

The Hierarchical Clustering of Clinical Psychology Practicum Competencies: A Multisite Study of Supervisor Ratings

Craig J. Gonsalvez, Clinical and Health Psychology Research Initiative, School of Social Sciences and Psychology, Western Sydney University

Frank P. Deane, Russell Blackman, and Michael Matthias, Illawarra Institute for Mental Health & School of Psychology, University of Wollongong

Roslyn Knight, School of Psychology, Macquarie University

Yasmina Nasstasia, School of Psychology, University of Newcastle

Alice Shires, Department of Psychology, University of Technology Sydney

Kathryn Nicholson Perry, Department of Psychology, Australian College of Applied Psychology

Christopher Allan, Illawarra Institute for Mental Health & School of Psychology, University of Wollongong

Vida Bliokas, Department of Psychology, Illawarra Shoalhaven Local Health District

Competency evaluation rating forms are widely used to assess a range of global and specific psychology practitioner competencies during and at the end of clinical placements. Surprisingly, there is little research examining the dimensional structure or the hierarchical clustering of items on these ratings. The current, multisite study examined supervisor ratings of clinical psychology trainees (N = 204) on the Clinical Psychology Practicum Competencies Rating Scale (CΨPRS). Based on the proximity criterion chosen, hierarchical clustering yielded either nine clusters or four super clusters: Good Practitioner Attributes and Conduct, Scientist Practitioner and Professional Management, Assessment and Intervention, and Psychological Testing. The study also tracked the developmental trajectory of competency attainment. CΨPRS ratings differentiated groups between early but not between later stages of training. Measurement issues and implications for training and practice are discussed.

Key words: competency assessment, field placement, halo bias, leniency bias, psychology internships, psychology practitioner competencies, supervisor evaluations, supervisor ratings. [*Clin Psychol Sci Prac* 22: 390–403, 2015]

Field placements are a central aspect of training programs in professional psychology. The structure, duration, casework, and supervision requirements of these placements vary across programs and across countries, but multiple placements are typically required by training programs and mandated by regulatory bodies to provide a breadth of professional experiences for trainees

Address correspondence to Craig J. Gonsalvez, School of Social Sciences and Psychology, Locked Bag 1797, Western Sydney University, NSW 2751, Australia. E-mail: c.gonsalvez@westernsydney.edu.au.

[The copyright line for this article was changed on March 30, 2016 after original online publication.]

doi:10.1111/cpsp.12123

(Nelson, 2007; Tweed, Graber, & Wang, 2010). In fact, a developmentally sequenced program of placements is a training template applied across specializations within psychology, and across allied health disciplines (Bogo, Regehr, Hughes, Power, & Globerman, 2002; Hatcher & Lassiter, 2007). This pedagogic model is designed to bridge the gap between theoretical knowledge, typically acquired within an academic institution, and competence in the world of the practitioner (Elman, Illfelder-Kaye, & Robiner, 2005). A wide range of terms are used to describe field placements (e.g., externships, rotations, internships); the generic term “placement” will be used in the current article. We adopt Epstein and Hundert’s (2002, p. 226) definition of professional competence, namely, “the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice.” Competencies refer to “measurable human capabilities involving knowledge, skills, and values, which are assembled in work performance” (Falender & Shafranske, 2007, p. 233).

The systematic monitoring of progress during and evaluation of performance at placement completion are integral components of assessment. Ongoing supervision paired with regular and systematic feedback helps shape, consolidate, and enhance knowledge and practitioner skills. In addition, structured evaluation at mid- and end placement provides summative feedback, meets requirements of training institutions, and serves as a mechanism to ensure the attainment of competence to an acceptable standard (Australian Psychology Accreditation Council, 2010; Kaslow, 2004; Kaslow et al., 2009).

Several important developments in the past decade have led to a greater emphasis on the nature, methods, and tools of field supervisor assessment (Falender & Shafranske, 2007; Gonsalvez & Freestone, 2007; Roberts, Borden, Christiansen, & Lopez, 2005). One such development is the recognition that the competency paradigm has the potential to improve professional training and practice (Kaslow et al., 2007; Roth & Pilling, 2008). Competencies across foundational and functional domains have been defined, organized, and benchmarked for different developmental stages (Fouad et al., 2009). The recognition that regular, systematic, and ecologically valid assessments constitute an essential aspect of competency-based training (Kaslow et al.,

2009; Leigh et al., 2007; Lichtenberg et al., 2007) has led to a closer scrutiny of the reliability and validity of competency assessments (Kaslow et al., 2007).

Competency assessment is a key challenge to the implementation of competency approaches (Kaslow et al., 2007; Lichtenberg et al., 2007). As a profession, we seem “better able to assess knowledge than skills or attitudes, more effective at evaluating skills than attitudes, and generally to have few established methods for assessing critical professional attitudes” (Lichtenberg et al., 2007, p. 476). Although professional psychology has tools for evaluating knowledge and skills (e.g., essays, supervisor reports), these assessments may have poor ecological validity and lack data to demonstrate good inter-rater reliability.

At the end of a placement, field supervisors typically complete a structured competency evaluation rating form (CERF) that employs a Likert scale to rate the trainee’s competence across a range of domains. CERFs are user-friendly, inexpensive to administer, easy to score, and are sufficiently versatile to measure a range of global and specific competencies (Gonsalvez et al., 2013). They are extensively used in psychology and other health disciplines, both within the United States and internationally (Baird, 2005; Gonsalvez & Freestone, 2007; Kaslow et al., 2009; Tweed et al., 2010). However, recent research has raised major concerns regarding the reliability and validity of such assessments, in particular their vulnerability to rater leniency and halo effects (Bogo et al., 2002; Gonsalvez & Freestone, 2007; Robiner, Saltzman, Hoberman, Semrud-Clikeman, & Schirvar, 1998).

