9999 = unknown	were supervisors and supervisees intentionally matched by race, gender, experience, etc.?, fair to assume 0 if not explicitly stated, 8888 if single supervisor	90.24%
Supervision Meeting Format: In-Person	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided, assume 1 unless otherwise stated	95.12%
Supervision Meeting Format: Phone	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided, assume 0 unless otherwise stated	95.12%
Supervision Meeting Format: Video Conference	1= yes, 0 = no, 9999 = unknown	8889 if supervision not provided, assume 0 unless otherwise stated	97.56%
Supervision Review Format: Video	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided, 9999 if description is more vague/unclear	80.49%
Supervision Review Format: Audio	1= yes, 0 = no, 9999 = unknown	8889 if supervision not provided, 9999 if description is more vague/unclear	82.93%
Supervision Review Format: Live Observation	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided, assume 0 unless otherwise stated	90.24%
Supervision: Orientation Reported?	1= yes, 0 = no, 9999 = unknown	supervision NOT treatment orientation (e.g., competency-based, developmental, treatment-specific supervision if specified [i.e., in the style of the intervention, using homework in CBT supervision]); description of supervision framework or guiding principles given	90.24%

Supervision: Treatment Training Provided to Supervisors Before or During the Study?	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided; if supervisors are treatment experts and/or have had training in treatment prior to the study BUT DO NOT attend/complete training during the study (including prior to supervision period of the study) then code as 0	85.37%
Supervision: Treatment Training Provided to Supervisees Before or During the Study?	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided	87.80%
Supervision: Ongoing Quality Assurance?	1= yes, 0 = no, 9999 = unknown	for supervisors (e.g., consultation, review of supervision sessions)	87.80%
Supervision Content: Administrative Issues	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided, 0 if thorough description of supervision given and administrative issues NOT included in that description, 9999 if description is more vague/unclear	87.80%
Supervision Content: Treatment-Specific Techniques/Strategies	1= yes, 0 = no, 9999 = unknown	1/yes if treatment-specific techniques/strategies are explicitly included in supervision meetings, 0 if thorough description of supervision given and techniques or strategies NOT included in that description, 9999 if description is more vague/unclear	78.05%
Supervision Content: Focus on Fidelity to Specified Treatment Model	1= yes, 0 = no, 9999 = unknown	1/yes if a stated goal of supervision is fidelity to identified manualized/named treatment model	80.49%

Supervision Content: Role-Playing/Active Learning	1= yes, 0 = no, 9999 = unknown	8888 if supervision not provided, 0 if thorough description of supervision given and active learning/role playing NOT included in that description, 9999 if description is more vague/unclear, does NOT include observing skills modeling, MUST include the trainee practicing skills	65.85%
Supervision: Is at Least One Common Element/Theme Included?	1 = yes, 0 = no	for codes below, 8888 for groups where supervision was not provided	97.56%
Common Element/Theme 1: Beneficence/ supervisor responsibility for the client	1= yes, 0 = no, 9999 = unknown	· Supervisor assigns supervisee developmentally appropriate cases · Supervisor increases supervision time for challenging cases · Supervisor sees supervisee's clients as their responsibility	90.24%
Common Element/Theme 2: Coordination among relevant professionals	1= yes, 0 = no, 9999 = unknown	· If appropriate, supervision group includes other relevant professionals (e.g., primary care providers, psychiatrists)	100.00%
Common Element/Theme 3: Modeling professionalism	1= yes, 0 = no, 9999 = unknown	· Stated adherence to licensure or accreditation standards (e.g., referencing APA or NASW standards for licensure or accreditation in description of supervision format or processes)	100.00%

