EVALUATION OF PSYCHOLOGY PRACTITIONER COMPETENCE IN CLINICAL SUPERVISION

Craig J. Gonsalvez[1]

Trevor P. Crowe[2]

[1]School of Social Sciences and Psychology, University of Wollongong

[2]School of Psychology, University of Wollongong

Current contact for Craig Gonsalvez:

C.Gonsalvez@uws.edu.au

Gonsalvez, C. J., & Crowe, T. (2014). Evaluation of Psychology Practitioner Competence in Clinical Supervision. *American Journal of Psychotherapy, 68(2), 177-193*.

ABSTRACT

There is a growing consensus favouring the development, advancement and implementation of a competency-based approach for psychology training and supervision. There is wide recognition that skills, attitude-values, and relationship competencies are as critical to a psychologist's competence as are knowledge capabilities, and that these key competencies are best measured during placements, leaving the clinical supervisor in an unparalleled position of advantage to provide formative and summative evaluations on the supervisee's progression towards competence. Paradoxically, a compelling body of literature from across disciplines indicates that supervisor ratings of broad domains of competence are systematically compromised by biases including leniency and halo-biases. The current paper highlights key issues affecting summative competency evaluations by supervisors: what competencies should be evaluated, who should conduct the evaluation, how (tools) and when evaluations should be conducted, and process variables that affect evaluation. The article concludes by providing research recommendations to underpin and promote future progress, and practice recommendations to facilitate a more credible and meaningful evaluation of competence and competencies. [164 words]

Keywords: summative evaluations, competency evaluation, supervisor ratings, practitioner competence, clinical supervision

Practitioner competence[1] may refer to limited domain competence such as competence in a specific form of psychotherapy, or refer to overall professional competence extending across a range of foundational and functional domains (Fouad et al., 2009), developmental stages, and competency types such as knowledge, knowledge-application, skills, relationship, and attitude-values (Gonsalvez, et al., 2014). The current paper focuses primarily on the supervisor's summative evaluations of overall practitioner competence demonstrated by the supervisee within a formally supervised (as opposed to a peer consulting) context. The primary concern of the paper is evaluations that occur during early developmental stages before the psychologist practitioner becomes fully licensed to practice. We recognise that supervisor evaluations do not occur in a pristine laboratory setting but within a complex context involving diverse parties including employers, training institutions, supervisors, and supervisees, each with different agendas that compete for priority in a fairly limited supervision space. Further, relationship factors within the supervision room and outside it will undoubtedly influence and probably bias the nature and direction of evaluation.

The past two decades have seen a plethora of contributions on competencies, their definition and classification (Fouad, et al., 2009; Rodolfa et al., 2005), measurement, and range of implications for professional practice (Kaslow et al., 2007; Leigh et al., 2007). Regulatory authorities including licensure boards and professional societies have contributed to the impact of the competency momentum by prescribing frameworks of competencies that trainees have to demonstrate before becoming eligible to practice or attain a professional qualification and/or by revising accreditation standards. To a large extent, these contributions have been written for training institutions and, whilst they are valued additions to the

[1] For the purpose of the current paper, the definition of professional competence suggested by Epstein and Hundert (2002) is accepted. Professional competence is "the habitual and judicious use of communication, knowledge, technical skills,
clinical reasoning, emotions, values, and reflection in daily practice" (p. 226). Competencies refer to "measurable human capabilities involving knowledge, skills, and values, which are assembled in work performance" (Falender & Shafranske, 2007, p. 233).

scientific literature, they have not addressed a range of key issues relevant to the practising clinical supervisor.

Competency-based supervision paradigms are in the limelight and replacing the somewhat jaded interest in earlier developmental- and psychotherapy-based models of supervision that dominated the chat rooms and research forums in recent decades. Competency-based pedagogies constitute nothing short of a paradigm shift that will generate changes that will cascade down all levels of training and supervision, and will affect pre- and post-licensure stages. The most complex, time-intensive, and challenging aspect of these changes will undoubtedly be the summative assessment of competencies by the clinical supervisor. Undoubtedly, faithful adherence to the principles and applications of competency-based models is going to place an important but onerous responsibility on the shoulders of the clinical supervisor, who, for several reasons, is poorly prepared to confront the challenge.

First, a long and deeply respected tradition of past models of supervision is the emphasis on a facilitative-supportive role for the supervisor (Gould & Bradley, 2001). Formative feedback constitutes the primary mode of effecting change and promoting supervisee development. Competency-based paradigms demand a restructuring of supervisor roles and greater engagement with summative assessment. Second, there are few reliable and valid tools to measure supervisee competence (Ellis & Ladany, 1997). Further, training institutions have a long tradition of assessing knowledge through a range of tests, essays, and assignments but are much less resourced to examine skills and attitude-value competencies that are crucial to professional practice. Finally, most supervision within psychology has happened within a confidential space designed to foster a supportive and nurturing supervisor-supervisee relationship (Gonsalvez et al., 2013). The competency-based approach has ushered into the supervision room a third and powerful presence – that of regulatory and accreditation authorities who seek to prescribe the supervision agenda, offer guidelines for supervision methods and practice, and are invested with the authority to dictate the type, nature, and frequency of assessment within supervision. For instance within Australia, the Psychology Registration Board has revised standards to enforce accreditation for all supervisors and has recently prescribed guidelines with which supervisor training programs must comply (Psychology Board of Australia, 2013).