Attempts to define, elaborate, and classify competencies have led to a proliferation of items on CERF-type instruments (Baird, 2005; Fouad et al., 2009; Gonsalvez & Freestone, 2007). However, increasing the item pool does not necessarily improve discrimination between competence domains or levels. Despite their popularity, there is a striking dearth of research on the CERF-type measures. Ellis and Ladany (1997) lament that there is little evidence indicating how or what is being evaluated and that supervisor evaluation of supervisee competence “may consist of many flaws bringing into question its usefulness” (p. 484). It is therefore critical that we better understand how supervisors construe competence, how they make sense of arrays of

competencies, which competencies they see as clustering together, and whether there is evidence of systematic bias influencing rater judgments.

A pioneering study examining the dimensional structure underlying CERFs through principal components analysis (PCA) has been described in social work (Bogo et al., 2002). Supervisor ratings of 80 competencies in field placements from first- ($n = 227$) and second-year students ($n = 253$) were analyzed. The PCA yielded seven (Year 2) or eight factors (Year 1), including Intervention Planning and Implementation, Differential Use of Self, Empathy and Alliance, Values and Ethics, Presentation Skills, Assessment, and Report Writing. Although the factors were consistent across years, between-supervisor reliability for Year 1 and Year 2 ratings for the same cohort of students was poor. A good understanding of the dimensional or hierarchical structure of competencies has several important implications for practitioner training in psychology. It will clarify the number of factors, their relative independence, and their generic and specific status (in the same way that *g*-factor facilitated research on intelligence). An accurate conceptualization of the structure is essential for better informed and more accurate measurement, a lacuna that is of particular salience within the current context (see Kaslow et al., 2007, 2009). Further, it will enable more accurate tracking of developmental trajectories of independent competencies/clusters and provide a blueprint for the development of more efficient practitioner-training programs. Finally, such an initiative will also provide a more informed, empirical definition of competency set boundaries, thereby helping differentiate among specializations within psychology, and between psychology and other allied disciplines. We are unaware of any study in psychology that has examined the dimensional structure underlying CERFs through principal components analysis or clustering of items through statistical clustering techniques. This study will attempt to address this issue.

Within the discipline of psychology, Gonsalvez and Freestone (2007) examined results from 291 end-placement reports on 131 clinical psychology trainees evaluated by 130 supervisors over a 12-year period. They reported that a single, “generic clinical skills” factor accounted for a large proportion of the variance. However, these results were obtained from overall domain scores (11 domains) and not itemwise scores. In a second

analysis using hierarchical clustering, two large clusters were identified: Assessment and Intervention Skills, and Professional Conduct and Interpersonal Skills.

Two studies have subjected specific domains to psychometric scrutiny. Dohrenbusch and Lipka (2006) examined 12 supervisors’ ratings of professional skills of 22 trainee therapists. Four factors were identified from a 36-item scale: 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 Behavior. More recently, Tweed et al. (2010) videotaped clinical assessment interviews conducted by clinical psychology trainees on simulated patients. Supervisors used a 33-item structured rating scale to evaluate competence from which five factors were identified: Demonstrating Professional Therapeutic Engagement, Creating a Secure Base, Formulation, Facilitating Mutual Understanding, and Session Structure.

Although attempts to define and classify competencies in terms of theoretically meaningful clusters and domains are laudable (Fouad et al., 2009) and constitute an essential first step, empirical validation of these categories is also important but has received much less attention. Most competency-based approaches to professional training espouse a developmental model that assumes a relative independence among domains. The implication is that different competency domains may have different developmental trajectories across time for both groups and individuals. For instance, it is feasible that a trainee who is yet to develop competence in intervention skills manifests appropriate knowledge, judgment, and respect for ethical principles and behaviors. In contrast, a certified professional, competent on intervention competencies, may manifest a blatant disregard for ethical values and conduct. Additionally, competencies such as case conceptualization and meta-competencies such as reflective practice and scientist practitioner attitudes may develop later, possibly even after the first developmental stage. This may be due to trainee anxiety and the challenge of unfamiliar client work early in training (Stoltenberg & McNeill, 1997). We are unaware of research that has attempted to plot these developmental trajectories. The current, multisite project was designed to address key lacunae within the competency assessment literature and had two main objectives: (a) to subject the

currently used competency rating scale to empirical scrutiny by employing a hierarchical clustering technique to determine the emergent pattern of clusters and higher-order super clusters, the advantage of the technique being that it allows the examination of either the clustering of items within a scale (relevant to this study) or the clustering of cases; and (b) to chart the profile of competencies demonstrated by four groups of trainees at different developmental levels. We predicted a stepwise increase in competence as trainees undertook four clinical placements. Further, because professional misconduct and ethical breaches are relatively uncommon, we predicted that compared to ratings on functional competencies, trainees would attain higher ratings on foundational competencies such as ethical behavior earlier in their training sequence.

METHOD

Participants

Participants were the supervisors of psychology trainees ($N = 204$) enrolled in one of the five participating universities that had clinical psychology training programs accredited by the Australian Psychology Accreditation Council (APAC) and the Clinical College of the Australian Psychological Society (APS). The trainees were enrolled in either a master's or doctoral clinical program after completion of four years of full-time psychology training at the undergraduate level. Of 204 trainees assessed in 2011, Data Set I comprised 194 trainees who had data on eight of the nine domains. Psychological Testing Skills was often not the focus of training, particularly during the first placement, so 71 trainees were not rated on this domain. Data Set II, a subset of Data Set I, consisted of the 123 participants who had ratings across all nine domains, including the Psychological Testing domain. Participant information concerning age and sex was deleted in research to ensure anonymity of students rated.

As part of their clinical training, trainees completed intensive coursework at their respective universities and concurrently enrolled in three or more field placements during a two-year period. The initial placement was usually in the university's psychology clinic, and subsequent placements occurred in external agencies. Each placement included between 200 and 300 placement hours, including a minimum of 80–100 hours of face-

to-face client contact during each placement. The vast majority of placements occurred as a two- or three-day per week commitment to working in an agency that provided psychological services. The type and nature of placements varied widely across client populations (e.g., child, adult), disorder (anxiety, mood, eating disorders), and severity levels (e.g., in- and out-patient services).