Common Element/Theme 4: Supervisor assesses supervision/ supervisee (formal/ ongoing)	1= yes, 0 = no, 9999 = unknown	· Formal evaluation of supervisee performance (separate from feedback provided in supervision meetings) · Completion of quantitative measures (e.g., competence, adherence, etc.) that are used to guide supervision feedback · Use of live observation and video/audio review techniques in formal observation	78.05%
Common Element/Theme 5: Supervisors set expectations/ contract	1= yes, 0 = no, 9999 = unknown	· Supervisors identify expected program competencies and performance standards, and assist the supervisee to formulate individual learning goals · Supervisor and supervisee make a written contract (in the case of social work, this would be a legal contract) · Supervisor engages in informed consent with supervisee at onset of supervision · includes collaborative agenda setting · supervisor communicates expectations of supervision format (i.e., setting, medium)	87.80%

Common Element/Theme 6: Technology considerations/supervisor uses technology to enhance supervision	1= yes, 0 = no, 9999 = unknown	· Supervisor uses technology to enhance the supervision feedback process (e.g., watching session videos, reviewing audio recordings) · Supervisors engage in learning (e.g., reading guidelines) to learn about the use of technology for supervision ·Supervisor uses technology to enhance the supervision format (e.g., using bug-in-the-ear/eye technology, providing supervision via videoconferencing software) · Supervisors discuss the use of technology with clients with their supervisees (e.g., use of internet searches, social media) •Supervision format uses technology to provide supervision remotely (e.g., teleconferencing, conference calls across supervision sites)	92.68%
Common Element/Theme 7: Self-awareness of supervisor competence	1= yes, 0 = no, 9999 = unknown	· Supervisor pursues continuing education · Supervisors only provide supervision within the limits of their own competence · Supervisor self-evaluation of competence · Emphasis on supervisor continuing education · Supervisor receives supervision of supervision or consultation on supervision (including peer supervision)	82.93%
Common Element/Theme 8: Adherence to established guidelines	1= yes, 0 = no, 9999 = unknown	· Reference to existing guidelines in description of supervision format (e.g., APA, NASW, ACES guidelines) ·Does NOT include reference to psychotherapy models or theoretical orientations	100.00%

Common Element/Theme 9: Maintaining appropriate relationships	1= yes, 0 = no, 9999 = unknown	· Supervisor does not engage in inappropriate relationships with supervisee (e.g., a romantic relationship)	100.00%
Common Element/Theme 10: Modeling ethical practices	1= yes, 0 = no, 9999 = unknown	· Supervision format references managing conflicts with professionalism	97.56%
Common Element/Theme 11: Providing multicultural supervision	1= yes, 0 = no, 9999 = unknown	· Supervisor infuses multicultural considerations regarding the supervisee into supervision · Supervisor builds their own multicultural awareness through various means (e.g., trainings)	95.12%
Common Element/Theme 12: Documentation of supervision	1= yes, 0 = no, 9999 = unknown	· Supervisory progress notes	100.00%
Common Element/Theme 13: Collaborative supervisory relationship	1= yes, 0 = no, 9999 = unknown	· Supervisor and supervisee creating contract together · Supervisee provides feedback regarding supervisor or supervision format	92.68%
Common Element/Theme 14: Supporting client advocacy	1= yes, 0 = no, 9999 = unknown	· Supervisor models advocacy for human rights and intervention with institutions and systems	100.00%
Common Element/Theme 15: Goal setting	1= yes, 0 = no, 9999 = unknown	· Creating a supervision contract with identified, developmentally appropriate goals · Referencing identified goals during supervision meetings	92.68%
Common Element/Theme 16: Supervisor encourages supervisee self-evaluation	1= yes, 0 = no, 9999 = unknown	· Supervisor solicits ongoing feedback from supervisee on their own competence · Supervisor incorporates that feedback into supervisee evaluation · Any mention of supervisee self-evaluation	87.80%