What Competencies Constitute Global Practitioner Competence?

It stands to reason that reliable, valid, and accurate evaluation of competence is predicated on establishing "what" core competencies need to be measured. There are several questions that are of paramount importance: What are the core domains of practitioner competence? How are they structured in relationship to one another? Are the various competency domains independent, and do they have different developmental trajectories? Recently, there have been notable strides towards a better conceptualisation and organisation of the competencies that together might constitute the competence required of a psychologist (Fouad, et al., 2009; Rodolfa, et al., 2005). There appears to be at least some consensus around a three-dimensional model of competence that involves foundational and functional domains varying across a developmental trajectory (Rodolfa, et al., 2005). There has also been an extended and more detailed explication of this model, and suggested benchmarks for competency attainment (Fouad, et al., 2009). These frameworks constitute valuable contributions to professional pedagogies and are essential for progress. Nevertheless, these frameworks have been arrived at through expert consensus and do not, at the current time, have empirical support, and are yet to demonstrate adequate predictive and construct validity. Specifically, it is unclear whether the 12 domains suggested by Rodolfa et al. (2005) or the 15 domains suggested by Fouad et al. (2009) warrant independent status; whether within domain

items are more closely linked with each other than they are with other domains; and whether a more meaningful organisational structure would emerge if, metaphorically, the many competency cards were allowed to fall based on empirical principles of coherence or proximity.

Although competencies have been evaluated by supervisors for at least several decades by competency evaluation rating forms (CERFs; Baird, 2005; Gonsalvez & Freestone, 2007; Kaslow et al., 2009; Tweed, Graber, & Wang, 2010), no more than a few studies have attempted to look at the inherent factorial structure of these ratings. There is limited research on outcomes from specific domains such as intervention (see Barber, Sharpless, Klostermann, & McCarthy, 2007; Muse & McManus, 2013) or clinical assessment (Tweed, et al., 2010), and further effort is required to evaluate the psychometric properties of specific instruments such as the Cognitive Therapy Scale (Barber, et al., 2007; McManus, Rakovshik, Kennerley, Fennell, & Westbrook, 2012).

An early study addressing the factorial structure of competency ratings has been described for social work (Bogo, Regehr, Hughes, Power, & Globerman, 2002). From field supervisor ratings of 80 competencies, seven factors (Year 2 of training) or eight factors (Year 1 of training) were identified. It is of note that despite poor predictive validity of Year 1 ratings for Year 2 ratings, the factor structure remained fairly consistent across Year 1 and Year 2 of training. Three of these factors are akin to domains within the foundational dimension of the psychology framework (Differential Use of Self, Empathy and Alliance, Values and Ethics), three factors are akin to functional domains (Intervention Planning and Implementation, Assessment, and Report Writing), and one factor has both foundational and functional elements (Presentation Skills).

Our extensive review of clinical and counselling psychology has yielded three studies that examined the dimensional structure underlying CERFs through principal components analysis (PCA) or clustering of items through statistical clustering techniques. One study examined supervisor ratings of "professional skills" from 36-item scale and reported four factors: Open-Mindedness and Social Competence in the Supervision Session, Systematic and Goal Oriented Approach to Therapy, Capacity to Create a Professional Therapeutic Relationship, and Motivating and Supporting Behaviour (Dohrenbusch & Lipka, 2006). A second study examined results from supervisor evaluation reports on clinical psychology trainees over a 12-year period. They reported that a single "generic clinical skills" factor accounted for a large proportion of the variance across the 11 broad competency domains. When the same data set was subjected to a secondary analysis using a hierarchical clustering technique, two broad sets of skills emerged: Assessment and Intervention Skills, and Professional Conduct and Interpersonal Skills (Gonsalvez & Freestone, 2007). Both studies have serious limitations. The first study was based on repeated measures from a small sample of supervisors (n=12) and trainee therapists (n=22). The second study had a larger sample (291 supervisor evaluation reports at placement[2] completion for 131 trainees), but emanated from a single training program.[2]

More recently, we reported results from a multi-site study involving five universities in Australia (Gonsalvez, et al., 2014). Supervisor evaluations at placement completion from a 60-item, clinical psychology practicum rating scale (CΨPRS) were subjected to a hierarchical clustering procedure. The results yielded preliminary validation for nine domains. These were five foundational domains: (i) Ethical Practice, (ii) Personal Capacities and Attributes, (iii)

[2]A wide range of terms are used to describe field placements (e.g., externships, rotations, internships) and the generic term 'placement' will be used in the current paper.

Response to Supervision, (iv) Scientist-Practitioner Approach, and (v) Organisation and Management Capabilities; and four functional domains: (i) Formulation and Intervention Skills, (ii) Clinical Assessment Skills, (iii) Relational Skills, and (iv) Psychometry Skills. Four super clusters were obtained from the 60 items when the proximity criterion was further relaxed: (i) Appropriate Practitioner Attributes and Conduct, (ii) Organisation and Management, (iii) Assessment and Intervention, and (iv) Psychometry capabilities.