Competency ratings were completed by university clinic and field supervisors ($N = 113$) who satisfied academic and professional requirements for supervision mandated by the accrediting bodies. All supervisors were clinical psychologists who held the requisite qualifications (clinical psychology master's or doctoral degree from an accredited training institution), and who had the relevant postqualification clinical psychology experience to become eligible for full membership of the APS College of Clinical Psychologists. Summative evaluations were completed by principal field supervisors at mid- and at end placement. End-placement data from consenting supervisor–trainee dyads are presented in this study.

Materials

Clinical Psychology Practicum Competencies Rating Scale (CΨPRS). The CΨPRS is a 69-item rating scale comprising 60 individual items and nine overall domain (*Dm*) items. The scale was developed from earlier versions of similar scales used by the participating universities and the list of practicum competencies identified by Hatcher and Lassiter (2007). CΨPRS ratings are based on a four-stage developmental framework ranging from Beginner (Stage 1) through to Competent (Stage 4). Each item is rated on a 0–10 point visual analog scale ranging from Beginner (0, *Stage 1*) to Competent (10, *Stage 4*), with intermediate, equidistant anchors being *Stage 2* and *Stage 3*. Stage descriptions and sample items are included as supporting information. Supervisors rated trainees in reference to a notional absolute standard of competent professional practice, defined as comprising capabilities and skills on par with clinical psychologists working in their first job following completion of their master's degree.

Procedure

All supervisors completed the CΨPRS online on a web-based application at the completion of the place-

ment. The online format ensured that all raters completed the scale in a uniform sequence. For each of the nine domains, supervisors locked in their overall ratings of competence before completing individual items within the domain. All items within a domain were completed before the next domain was presented. Following completion of the CΨPRS, participants endorsed an option to provide or withhold consent for their de-identified data to be included in the research. The project was approved by the ethics committees of each of the participating universities.

RESULTS

Analyses Clusters and Super Clusters

Descriptive statistics for the CΨPRS end-placement data, both overall and mean scores, are provided in the supporting information. An important objective of the study was to allow an empirical process to determine the classification of items into subclusters, clusters, and super clusters. Therefore, we used a hierarchical clustering statistical technique to determine the relative proximities of the relationship between the items. A tree-clustering approach (Statistica, 2012) was employed whereby items are joined into successively larger groupings based upon the successive relaxation of the measure of similarity that initially defined their separation. As the clustering algorithm progresses through successive iterations, larger and larger clusters of increasingly dissimilar elements are aggregated. The measure of the proximity or tightness among items and clusters is termed the rescale distance unit, and ranges from 1 to 25, with shorter distances indicating greater proximity/similarity. The rescale distance is a good metric of item/cluster relatedness, in a similar way that a correlation coefficient is a good metric of the relationship among items in PCA. The clustering technique has an advantage over principal components analyses because it can be reliably applied with smaller sample sizes, and because it provides a clearer depiction of the relationship among items as they progressively link with one another to form clusters and super clusters. To examine the reliability of the results, the analyses were conducted on Data Sets I and II. Readers interested in the stepwise progression of the

clustering may view these results in the supporting information.

Competency Domains Determined by Hierarchical Clustering

Data Set I (N = 194; 8 domains; 54 items). The 54 items from eight original domains were reduced to 25 subclusters at distance unit 1, to 13 clusters at distance unit 2, to nine clusters at distance units 3 and 4 (designated as A1–A9), to eight clusters at distance unit 4 (A1–A8), to six clusters at distance unit 5 (designated as B1–B6), to four clusters at distance units 6 and 7, and to three clusters at distance unit 8 (designated as C1–C3 and termed super clusters for the current article; see Table 1).

The three-cluster solution remained stable across further distance manipulations until they reduced to two clusters at distance unit 12. Further, all five items under Ethical Practice (7a–e) were more akin to each other than they were to items on Personal Capacities (Dm 6), and the Ethics cluster was more closely linked to Personal Capacities than that was to Scientist Practitioner competencies (Dm5). Finally, items within the Ethics cluster were most dissimilar to the Psychological Testing cluster (Dm4).

The item membership structure at a rescaled distance unit of 3 (A-series, eight clusters) and 5 (B-series, six clusters) generated a number of clusters (six to eight clusters) that approximated the number of domains in the original data (eight domains, because no data were available for Psychological Testing). Specifically, at a rescale distance unit of 3 and 4, the individual items that constituted six of the original domains Dm1(A6), Dm2(A7), Dm3(A8), Dm5(A4), Dm6(A2), and Dm7 (A1) remained unchanged. There were minor changes to domain structure for two domains including Professional Skills (Dm8) and Response to Supervision (Dm9).

Specifically, Professional Skills that originally comprised nine items were subdivided into three sections: five items comprising Organization and Management Skills clustered in one domain (A5 in Table 1), two items that reflected collaborative interactions with other professionals and professional dress and demeanor clustered with the Response to Supervision domain

Table 1. Comparison between original categorization of competencies into domains (Dm) and the empirical categorization into clusters (A- and B-codes) and super clusters (C-codes) from Data Set I (e.g., A1, A2, B1, B2) and Data Set II (#codes, e.g., #A1, #A2). The differentiation between clusters and super clusters is based on the relative distance (RD) statistic metric.