Common Element/Theme 17: Supervisor provides feedback to supervisee (general)	1= yes, 0 = no, 9999 = unknown	· Any review of taped/audio-recorded sessions during supervision · Any discussion of supervisee performance, default to 9999 over 0 unless explicitly stated otherwise	85.37%
Common Element/How supervision theme(s) was/were operationalized:	briefly summarize how present themes were described	use numbers in order to identify each theme (e.g., 5) supervisors set contracts with supervisees at first meeting, 12) supervision consultation record was used at each meeting)	n/a
Supervision Themes Analyzed Separately?	1= yes, 0 = no, 9999 = unknown, 8888 if N/A	1/yes if it is possible to associate specific theme with an outcome, if we can associate specific themes with multiple outcomes, 8888 if no supervision provided	97.56%
Measure Level			
MEASURE TYPE	1 = therapist adherence 2 = therapist competence 3 = therapist self-efficacy		86.21%
MEASURE NAME	full name if available, domain description if full measure name not given		n/a
MEASURE REPORTED BY:	1 = therapist, 2 = supervisor, 3 = client or client caregiver, 4 = observer, 5 = records/count-based data, 6 = performance test		82.76%
IS HIGH GOOD OR BAD?	1 = good (i.e., higher numbers are better), 0 = bad (i.e., lower numbers are better)		93.10%

PRIMARY OUTCOME MEASURE?	1 = yes, 0 = no/secondary outcome, 9999 if unknown/ unclear (default to 1/yes if only/primary measure discussed; default to 1/yes if no reason given that it is secondary to other outcomes)		93.10%
TIME SINCE FIRST/PRE MEASUREMENT	time in weeks (e.g., 2 for 2 weeks, 24 for 6 months [6 x 4 weeks]) since the first/pre assessment measure; if range given, give the average (e.g., between 26-52, it would be 39)		62.07%
IS THIS THE "TRUE" POST MEASURE?	1 = yes, 2 = FU assessment after supervision and/or treatment has ended, 0 = peri/mid-treatment		86.21%
SAME MEASURE USED AT BOTH ASSESSMENT TIME POINTS BEING COMPARED?	1 = yes, 0 = no (cannot be calculated)		100.00%
WAS NON-SUPERVISION TRAINING IN TREATMENT PROVIDED IN BETWEEN MEASUREMENTS?	1 = yes, 0 = no (0 if non-supervision training only provided PRIOR to pre-supervision/ baseline assessment), specific to treatment techniques over adoption/general implementation		85.71%
N FOR THIS MEASURE FOR SUPERVISION GROUP AT FIRST/PRE ASSESSMENT	number or 9999 for missing		79.31%

MEASURE MEAN FOR THIS SUPERVISION GROUP AT FIRST/PRE ASSESSMENT	number or 9999 for missing		89.66%
MEASURE SD FOR THIS SUPERVISION GROUP AT FIRST/PRE ASSESSMENT	number or 9999 for missing		89.66%
N FOR THIS MEASURE FOR THIS SUPERVISION GROUP AT POST/FOLLOW UP MEASUREMENT	number or 9999 for missing		75.86%
MEASURE MEAN FOR THIS SUPERVISION GROUP AT POST/ FOLLOW UP	number or 9999 for missing		89.66%
MEASURE SD FOR THIS SUPERVISION GROUP AT POST	number or 9999 for missing		82.76%

VITA

Siena Keller Tugendrajch was born in Manhattan, New York on January 3, 1992. She attended Skidmore College from 2010 to 2014, and received a Bachelor of Arts in Psychology and English in 2014. She worked as a research assistant at the Geisel School of Medicine at Dartmouth from 2014 to 2016. In 2016, She started graduate school at the University of Missouri. She received her Master of Arts in Clinical Psychology in 2018. She is currently completing her clinical internship at the Rochester Institute of Technology, which will be completed in June 2022. She will receive her Doctor of Philosophy in Clinical Psychology at the University of Missouri in July 2022. She will begin a postdoctoral research fellowship with TRAILS (Transforming Research in Action to Improve the Lives of Students) in August 2022.