Evaluation of Psychotherapy Competencies

Supervisor evaluation of psychotherapy competence is a complex and daunting challenge because a large diversity of psychotherapies, each founded on different theoretical assumptions and principles, offer diverse conceptualisations of what might constitute best-practice. Further, there is considerable variability among therapeutic orientations in terms of their readiness to define and gain consensus on core competencies and perhaps also on whether such an approach will yield meaningful progress. What appears obvious at this stage is that evaluation of intervention competence will, at least in part, have to be orientation specific (see Simons, Rozek, & Serrano, 2013). If a competency model is adopted, we will have to commit time and resources towards answering several questions: What competencies are central to psychotherapy competence within a specific theoretical persuasion? How are these competencies related to one another? Is there evidence for the existence of a trans-theoretical *g*-factor? What and how influential are the elements of the g-factor and the additional specific factors for the various empirically validated psychotherapies as they are delivered across client demographics, disorder types, and severity levels?

Attitudes, Attributes and Values

Demonstration of knowledge-based competencies is not the crucial problem for the psychology trainee. However, this cannot be said of attitude, attribute, and value competencies (Gonsalvez, 2014). Psychologists who have demonstrated adequate knowledge competence in terms of passing examinations continue to be disciplined because dysfunctional attitudes have led to unprofessional conduct. At the heart of ethical behaviour is a mind-set that respects and values the welfare of a client or supervisee. It is of salience that we have a considerable literature on ethical guidelines and professional conduct (in terms of the content, principles, and vignettes that will facilitate discrimination between appropriate and inappropriate ethical judgements), but have shied away from addressing a question more vital to the discipline of psychology: How might psychology training in general and supervision in particular germinate and grow healthy professional attitudes and values? To do this we should have a clear conceptualisation of the attitudes, attributes, and values that constitute the essential and desirable sets of practitioner competencies. It is fascinating that the "big" competencies with deep impact are attitude-value attributes, not knowledge competencies. For instance, the quintessence of the scientist-practitioner mind-set that is arguably the signature competency of the psychologist is not a knowledge or skill competency. An extensive knowledge of the scientific literature on empirically validated treatments is not the seminal characteristic that sires a scientist-practitioner. A deep respect for scientific evidence and high regard for its methods – both attitudes – come closer to capturing the essence of the competency. In the same vein, an expert reflective practitioner is capable of an awareness of one's experience and the cognitive and emotional experience of the "other" in the room, and is capable of a non-defensive openness to accurately appraise the impact of one's responses within the given situation. These are not knowledge, but relationship competencies that reflect healthy attitudes to self and other.

In summary, a clear notion of what competencies are important, how they congregate into domains, how domains converge into super-clusters, and whether and the extent to which key metacompetencies can drive change in other foundational and functional competencies is critical to psychology's bid to adopt a competency-based model in training and supervision. In our opinion, the frameworks of competencies currently available (e.g., Rodolfa, et al., 2005), capture well the domains of knowledge and skills relevant to the psychologist practitioner. There is a need for better clarity about attitude-value competencies, their definition and hierarchical structure, their monitoring, and measurement.

Who Should Evaluate the Supervisee's Competence?

Supervisor Evaluations

The clinical supervisor is able to observe supervisees in interaction with clients across a spectrum of disorders, a range of severity levels and complexities, and over an extended period of time. Further, the supervisor is also privy to information about professional behaviours and the supervisee's interactions with both peers and superiors within and across disciplines. In terms of the evaluation of the overall clinical competence of a supervisee, the clinical supervisor occupies an unrivalled position to form an appraisal across the many domains of the supervisee's competence. In fact, there has been a longstanding tradition to hold supervisor judgements, especially in year-long internships, with high regard. For instance, directors of training ranked internship supervisor evaluations of supervisees first, above 36 other quality assurance indices of professional training (Norcross, Stevenson, & Nash, 1986).

However, there is a surprising lack of systematic research that testifies to the reliability of supervisor judgements. Conversely, a growing body of evidence suggests that supervisor judgments are systematically compromised by halo and leniency biases (Borders & Fong, 1991; Gonsalvez, Bushnell, et al., 2013; Gonsalvez & Freestone, 2007; Lazar & Mosek, 1993). Rater reliability between judgments of supervisors and external experts about the supervisee's competence in specific domains (e.g., counselling skills) are far from satisfactory. Borders and Fong (1991) compared supervisor ratings of a trainee's global counselling skills with ratings derived from expert judges who reviewed an audiotaped counselling session and reported a weak and insignificant relationship (r = .12). It could be argued that the supervisor's ratings might capture a more global and hence more accurate profile of the supervisee's competencies whereas the objective expert rater is constrained by a limited sample (e.g., one video tape). However, there is also evidence that inter-rater reliability between supervisors and expert judges based on the same sample of supervisee performance fails to be satisfactorily high (r = .24; Dienst & Armstrong, 1988). Supervisor ratings of competencies within and across domains are very highly correlated for the same individual (Gonsalvez, et al., 2013; Gonsalvez & Freestone, 2007), although supervisor ratings of competence from a previous placement have low predictive validity, failing to predict performance of the same student on subsequent placements (Bogo, et al., 2002; Lazar & Mosek, 1993). This pattern of ratings is strongly suggestive of leniency and halo biases. Finally, the majority of supervisors themselves acknowledge that their ratings are biased and also believe that the ratings of their colleagues are biased (Robiner, Saltzman, Hoberman, Semrud-Clikeman, & Schirvar, 1998). It is possible that the supportive and nurturing role supervisors are called to play in their own therapy with clients, and the formative role they play in building up skills and confidence in an often anxious and sometimes vulnerable trainee, conspire against objective and critical supervisory judgments. From a psychodynamic perspective, it might be hard for the supervisor to remain neutral and objective as