| Original Domain Classification | | Results for Data Set I (N=194) in comparison with original domain structure | | | Results for Data Set II (N=123) in comparison with Data Set I | |
|--|---|---|--|---|--|--|
| Domain Description | Domain name and items | RD39 clusters | RD57 clusters | RD83 clusters | RD39 clusters | RD84 clusters |
| Knowledge of and commitment to ethical/professional codes, standards, and guidelines, and recognition of applicable circumstances. Maintains appropriate and respectful boundaries and seeks consultation on ethical issues. | Dm7. <i>Ethical Practice</i> five items | A1 <i>Ethical Practice</i> five items No change | B1 five items | C1 Super-cluster <i>Good Practitioner Attributes and Conduct</i> 24 items | #A1 No change | #C1 Super-cluster 32 items (C1 + A4) Good Practitioner Attributes and Conduct |
| | Dm6. <i>Personal Capacities and Attributes</i> 10 items | A2 <i>Personal Capacities and Attributes</i> 10 items No change | B2 19 items | | #A2 10 items No change | |
| | Dm9. <i>Response to Supervision</i> seven items | A3 <i>Response to Supervision</i> nine items 7 original items + 2 items from Dm8 (8g,8i) | | | #A3 10 items one additional item from Dm8 | |
| Knowledge of theoretical and research evidence related to diagnosis, assessment and intervention. Respect for scientific methods and empirical evidence and commitment to their application to clinical practice. | Dm5. <i>Scientist Practitioner Approach</i> three items | A4 <i>Scientist Practitioner Approach</i> three items No change | B3 eight items | C2 Super-cluster <i>Scientist-Practitioner & Professional Management Capabilities</i> 10 items | #A4 three items No change | #C2 Super-cluster <i>Organisation and Management</i> six items |
| | Effective organisation and time management for client care and management. Clear and professional expressive skills, professional dress and demeanour. Good interactional skills with colleagues and other professionals. | Dm8. <i>Professional Skills</i> nine items | A5 <i>Organisation and Management</i> five items | | #A5 three items Subset of A5 | |

| | | | | | | |
|---|---|---|-------------------------|---|--|---|
| Includes ability for empathic understanding, application of basic counselling techniques, and collaborative goal formulation with clients. | Dm1. Relational four items | A6 Relational four items No change | B4 four items | C3 Super-cluster Assessment and Intervention 20 items | #A6 four items No change | #C3 Super-cluster Assessment and Intervention 20 items |
| | Dm2. Clinical Assessment six items | A7 Clinical Assessment six items No change | B5 six items | | #A7 16 items A7 + A8 | |
| | Dm3. Formulation and Intervention 10 items | A8 Formulation and Intervention Skills 10 items No change | B6 10 items | | | |
| | Ability to appropriately conceptualise and formulate cases, generate realistic treatment plans, and monitor treatment progress and outcomes. Knowledge and skills required to conduct a range of empirically supported treatment interventions. | | | | | |
| Ability to apply knowledge to correctly select, administer, score, and interpret relevant psychometric tests. Good reporting skills. Knowledge of psychometric issues and testing theory. | Dm4. Psychological Testing six items | No data | | | #A8 Psych Testing six items No change | #C4 Super-cluster Psych Testing six items |

(Dm9), and two items (e.g., intake capabilities) that remained isolated and that were dropped from further analyses. The above structure remained stable in that it was unchanged at rescaled distance unit 4. The structure changed in minor ways at rescaled distance unit 5 (B-series). Specifically, Response to Supervision and Personal Capacities/Attributes merged into a large cluster comprising 19 items, and Scientist Practitioner Approach and Professional Skills merged into a larger cluster comprising eight items (B2 and B3 clusters in Table 1).

At rescaled distance unit 8, the domains converged into three super clusters (C-series): Good Practitioner Attributes and Conduct (C1, 24 items), Scientist Practitioner and Organization and Management Skills (C2, 10 items), and Assessment and Intervention Skills (C3, 20 items). Internal consistency measures for A-series clusters were high (Cronbach's $\alpha = 0.91$ or higher for each of the clusters).

Data Set II ($N = 123$; 9 domains; 60 items). When the clustering procedure was repeated for Data Set II, the results were a close approximation of results from Data Set I. A comparison between results obtained for the two data sets is also summarized in Table 1. The eight A-clusters (A1–A8) formed seven clusters (#A1–A7; clusters from Data Set II are designated by the #-code). Four were unchanged (A1, A2, A4, A6), and one changed marginally with a nine-item A3-cluster incorporating an additional item (#A3). The Clinical Assessment (A7) and Formulation and Intervention (A8) clusters merged earlier in the agglomeration process (#A7). Finally, the five-item Organization and Management Skills (A5) separated into Organization (#A5a, three items) and Management Skills (#A5b, two items). The six items from the Psychological Testing domain (#A9) congregated into one cluster at distance unit 3 and remained both stable and independent of items and clusters emanating from other domains, eventually constituting an independent super cluster (#C4; see details in the supporting information). Thus, adding the data from the Psychological Testing into the analysis confirms the clusters identified earlier and also suggests that Psychological Testing constitutes a separate cluster at A- and C-levels.

Split-Case Analysis. To examine the reliability of the cluster analysis, a split-case analysis was run. The results yielded strikingly similar structures, so further comments refer to the full data set.

Summary

In summary, the hierarchical clustering technique yielded results that were relatively stable across Data Sets I and II, and across the split-case analyses. There was empirical justification for the use of a nine-cluster solution (A1–A8, #A9) at a fairly strict proximity criterion (distance unit of 3), or a seven-cluster solution (B1–B6, #A9) when a more relaxed criterion was adopted (distance unit 5). When the proximity criterion was relaxed further, a four super cluster solution emerged: Good Professional Attributes and Conduct (C1), Scientist Practitioner and Professional Management capabilities (C2), Assessment and Intervention Skills (C3), and Psychological Testing Skills (#C4).