transference and countertransference issues in the therapy room begin to be played out in supervision through parallel processes. Further, a growing body of literature testifies that concerns about the accuracy and objectivity of supervisor ratings extend to other disciplines including social work, medicine and nursing. Lazar and Mosek (1993) conclude that supervisor grades evaluating fieldwork performance are "contaminated", and that the "influence of the supervisor-trainee relationship on the grading is such that the evaluation is invalid" (p. 119).

Why has psychology, for so long, often considered the supervisor's judgment as both necessary and sufficient? Why has the discipline that prides itself as being a science and committed to the scientist-practitioner model, not made concerted efforts to enhance reliability of critical clinical competencies ratings by requiring that supervisor evaluations be complemented by ratings from other sources? Gonsalvez and Freestone (2007) lament that psychology as a discipline, has a longstanding tradition that acknowledges that evaluation of a doctoral thesis by the candidate's research supervisor may seriously compromise the objectivity of the evaluation process, yet has considered it adequate to have clinical competence at placements evaluated solely by the clinical supervisor. It is surprising that there is little research targeted at determining whether evaluation by different supervisors or repeat evaluations by the same supervisor improves the reliability and validity of supervisor evaluations. Based on the modest rater agreement among supervisors within one training program in the United Kingdom, statistical models suggest that to attain acceptable reliability and validity, evaluations such as video tape review would have to be repeated a large number of times (19 videotapes) by two examiners to ensure adequate levels of generalizability, making this impracticable in terms of staff time and resources (Keen & Freeston, 2008).

Self and Peer Evaluations of Professional Competence

There has been some research on self and peer supervisee assessments. From a pedagogic and quality assurance perspective, the use of self and peer evaluations by trainee psychologists to bolster less than reliable supervisor ratings is flawed. First, it assumes that the trainee psychologist who has not developed a sufficient level of competence to become a licensed practitioner is already endowed with the metacompetence to accurately identify what he or she (or a peer) knows and does not know. Second, whilst this assumption is problematic for evaluation of competencies within the knowledge domain, it becomes even more problematic for attitude-value competencies. A trainee who is overly threatened by and unreceptive to negative feedback is also likely to have a less than accurate evaluation of his or her own level of competence. It is no wonder that the minimal research available in this area highlights discrepancies between self and supervisor evaluations and that these discrepancies are significantly accentuated for supervisees who actually lack competence (McManus, et al., 2012). In effect, the current evidence does not support the use of self or peer evaluation as part of summative assessments, although they are certainly valuable within formative evaluation. The ability to accurately appraise one's performance, including what and how one's interactions impact on significant others in professional interactions, is a core aspect of reflective practice and there is little doubt that this important competency requires development and that supervision accords the best context for this. Thus, self-evaluation alongside supervisor evaluation is a valuable method to foster improved accuracy, depth, and effectiveness of reflective practice competencies, but at the current time we lack the pedagogical or empirical justification for self or peer evaluation to replace supervisor evaluation (see Borders, 2012).

How and When Should Competencies be Evaluated?

Kaslow et al. (2009) provide a comprehensive "toolkit" for competency assessment for professional psychology. This toolkit spans formative and summative assessments across core foundational and functional competencies. One key message is that although there are multiple, well-established assessment methods (e.g., live and delayed observation; self-assessment; reflective practice journaling; competency evaluation rating forms; case studies and reports; simulations/roleplays; 360-degree evaluations; client process and outcome data; consumer feedback; objective structured clinical examinations; portfolios; etc.), supervisors fail to ensure that pedagogic and evidence-based principles determine how and when these methods are used. First, an important tenet of competency-based approaches is that assessment methods and tools must fit the competency type under evaluation (Falender & Schafranske, 2004; Gonsalvez et al., 2014). At a basic level, essays, short-answers and multiple choice questions may be appropriate methods to evaluate knowledge acquisition but are inappropriate to evaluate skills or knowledge application. Knowledge application (e.g., ability to apply knowledge of diagnostic and differential diagnostic criteria to a specific case) and judgment (e.g., ability to apply known ethical principles to a specific situation) may be evaluated by a series of standardised scenarios in graduated steps. Skills-based competencies may be shaped and evaluated by demonstration and observation (e.g., videotapes) whereas and skills generalisation may require a sequence of role play exercises or a series of repeated observations of the specific skill across a range of situations. Consistent with these principles, a combination of self- and observer-evaluation would be required to evaluate case conceptualisation and reflective practitioner competencies (Gonsalvez et al., 2014). Regrettably, the evidence indicates that supervisors fail to select supervision methods based on sound pedagogic principles. Specifically, they overuse self report and under use observation methods (Gonsalvez et al., 2002) and these preferences may actually be replicated in the choice of evaluation methods.

In addition, normative trajectories governing the development of the evaluated competency, the evaluation purpose (e.g., formative vs. summative), and the developmental stage of the supervisee are considerations that should determine how, how often, and when evaluation occurs (Kaslow, et al., 2009). Many summative assessment tools can be used to provide formative feedback, but some formative assessment methods may lack capacity for summative assessment requirements (Muse & McManus, 2013). Considering the concept of life-long learning and required ongoing professional development (Pelling, Barletta, & Armstrong, 2009; Rodolfa, et al., 2005), the question remains, what is "competent enough" under which circumstances? Perhaps a better question to ask is, what are the learning needs of this supervisee at different stages (Watkins & Scaturo, 2013)? Such an approach would warrant repeat evaluations including at least baseline, mid- and end-placement evaluations.

Finally, there is an urgent need to establish the psychometric properties of the tools clinical supervisors typically employ for competency evaluation (Ellis & Ladany, 1997; Gonsalvez & McLeod, 2008). This lacuna is particularly evident for evaluation of overall or multiple domains of competence. The poor reliability observed for supervisor ratings gives this issue additional salience and urgency.

With regard to intervention-specific competencies, different therapeutic approaches offer a diversity of competency measures. Muse and McManus (2013) completed a systematic review of assessment methods for CBT and reported significant evidence base limitations for the use of existing CBT competence assessment methods. They concluded:

> Further psychometric evaluation and refinement of existing measures and/or the development of novel assessment tools with validated benchmarking and clear implementation protocols is needed….[and] given the complex, multi-faceted nature

> of CBT competence, multi-method assessments may ultimately be necessary in order to provide adequate assessment of all aspects of CBT competence. (p. 496).

Some researchers argue that several inherent characteristics of Likert-type scales make the scale a poor instrument for competency measurement, and have advocated that research should focus on the development of new instruments rather than on repeated attempts to improve a flawed instrument (Bogo et al., 2002; Gonsalvez et al., 2013, 2014).
One such novel assessment method is the vignette-matching procedure that employs a catalogue of standardised and calibrated vignettes to assess competencies. The vignette procedure has been shown to be superior to a conventional competency rating scale, reduced supervisor leniency and halo biases, and better differentiated levels of competence (Gonsalvez et al., 2014). Further, the method has "idiographic merits that enable the monitoring and tracking of an individual's attainment of competencies, and normative applications that help benchmark outcomes across cohorts and training programs" (p.109). Future development work is required with multi-trait and multimodal assessment methods that can be completed by service consumers and external assessors (Kaslow, et al., 2007).

There is still some debate about what should be considered the right developmental ordering of competencies even within what might be considered foundational competencies. Some of this difference of opinion may be related to parallels between the different supervisors' therapeutic allegiances and associated conceptualisations of therapeutic process and conceptualisations of supervisee learning and development processes. Further difference might be associated with the nature, custom and traditions of specific clinical placement settings and the value placed upon the various competencies. Sometimes what might be judged as competent or not might be as much an indication of the degree of "fit" or ease of assimilation the supervisee has with the assessor and/or the culture of an organisation as it is about actual competence. A supervisee who might challenge the beliefs, expectations and existing practices of a supervisor or placement organisation might trigger defensiveness, protectiveness and perhaps more critical judgement than a supervisee who mirrors sameness, thus rewarding assimilation. This might explain the finding that competence assessment in one placement does not predict competence in assessment in other placements (Bogo et al., 2002; Gonsalvez & Freestone, 2007). Therefore, objectivity in the assessment of competence may prove difficult, even if assessed externally. The challenge then is to continue to orient towards objectivity whilst recognising and directly working with the bi-directionality of dialogical engagement within the supervisory relationship and placement context.

It also seems important to acknowledge that evaluation processes and outcomes are likely to be influenced by interactive and bidirectional forces, and to examine strategies aimed at maximising the benefits of the relational elements of supervision, such as the supervisory alliance (Pearce, Beinart, Clohessy, & Cooper, 2013), directed parallel processing (Crowe, Oades, Deane, Ciarrochi, & Willliams, 2011; Tracey, Bludworth, & Glidden-Tracey, 2012), isomorphism and perpendicular intervention (Haber et al., 2009), and dialogical reflexivity (e.g., Hill, Crowe, & Gonsalvez, 2013), while simultaneously working to understand and reduce leniency and halo effects of supervisor assessment. The later point implies a bi-directional learning and assessment process for both supervisee and supervisor, as well as the supervision itself. That is, assessing supervisee competence outside of the relational and institutional context perpetuates the assumption of independence of the supervisee's performance. It fails to account for interdependence processes (including issues of non-disclosure, attachment, power, scaffolding learning and readiness for different competencies, e.g., Kaslow, et al., 2009; Ladany, Hill, Corbett, & Nutt, 1996), and perhaps favours assimilation and inculcation over reflexivity (Henry & Bruland, 2010) and transformative learning (Mezirow, 2009).