Developmental Stage by Cluster Effects

Following the determination of clusters, we assessed whether supervisors rated trainees differently across clusters and across placements. Placements occurred in

sequence and were used as a proxy for developmental stage, with earlier placements representing earlier developmental stages. Thus, developmental stage varied at four levels, determined by which of the four placements (P) were completed by the group of trainees: P1, ($n = 33$), P2, ($n = 32$), P3, ($n = 39$), and P4, ($n = 53$). Fewer trainees completed P5 and P6 ($n = 24$ in total) and were excluded from this analysis. The main analysis comprised a Placement \times Cluster ANOVA, with repeated measures for the Cluster factor ($N = 157$). The main effects for Placement and Cluster were significant: for Placement, $F(3,146) = 19.88$, $p < .001$; for Cluster, $F(7,141) = 29.02$, $p < .001$. To clarify the main effects, two Placement \times 9 Cluster (eight domains + grand mean domain score) ANOVAs were conducted for three separate contrasts: P1 versus P2, P2 versus P3, and P3 versus P4. For the Cluster factor, eight planned contrasts were performed, comparing each of the eight cluster scores against the cluster mean score. Because there were several missing values for the Psychological Testing cluster (#A9), this was analyzed separately with a smaller sample ($n = 103$) in a four Placement \times 2 Cluster (Psychological Testing and cluster mean score) ANOVA. The results are presented in Figure 1.

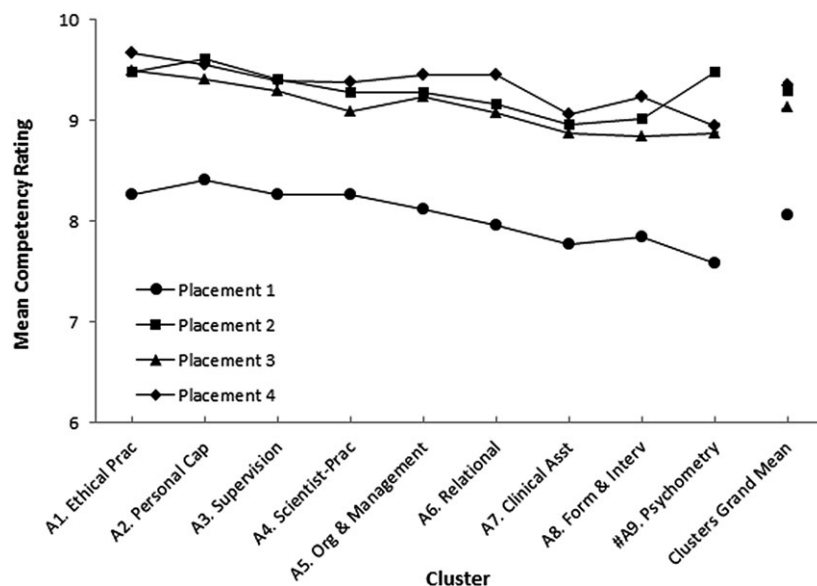


Figure 1. Mean competency ratings for clusters A1 to #A9. Note. Ethical Prac = Ethical Practice; Personal Cap = Personal Capacities; Supervision = Response to Supervision; Scientist-Prac = Scientist Practitioner Approach; Org & Management = Organization and Management; Clinical Asst = Clinical Assessment; Form & Interv = Formulation and Intervention.

Developmental Stage (Placement) Effects on Competency Scores

P1 Versus P2. As predicted, competency scores showed significant gains from P1 to P2, as demonstrated by a significant main effect for Placement, $F(1,60) = 40.36, p < .001$. The separate analysis conducted for Psychological Testing Skills yielded the same pattern of results, with higher scores for P2. Consistent with our predictions, competency scores for Ethical Practice (A1, $p < .001$), Personal Capacities and Attributes (A2, $p < .001$), and Response to Supervision (A3, $p < .005$) were higher than the grand mean scores across clusters, whereas scores on Relational Skills (A6, $p < .05$), Clinical Assessment (A7, $p < .001$), Formulation and Intervention Skills (A8, $p < .001$), and Psychological Testing Skills (#A9, $p < .001$) were lower than grand mean cluster scores (Figure 1). Scores for Scientist Practitioner and Organization and Management Skills (A4, A5) were comparable to the grand mean scores. None of the interactions between Placement and Cluster were significant.

P2 Versus P3 and P3 Versus P4. For both of these comparisons, the main effect for Placement across the eight clusters was not significant ($p > .05$; Figure 1), and similar results were obtained for the Psychological Testing cluster in independent analyses. When between-cluster comparisons were made, scores for Ethical Practice (A1, $p < .001$) and Personal Capacities and Attributes (A2, $p < .001$) were higher than grand mean cluster scores attained by the groups (for P2 versus P3, and P3 versus P4 comparisons). Response to Supervision scores were higher for the P2 versus P3 (A3, $p < .005$), but not for the P3 versus P4 comparison. In contrast, scores on Clinical Assessment (A7, $p < .001$), Formulation and Intervention Skills (A8, $p < .001$), and Psychological Testing Skills (#A9, $p < .001$) were lower than grand mean scores. Scientist Practitioner (A4), Organization and Management Skills (A5), and Relational Skills (A6) were no different from grand mean scores.

For P3 versus P4, a significant Placement by Cluster interaction further qualified between-cluster results, indicating greater improvement for Formulation and Intervention skills and Relational Skills at P4 compared with the minimal changes observed among other clusters.

For P2 versus P3, none of the interactions were significant.

Analysis of Super Clusters

The analytic strategy described above for the clusters constituting the A-series was repeated for super clusters (C-series). The results showed improvement across clusters between P1 and P2, $F(1,60) = 38.42, p < .001$, and no further changes from P2 to P3 ($p > .05$) or from P3 to P4 ($p > .05$). Within placements, competency scores for Cluster 1 were higher than the grand mean cluster score, whereas scores for Clusters 3 (Assessment and Intervention) and Cluster 4 (Psychological Testing Skills) were lower than mean scores. Competency scores for Cluster 2 (Scientist Practitioner and Professional Management capabilities) were comparable to the mean cluster score.