Bi-directional learning (reciprocity includes the supervisee teaching the supervisor how to supervise her/him and embracing growth potential related to the contact; e.g., Miller, et al., 2006) and assessment may be the paradigm shift that drives the required change in competency assessment. Striving to provide objective competence assessment requires supervisors to take into account the interactional influence of and on the assessment. The supervisors need to reflect on their own practice to understand their automatic or elicited responses (righting reflex, reassurance, advice-giving, anxiety relief via structure provision/skills focus, frustration, criticism etc.) and modelling responses that espouse core discipline and therapeutic values/principles. There also needs to be a reflection and dialogue about the different parts of self that are left out (non-disclosure) and how this non-acceptability (Falender & Shafranske, 2004) might perpetuate the hampering of the development of relational competencies. Dialogical reflexivity is here-now, I-Thou engagement with, and conversation about, the relational interactions between the supervisee and supervisor (Hill, et al., 2013). This is much deeper than reflection after the fact or even practicing reflexivity in process (Henry & Bruland, 2010). Dialogical reflexivity resonates with the psychodynamic self-reflective competence (Sarnat, 2010); and demands a tripartite model of supervision integrating alliance, transference, and the real relationship (Watkins, 2011). Striving for objectivity in competence assessment requires an honouring of the transformational potentials of difference/diversity, naivety (fresh set of eyes), thoughts of teaching and learning, scaffolding needs, and relational safety (e.g., disclosure anxiety; Ladany et al., 1996). Consequently, assessment of competence in supervision should parallel an assessment of supervision itself (Milne & Reiser, 2011).

The ideal competency assessment paradigm in supervision may then have both corrective emotional, cognitive, behavioural and expansive functions, is trans-theoretical in nature, and can hold non-linear learning processes (Watkins & Scaturo, 2013). We further suggest that assessment within supervision reflects elements of assimilative and transformational learning, and that this learning and assessment is bi-directional, dialogical and interpersonally vital. Put differently, the evaluation of relationship capabilities and processes, and of attitude and value competencies are critically important but also pose a major challenge for the supervisor and the supervision context.

Summary and Recommendations

There has been recent progress in terms of formulating a framework of competencies, organising them into domains and sub domains, and outlining benchmarks for their attainment. However, these frameworks are based on expert consensus. It is important to determine the dimensional structure and clustering of these many competencies and to chart the normative developmental trajectories for the diverse dimensions. Additionally, to improve effectiveness and enhance efficiency, it will be critical to gain a good understanding of whether, the circumstances when, and the extent to which changes to metacompetencies transfer across to other domains. For training institutions and the practising supervisor it will be of value to understand whether and which domains need to be targeted first. Attitude-value competencies constitute the quintessence of a competent practitioner, but are more difficult to evaluate and are less poorly understood. A better understanding of how these competencies may be monitored and enhanced during supervision is a professional issue of top priority.

The evidence that summative evaluations of global competence by a single supervisor is vulnerable to a range of rating biases has far-reaching implications for training and supervision, and warrants a review evaluation protocols and processes. First, there is an urgent need to design better new instruments and improve existing evaluation tools. In this regard, the vignette-matching procedure employed a catalogue of standardised vignettes to

deliver better evaluation outcomes than those derived from a conventional competence rating scales. Second, there needs to be a better match between what is evaluated and how the evaluation is conducted. For instance, counselling and psychotherapy skills must incorporate the use of observational methods, and conceptualisation and reflective practitioner capabilities require both self- and supervisor- evaluations of the same set of interactions. The use of appropriate assessment strategies and tasks (e.g., self report, observation, role play, self-reflective exercises) is likely to enhance discrimination between competencies and improve reliability and predictive validity of supervisor ratings. Supervisor education and training may be necessary to facilitate a more faithful adherence to competency-based evaluation practices in supervision. Third, where available, supervisors should use instruments with established psychometric properties. This is likely to be of relevance to the evaluation of psychotherapy competences (e.g., the use of the Cognitive Therapy Scale). Four, the use of the constant stimulus method or standardised cases may be of value. For instance, a series of cases ordered in progressive levels of difficulty may be used to examine diagnostic and differential diagnostic competencies or to evaluate the supervisee's capability to make appropriate ethical judgments. Finally, considering the poor reliability and validity of competency evaluations completed by field supervisors, it seems important to incorporate triangulated assessment strategies (e.g., supervisor, supervisee, external assessor, and where possible, client feedback; e.g., Kaslow, et al., 2009; Miller, Duncan, Brown, Sorrell, & Chalk, 2006; Muse & McManus, 2013). At the very least, it seems important that at least core competencies (e.g., assessment and intervention strategies and core attitude-values) are also evaluated by a third party (e.g., co-supervisor, training institution) at pre-determined times (e.g., pre-, mid, and post-placement).

Further, it seems unrealistic to expect supervisors to generate complex evaluation tools when they supervise no more than a few supervisees at a time. It is obvious that substantive progress can only be made by collaborative initiatives that engage researchers at training institutions and supervisors at placement sites. Future research could determine whether it would be advantageous to restructure the formative and summative supervisory roles concerning competence evaluation. For instance, a program of summative evaluation tasks and processes could become a joint responsibility between the training institution and the clinical supervisor with the training institution taking primary responsibility to design, in consultation with the supervisor, a program of summative evaluation tasks and processes relevant to the placement and consistent with best practice evaluation principles.

Finally, a proliferation of competencies and domains does not necessarily translate to improved evaluation outcomes. Not everything that can be measured is worth measuring. In our zeal to monitor and evaluate an ever increasing inventory of competencies, the challenge is to be innovative to ensure that we do not discount important competencies. Attitude-value and relationship-process competencies remain central to the competence of the psychology practitioner. It will be a great loss if we gradually lose sight of the true essence of the competent psychologist because these capabilities could not be reliably captured by our evaluation paradigm or measurement instruments.

References

Baird, B. N. (2005). *The internship, practicum, and field placement handbook: A guide for the helping professions* (4 ed.). New Jersey: Pearson Prentice Hall.

Barber, J., Sharpless, B., Klostermann, S., & McCarthy, K. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology - Research & Practice, 38*(5), 493-500.

Bogo, M., Regehr, C., Hughes, J., Power, R., & Globerman, J. (2002). Evaluating a measure of student field performance in direct service: Testing reliability and validity of explicit criteria. *Journal of Social Work Education, 38*(3), 385-401.

Borders, L. D. (2012). Dyadic, triadic, and group models of peer supervision/consultation: What are their components, and is there evidence of their effectiveness? *Clinical Psychologist, 16*(2), 59-71.

Borders, L., & Fong, M. L. (1991). Evaluations of supervisees: Brief commentary and research report. *The Clinical Supervisor, 9*(2), 43-51.

Crowe, T. P., Oades, L. G., Deane, F. P., Ciarrochi, J., & Williams, V. C. (2011). Parallel processes in clinical supervision: Implications for coaching menthal health practitioners. *International Journal of Evidence Based Coaching and Mentoring, 9*(2), 56-66.

Dienst, E. R., & Armstrong, P. M. (1988). Evaluation of students' clinical competence. *Professional Psychology: Research and Practice, 19*, 339-341.

Dohrenbusch, R., & Lipka, S. (2006). Assessing and predicting supervisors' evaluations of psychotherapists: An empirical study. *Counselling Psychology Quarterly, 19*(4), 395-414.

Ellis, M. V., & Ladany, N. (1997). Inferences concerning supervisees and clients in clinical supervision: An integrative review. In C. E. Watkins, Jr. (Ed.), *Handbook of psychotherapy supervision* (pp. 447-507). New York: Wiley.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *The Journal of the American Medical Association, 287*(2), 226-235.

Falender, C. A., & Shafranske, E. P. (2004). Alliance in Therapeutic and Supervisory Relationships. In C. A. Falender (Ed.), (pp. 95-114). Washington, DC, USA: American Psychological Association.

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.

Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B., . . . 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), S5-S26.

Gonsalvez, C. J. (2014). Establishing supervision goals and formalising a supervision agreement: A competency-based approach. In C. L. Watkins & D.L.Milne (Eds.), *An International Handbook of Clinical Supervision*, NY: Wiley. In Press.

Gonsalvez, C. J., Bushnell, J., Blackman, R., Deane, F., Bliokas, V., Nicholson-Perry, K., . . . Knight, R. (2013). Assessment of psychology competencies in field placements: Standardized vignettes reduce rater bias. *Training and Education in Professional Psychology, 7*(2), 99-111.

Gonsalvez, C. J., Deane, F., Matthias, M., Blackman, R., Knight, R., Allan, C., . . . Bliokas, V. (2013). Measurement of clinical psychology practicum competencies: Preliminary validation of a rating instrument. Manuscript under revision. *Training and Education in Professional Psychology*.

Gonsalvez, C. J., & Freestone, J. (2007). Field supervisors' assessments of trainee performance: Are they reliable and valid? *Australian Psychologist, 42*(1), 23-32.

Gould, L., & Bradley, L. J. (2001). Evaluation in supervision. In L. J. Bradley (Ed.), *Counselor supervision: Principles, process, and practice (3rd ed., pp. 271-303)*. New York: Brunner-Routledge.

Haber, R., Marshall, D., Cowan, K., Vanlandingham, A., Gerson, M., & Fitch, J. C. (2009). "Live" supervision of supervision: "Perpendicular" interventions in parallel processes. *The Clinical Supervisor, 28*(1), 72-90.

Henry, J., & Bruland, H. H. (2010). Educating reflexive practitioners: Casting graduate

teaching assistants as mentors in first-year classrooms. *International Journal of Teaching and Learning in Higher Education, 22*(3), 308-319.