DISCUSSION

The study makes a valuable contribution by offering preliminary insights into the internal structure of competency ratings and how individual clusters blend together to form super clusters. As far as we are aware, this is the first study in clinical psychology that attempts to analyze the inherent clustering of competencies and to track competency profiles across developmental stages. An empirical technique (hierarchical clustering) was employed and a close level of similarity (distance unit 3) yielded a nine-cluster solution that closely replicated the nine original domains, although their constituent items were reorganized in minor but salient ways. Specifically, the reorganization produced a narrower set of items best described as Organization and Management Skills from a broader mix of items included under Professional Skills. Second, the narrower Response to Supervision domain reorganized into a broader set of items relabeled Reflective Practice and Openness to Feedback. The domains that emerged from the clustering of items included the following competency domains: Ethical Practice, Personal Capacities and Attributes, Reflective Practice and Openness to Feedback (Response to Supervision, relabeled), Scientist Practitioner, Organization and Management, Relational Skills, Clinical Assessment, Formulation and Intervention, and Psychological Testing. As might be expected, the items within the nine clusters have high internal consistencies. These resulting clusters

and super clusters were reliable in that strikingly similar clusters were obtained for Data Sets I and II and for split-case analyses.

There is broad overlap between the domains identified in this study and those outlined by Bogo et al. (2002) among social work trainees. For instance, there is obvious overlap between Ethical Practice and *Values and Ethics*, Formulation and Intervention and *Intervention Planning and Implementation*, Relational Skills and *Empathy and Alliance*, and Clinical Assessment and *Assessment* (factors identified by Bogo are in italics). There is also some overlap between the domains Reflective Practice and Openness to Feedback and *Differential Use of Self*. Scientist Practitioner and Psychological Testing emerge as clusters in clinical psychology but not in social work.

Super Clusters

The relative affiliation of competencies among themselves is enlightening, and the structure of the four super clusters has intuitive appeal. A range of important attitudes and values including a respect for the beliefs and welfare of clients and professionals (including cross-cultural values), commitment to client care, professional responsibilities, openness to feedback, a commitment to growth, and reflective practice capabilities merge into the first super cluster, Good Practitioner Attributes and Conduct. This core set of practitioner attitudes and values is likely to underpin good and ethical clinical psychology practice and is also likely to form the bedrock for good practitioners of other psychology specializations and indeed other health disciplines. Second, scientist practitioner capabilities form a kinship with effective management and organizational capabilities including effective management of time, professional demeanor, and the ability to work professionally with colleagues to comprise the Scientist Practitioner and Professional Management super cluster. Third, although an increasingly large number of discrete assessment and interventions skills are often delineated and differentiated for different client populations, Clinical Assessment, Formulation, and Intervention clusters gel into a large Assessment and Intervention super cluster. Finally, the capabilities to conduct, interpret, and report on psychological tests emerged as an independent cluster, separate from Assessment and Intervention.

Notably, the Assessment and Intervention and the Good Practitioner Attributes and Conduct super clusters were evident in a previous study that found two large clusters, Assessment and Intervention, and Interpersonal and Professional Skills (Gonsalvez & Freestone, 2007). Taken at face value, Super cluster 1 may represent a set of ethical attitudes and practitioner values that may be desirable of good psychologists and good practitioners across health disciplines. Super cluster 3 may represent knowledge and skill capabilities that underpin the acquisition of relevant assessment and intervention competencies. Of course, these core capabilities would be shaped by specialized training to evolve into independent configurations of discrete competencies relevant to specializations within and across disciplines. It is possible that the scientist practitioner mindset could be a cluster that separates psychologists from other allied health disciplines, and specific capabilities to understand and interpret psychological tests may constitute an independent set of competencies that may be required in ample measure for certain aspects of psychological practice such as educational, personality, and neuropsychological testing, and less essential to other aspects of practice such as counseling and other intervention techniques.

Admittedly, the current data provide no more than preliminary investigation into an important issue that requires systematic long-term research, and the above suggestions are offered as no more than tentative suggestions for future validation. For instance, although a fairly large sample of supervisors was used in the current study, the five clinical training programs were drawn from the state of New South Wales in Australia, where the scientist practitioner approach to professional practice and a cognitive-behavioral orientation to therapy are typically emphasized. Recent evidence points to variability in the commitment to and emphasis on evidence-based practice (e.g., Rodolfa et al., 2013), so concerns about the extent to which local, regional, and geographic factors influence outcomes at the cluster and super cluster levels are justified, and need to be pursued by future research. On the other hand, given the overlap observed between the results of the current study and a previous initiative in social work in Canada (Bogo et al.,

2002), it is likely that geographic differences play a relatively minor role.

Trajectory of Competency Development

In an overall sense, and against expectations, the CYPRS cluster scores did not support a stepwise enhancement toward competence during the two years of training. Instead, large initial competency improvements appeared to occur early in the developmental course of practitioner training (P1 versus P2), followed by relatively small and statistically insignificant changes (P3 and P4). The system of clinical psychology training adopted by the five training institutions involved in this study comprises an intensive and closely supervised program of training (incorporating regular and systematic observation and feedback) within a university clinic before additional field placements are undertaken (Gonsalvez, Hyde, Lancaster, & Barrington, 2008). At face value, the data suggest that large early gains may be followed by smaller gains later in training. This finding, if replicated, has the potential to have major implications for the way we currently conceptualize and conduct practitioner training. The assumption that progression toward competence can be charted in stepwise milestones is an attractive theoretical notion and makes for an elegant training paradigm, currently embraced by a range of disciplines (Epstein & Hundert, 2002; Fouad et al., 2009; Gonsalvez & Calvert, 2014). However, whether the empirical cards actually fall into neat, stagewise stacks is yet to be determined and certainly warrants further investigation. Admittedly, several important caveats and alternative interpretations of the current data set deserve mention and are discussed later.

The study also sought to determine whether training differentially affected competency attainment across domains. Our data suggest that at the same cross-sectional point in time, trainees were rated higher on Good Practitioner Attributes and Conduct (e.g., Ethical Practice, Personal Capacities, Response to Supervision) than on Assessment and Intervention (Super cluster 3: Relational Skills, Clinical Assessment, Formulation and Intervention) and Psychological Testing competencies (Super cluster 4). This pattern was consistent in each of the four placements examined. Further, although in an overall sense, there was little change evidenced

between P2, P3, and P4 scores, small but significant improvements were observed for the P3 versus P4 comparison on two domains—Formulation and Intervention and Relational Skills. This pattern might reflect differential growth rates among competencies within trainees, or supervisors prioritizing foundational competencies such as desirable practitioner attitudes and values (see Fouad et al., 2009) early in training. Thus, there is some evidence that developmental trajectories may vary, at least marginally, across competency domains. Further research in the area is warranted.

Are Supervisor Ratings Biased?

Frequency distributions computed for itemwise competency ratings suggest leniency effects. The lower half of the scale (ratings from 0 to 5) was used for no more than 1.6% of the ratings (range = 0.5–3.0%) across the 60 items. Although scores around the competent level (above 9) were expected at the end of P4 ($M = 9.41$) when most postgraduates would expect to graduate from their professional master's course, competence levels attained after P1 ($M = 8.39$) and P2 ($M = 9.26$; see Figure 1) are somewhat difficult to reconcile with the relatively short training periods. The possibility of leniency effects is consistent with previous research in psychology (Gonsalvez & Free-stone, 2007; Robiner et al., 1998) and in other disciplines (Bogo et al., 2002). Leniency biases affecting supervisor ratings could create a ceiling effect early in training ($M = 8.39$ in P1 and $M = 9.26$ in P2), obscuring true differences during later stages of training (P3 and P4). High ratings following initial placements are unlikely to be the result of extensive practicum experience before commencing clinical training, a practice that is common in some regions of the United States and the United Kingdom. In general, trainees in our sample did not undergo extensive practicum experience before commencing clinical training. It is possible that the formative role that supervisors are encouraged to espouse in supervision translates into a pattern of positive, encouraging, and affirming formative feedback during the placement, and to overly lenient summative ratings at the end of placement. It is of note that high supervisor ratings occurred despite our attempts to counter this trend by

explicit instructions, “ratings across placements during Clinical Masters Years 1 & 2 should reflect progression towards competency and most trainees will attain Stage 4 at course completion. Performance levels during earlier placements are likely to match Stages 1 and 2 and, as training progresses, move towards Stages 3 and 4.”

Although CERF-type rating scales are widely used in clinical psychology, several researchers have recently argued that these scales may be especially vulnerable to leniency and halo biases (Bogo et al., 2002; Gonsalvez & Freestone, 2007; Robiner et al., 1998). Among other reasons, Likert-type scales often lack sufficiently detailed behavioral anchors to facilitate discrimination between levels of competency attainment (Gonsalvez et al., 2013). Although the current study attempted to mitigate against this problem by providing a description of the four developmental stages, it is possible that broad, stagewise descriptions are insufficient and that a more satisfactory solution would require the formulation of a more systematic matrix of anchors that are domain and item-specific. The need to ensure that these benchmarks have some validity and that they can be reliably administered by supervisors in their evaluations makes such an initiative difficult and resource intensive.

In effect, it is possible that leniency effects have biased the supervisor ratings reported in our study and may also have compromised our efforts to determine differences between developmental stages. Given that the evidence indicates that leniency may be a relatively ubiquitous trend observed across countries and across disciplines, research initiatives designed to objectively monitor and measure the extent of the bias (e.g., through recorded assessment and therapy sessions evaluated by both supervisors and expert raters), as well as supervisor training specifically designed to reduce leniency, appear warranted.

Further Limitations and Future Directions

Besides an obvious need for better anchoring of rating points on the CYPRS (a problem endemic to all CERF instruments), the use of a fixed order for item administration (within and across domains) may have contributed to order and halo effects. In other words, it is possible that the temporal proximity of items

(e.g., items rated immediately before and after a specific item) inflated the kinship observed in the clustering outcomes for items within domains. A randomized or counterbalanced order across evaluators is cumbersome to administer and inconvenient for supervisors, but may help clarify and validate the structure of competencies determined by the current study. Further, the current study employed a cross-sectional design where the different developmental stages were represented by different groups of trainees. A longitudinal, within-subject design will be of value, despite the fact that longitudinal studies that examine training outcomes are often vulnerable to confounds arising out of ongoing initiatives by training institutions and supervisors to monitor and improve their training methods.

Despite the limitations, the current study offers the first investigation of the hierarchical structure of practicum competencies as they are perceived and rated by clinical field supervisors. Within a context where competency-based pedagogies and competency frameworks underpin major and systemic change to clinical training, supervision, and practice (Fouad et al., 2009; Kaslow et al., 2007, 2009; Roberts et al., 2005), the study raises (if not resolves) several fundamental but pivotal issues inherent to the competency paradigm: the true nature of their structure, the nature of their development, and problems with their measurement. Each of these issues has major theory and practice implications. The current study offers preliminary validation for a structure of competency assessment in field placements and suggests that rate of progress toward competency attainment may be both nonlinear and nonuniform across domains. It also draws attention to the likelihood of biases affecting competency evaluations. Until improved instruments and more efficient procedures facilitate the attainment of reliable and valid competency ratings, supervisors are encouraged to implement a more comprehensive assessment strategy. Such a strategy has been recommended by previous researchers and incorporates multitrait (assessing multiple domains and elements), multimethod (e.g., observation, role play, use of structured and calibrated test scenarios), and multiple raters (Gonsalvez et al., 2013; Kaslow et al., 2009; Leigh et al., 2007; Lichtenberg et al., 2007).