Hill, H., Crowe, T. P., & Gonsalvez, C. J. (2013, June). *Piloting a reflective practice intervention based on videos of supervision: Evaluation and experiences of supervisors and supervisees*. Paper presented at the Advances in Clinical Supervision: Innovation and Practice, Sydney, Australia.

Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency assessment toolkit for professional psychology. *Training and Education in Professional Psychology, 3*(4), S27-S45.

Kaslow, N. J., Rubin, N., Bebeau, M., Leigh, I., Lichtenberg, J., Nelson, P., . . . Smith, L. (2007). Guiding principles and recommendations for the assessment of competence. *Professional Psychology: Research & Practice, 38*, 441-451.

Keen, A. J., & Freeston, M. H. (2008). Assessing competence in cognitive–behavioural therapy. *The British Journal of Psychiatry, 193*(1), 60-64.

Ladany, N., Hill, C. E., Corbett, M. M., & Nutt, E. A. (1996). Nature, extent, and importance of what psychotherapy trainees do not disclose to their supervisors. *Journal of Counseling Psychology, 43*(1), 10-24.

Lazar, A., & Mosek, A. (1993). The influence of the field instructor-student relationship on evaluation of students' practice. *The Clinical Supervisor, 11*(1), 111-120.

Leigh, I., Smith, L., Bebeau, M., Lichtenberg, J., Nelson, P., Portnoy, S., . . . Kaslow, N. J. (2007). Competency assessment models. *Professional Psychology: Research & Practice, 38*(5), 463-473.

Mathieson, F. M., Barnfield, T., & Beaumont, G. (2009). Are we as good as we think we are? Self-assessment versus other forms of assessment of competence in psychotherapy. *The Cognitive Behaviour Therapist, 2*, 43-50.

McManus, F., Rakovshik, S., Kennerley, H., Fennell, M., & Westbrook, D. (2012). An investigation of the accuracy of therapists' self-assessment of cognitive-behaviour therapy skills. *British Journal of Clinical Psychology, 51*(3), 292-306.

Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, M. B. (2006). Using formal feedback to improve outcome and retention: Making ongoing, real time assessment feasible. *Journal of Brief Therapy, 5*(2), 5-22.

Milne, D. L., & Reiser, R. P. (2011). Observing competence in CBT supervision: A systematic review of the available instruments. *The Cognitive Behaviour Therapist, 4*(3), 89-100.

Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review, 33*(3), 484-499.

Norcross, J. C., Stevenson, J. F., & Nash, J. M. (1986). Evaluation of internship training: Practices, problems, and prospects. *Professional Psychology: Research and Practice, 17*, 280-282.

Pearce, N., Beinart, H., Clohessy, S., & Cooper, M. (2013). Development and validation of the Supervisory Relationship Measure: A self-report questionnaire for use with supervisors. *British Journal of Clinical Psychology, 52*(3), 249-268.

Pelling, N., Barletta, J., & Armstrong, P. (2009). *The practice of clinical supervision*. Bowen Hills, QLD, Australia: Australian Academic Press.

Psychology Board of Australia. (2013). *Guidelines for supervisors and supervisión training providers*. Retrieved from Psychology Board of Australia website: http://www.psychologyboard.gov.au/Standards-and-Guidelines/Codes-Guidelines-Policies.aspx

Robiner, W. N., Saltzman, S. R., Hoberman, H. M., Semrud-Clikeman, M., & Schirvar, J. A. (1998). Psychology supervisors' bias in evaluations and letters of recommendation. *The Clinical Supervisor, 16*(2), 49-72.

Rodolfa, E. R., Bent, R. J., Eisman, E., Nelson, P. D., L, R., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice, 36*, 347-354.

Sarnat, J. (2010). Key competencies of the psychodynamic psychotherapist and how to teach them in supervision. *Psychotherapy: Theory, Research, Practice, Training, 47*(1), 20-27.

Simons, A. D., Rozek, D. C., & Serrano, J. L. (2013). Wanted: Reliable and valid measures for the science of cognitive behavioral therapy dissemination and implementation. *Clinical Psychology: Science and Practice, 20*(2), 181-194.

Tracey, T. J., Bludworth, J., & Glidden-Tracey, C. E. (2012). Are there parallel processes in psychotherapy supervision? An empirical examination. *Psychotherapy, 49*(3), 330-343.

Tweed, A., Graber, R., & Wang, M. (2010). Assessing trainee clinical psychologists' clinical competence. *Psychology Learning & Teaching, 9*(2), 50-60.

Watkins, C. E., Jr. (2011). Toward a tripartite vision of supervision for psychoanalysis and psychoanalytic psychotherapies: Alliance, transference-countertransference configuration, and real relationship. *Psychoanalytic Review, 98*(4), 557-590.

Watkins, C., & Scaturo, D. J. (2013). Toward an integrative, learning-based model of psychotherapy supervision: Supervisory alliance, educational interventions, and supervisee learning/relearning. *Journal of Psychotherapy Integration, 23*(1), 75-95.