Above all, the current study highlights a crucial challenge to psychology's bid to align practitioner

training with the rigor of competency-based pedagogies. For the foreseeable future, it is clear that field placements will continue to remain essential and pivotal to practitioner training. It is also obvious that the lack of research on competency assessments is an important concern and that systematic analyses, innovation, and reform of competency assessment instruments are urgently required.

ACKNOWLEDGMENT

Funding for this project has been provided by the Australian Government Office for Learning and Teaching.

REFERENCES

- Australian Psychology Accreditation Council. (2010). *Rules for accreditation and accreditation standards for psychology courses*. Melbourne: Australian Psychology Accreditation Council.
- Baird, B. N. (2005). *The internship, practicum, and field placement handbook: A guide for the helping professions* (4th ed.). Upper Saddle River, NJ: Pearson-Prentice Hall.
- 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, 385–401. doi:10.1080/10437797.2002.10779106
- Dohrenbusch, R., & Lipka, S. (2006). Assessing and predicting supervisors' evaluations of psychotherapists: An empirical study. *Counselling Psychology Quarterly*, 19, 395–414. doi:10.1080/09515070601106737
- 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, NY: Wiley.
- Elman, N. S., Illfelder-Kaye, J., & Robiner, W. N. (2005). Professional development: Training for professionalism as a foundation for competent practice in psychology. *Professional Psychology: Research and Practice*, 36, 367–375. doi:10.1037/0735-7028.36.4.367
- Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, 287, 226–235. doi:10.1001/jama.287.2.226
- Falender, C. A., & Shafranske, E. P. (2007). Competence in competency-based supervision practice: Construct and application. *Professional Psychology: Research and Practice*, 38, 232–240. doi:10.1037/0735-7028.38.3.232

- 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 Suppl. 1), S5–S26. doi:10.1037/a0015832
- 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, 99–111. doi:10.1037/a0031617
- Gonsalvez, C. J., & Calvert, F. (2014). Competency-based models of supervision: Principles and applications, promises and challenges. *Australian Psychologist*, 49, 200–208. doi:10.1111/ap.12055
- Gonsalvez, C. J., & Freestone, J. (2007). Field supervisors' assessments of trainee performance: Are they reliable and valid? *Australian Psychologist*, 42, 23–32. doi:10.1080/00050060600827615
- Gonsalvez, C., Hyde, J., Lancaster, S., & Barrington, J. (2008). University psychology clinics in Australia: Their place in professional training. *Australian Psychologist*, 43, 278–285. doi:10.1080/00050060802413529
- Hatcher, R. L., & Lassiter, K. D. (2007). Initial training in professional psychology: The practicum competencies outline. *Training and Education in Professional Psychology*, 1, 49–63. doi:10.1037/1931-3918.1.1.49
- Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist*, 59, 774–781. doi:10.1037/0003-066x.59.8.774
- Kaslow, N. J., Grus, C. L., Campbell, L. F., Fouad, N. A., Hatcher, R. L., & Rodolfa, E. R. (2009). Competency assessment toolkit for professional psychology. *Training and Education in Professional Psychology*, 3(4 Suppl.), S27–S45. doi:10.1037/a0015833
- 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. doi:10.1037/0735-7028.38.5.441
- Leigh, I., Smith, L., Bebeau, M., Lichtenberg, J., Nelson, P., Portnoy, S., ... Kaslow, N. J. (2007). Competency assessment models. *Professional Psychology: Research & Practice*, 38, 463–473. doi:10.1037/0735-7028.38.5.463
- Lichtenberg, J., Portnoy, S., Bebeau, M., Leigh, I., Nelson, P., Rubin, N., ... Kaslow, N. J. (2007). Challenges to the assessment of competence and competencies.

- Professional Psychology: Research & Practice*, 38, 474–478. doi:10.1037/0735-7028.38.5.474
- Nelson, P. D. (2007). Striving for competence in the assessment of competence: Psychology's professional education and credentialing journey of public accountability. *Training and Education in Professional Psychology*, 1, 3–12. doi:10.1037/1931-3918.1.1.3
- Roberts, M. C., Borden, K. A., Christiansen, M. D., & Lopez, S. J. (2005). Fostering a culture shift: Assessment of competence in the education and careers of professional psychologists. *Professional Psychology: Research and Practice*, 36, 355–361. doi:10.1037/0735-7028.36.4.355
- 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. *Clinical Supervisor*, 16, 49–72. doi:10.1300/J001v16n02_04
- Rodolfa, E., Greenberg, S., Hunsley, J., Smith-Zoeller, M., Cox, D., Sammons, M., & Spivak, H. (2013). A competency model for the practice of psychology. *Training and Education in Professional Psychology*, 2, 71–83. doi:10.1037/a0032415
- Roth, A. D., & Pilling, S. (2008). A competence framework for the supervision of psychological therapies. Retrieved from https://iris.ucl.ac.uk/research/browse/show-publication?pub_id=88611&source_id=1
- Statistica. (2012). Clustering techniques and Statistica. Retrieved from <http://www.statsoft.com/portals/0/products/data-mining/clustering.pdf>
- Stoltenberg, C. D., & McNeill, B. W. (1997). Clinical supervision from a developmental perspective: Research and practice. In C. E. Watkins (Ed.), *Handbook of psychotherapy supervision* (pp. 184–202). New York, NY: Wiley.
- Tweed, A., Graber, R., & Wang, M. (2010). Assessing trainee clinical psychologists' clinical competence. *Psychology Learning & Teaching*, 9, 50–60. doi:10.2304/plat.2010.9.2.50
- Received December 13, 2014; revised May 21, 2015; accepted May 21, 2015.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1. Descriptive Data for Overall and Mean Competency Scores (mean of individual items) on the Clinical Psychology Practicum Competency Rating Scale (CΨPRS).

Figure S1. Hierarchical Clustering of the 54 Items from the Eight Domains ($N = 194$), with the Addition of Clustering of the Six Psychometry Items ($N = 123$).

Appendix S1. Clinical Psychology Practicum Competencies Rating Scale (CΨPRS).