RESEARCH ARTICLE

WILEY

Supervisor versus self-assessment of trainee competence: Differences across developmental stages and competency domains

Craig J. Gonsalvez[1] | Tanya Riebel[1] | Laura J. Nolan[1] |
Sonja Pohlman[2] | Warren Bartik[3]

[1]School of Psychology, Western Sydney University, Penrith, New South Wales, Australia

[2]School of Psychology, University of Newcastle, Callaghan, New South Wales, Australia

[3]School of Psychology, University of New England, Armidale, New South Wales, Australia

Correspondence
Craig J. Gonsalvez, School of Psychology, Western Sydney University, Penrith, NSW, Australia.
Email: c.gonsalvez@westernsydney.edu.au

Abstract

Objectives: This research aimed to systematically examine supervisor–trainee differences in assessments of trainee competencies across domains and developmental stages.

Methods: Trainees and supervisors ($N = 141$ dyads) independently rated trainee performance at the end of placements using the Clinical Psychology Competencies Rating Scale. Based on the number of placement hours completed at the time competence was assessed, the 141 trainees were assigned to three developmental levels (61, 42, and 31 in the groups, respectively). Trajectories of 10 different competencies and trainee–supervisor differences for these competencies were examined across three developmental levels.

Results: Compared to their supervisor ratings, trainees underestimated their competence during early stages of training, with this discrepancy reducing at Level 2 and reversing into an overestimation at Level 3. Compared to their own ratings for overall competence, trainees rated Relational and Communication, Reflective Practice, and Professionalism domains as relative strengths, and

rated their competence on assessment and intervention domains as relative weaknesses.

Conclusion: Growth trajectories derived from supervisor assessments were much flatter than trajectories derived from trainee assessments. As predicted by the impostor theory of practitioner development, trainees significantly underestimated their competence early in training. The trend for trainees to overestimate their competence toward the end of their training is a potential concern that warrants further research.

KEYWORDS
competency assessment, psychologist competence, psychology practitioner, supervisor evaluation, trainee self-assessment

1 | INTRODUCTION

Competency-based approaches have come to dominate the current training landscape and have become increasingly established internationally both in psychology and in other health disciplines. Considerable effort has been invested in the design, development, revision, and refinement of fit-for-practice competency frameworks that are meant to serve as blueprints for the education and training of competent psychology practitioners into the future (Fouad et al., 2009; Rodolfa et al., 2013).

1.1 | Developmental trajectory of competence attainment

Whilst models of competence may differ in terms of their dimensional and domain structure, all current models embrace a developmental approach to the attainment of competence (Fouad et al., 2009; Rodolfa et al., 2013). However, there is little systematic research on the milestones that anchor competency progress or a comparative evaluation of the developmental trajectories of the different competencies. Preliminary evidence suggests that competencies may not necessarily progress in a linear fashion and at a uniform rate. Instead, initial growth spurts early in training may be followed by more modest growth at later stages (Deane et al., 2018). Moreover, developmental trajectories may be domain- or context-specific (Gonsalvez et al., 2015), with some domains (e.g., functional domains such as assessment and intervention) perceived as growing more slowly than foundational domains (e.g., ethical conduct; Deane et al., 2018; Hitzeman et al., 2020). It is also possible that attainment of competence on some domains may be more challenging because of the need for specialized advanced knowledge and skills (e.g., cross-cultural and case-conceptualization competencies) or delayed because they demand additional experience, maturity, self-awareness, and attitudinal changes (e.g., scientist-practitioner and reflective practice competencies; Barrett al., 2020).

In effect, whilst there has been promising early research on competency development, there is a need for a greater focus on the mapping of the developmental courses of the discrete competencies.

1.2 | Importance of realistic self-assessment

There is a growing acknowledgment that realistic self-assessment is an important competency for practitioner training (Creed et al., 2016), and the capacity to alter one's appraisal of oneself and of others through supervisory feedback is an essential aspect of a broader, reflective practice competency (Falender & Shafranske, 2017; Fouad et al., 2009) and crucial to ensure career-long competence (APA, 2014). Practitioners are required to commit to continued education and professional development to maintain one's competence, a process that is largely self-monitored and self-directed (Falender & Shafranske, 2017). Therefore, it makes good pedagogic sense for training programs to systematically monitor self-assessments and to carefully investigate the factors that influence these assessments in both positive and negative ways.

A more compelling reason for the importance of self-assessment is its role in reflective practice and meta-competence. Reflective practice has been described as a conscious process of accurately observing, analyzing, and evaluating one's own proficiency (Boud et al., 2013). The capacity for reflection involves meta-cognitive and meta-affective elements that facilitate observation, analyses, and experience-informed growth of knowledge, skills, and attitudinal aspects of professional competence (Bennett-Levy et al., 2009; Gonsalvez & Crowe, 2014). Viewed from this perspective, reflective practice is a meta-competency involving higher order attitudes, knowledge, and skills underpinning and promoting maturity toward one's own competence. A realistic appraisal of one's performance including the ability to accurately self-assess what one knows or lacks becomes foundational to the development of reflective practice skills (APA, 2014; Falender & Shafranske, 2017). Whilst there may be different approaches to conceptualizing and differentiating the multiple elements of reflective practice (Lilienfeld & Basterfield, 2020) and a range of different strategies to enhance reflective practice competence in training (Bennett-Levy et al., 2009; Fisher et al., 2015), there is a strong consensus of the pivotal role it plays in the development and maintenance of competence (APA, 2014). Nevertheless, Kaslow and Ammirati (2020) point to a disconnect between research on basic psychological processes and reflective practice competency, limited evidence for efficacy of various reflective practices, and a significant opportunity for creating, implementing, evaluating evidence-informed educational methodologies, didactic curricula, and experiential training. Despite the longstanding interest in both the theory and practice of reflective practice (Lilienfeld & Basterfield, 2020), there has been much less attention devoted to a systematic investigation of self-assessments and their role in reflective practice.

1.3 | Differences between self-assessment versus supervisor assessment

In addition to examining how trainees evaluate their own competence, a careful examination of discrepancies between self-assessment versus supervisor assessment has relevance. First, it is unclear if trainees possess the knowledge and expertise to accurately identify thresholds that define competence and the ability to discriminate between performance levels that fall on either side of the threshold. Given the importance of self-assessment to the attainment and maintenance of competence, ongoing monitoring and evaluation of the extent to which one's self-assessment aligns with assessments by recognized experts including supervisors, becomes imperative. Second, self-assessment processes are susceptible to a range of rating biases both apparent and implicit (Karpen, 2018). On the one hand, there may be advantages for trainees to overestimate their competence. For instance, students have admitted to strategically "overestimating" self-assessment in a bid to influence their supervisor's perception of self and gain higher marks (Al-Kadri et al., 2012). An acknowledgment from a trainee that they are underperforming may be seen as an invitation for the supervisor to agree with such an assessment and follow through with a prescription of remedial action, leading to substantive financial and other disincentives for the trainee. Underestimation of one's ability may be a strategy employed by high performers to protect social support by downplaying their achievement (Tice et al., 1995) or using defensive pessimism to reduce anxiety, improve performance, and motivate themselves toward continued success (Norem & Cantor, 1986). Further, the trainee's developmental stage may have an impact on self-assessment. There is good evidence that novice practitioners often experience high levels of anxiety, self-doubt, and self-criticism, features of the impostor syndrome (Stoltenberg et al., 2014) leading to

underestimation of their actual competence across a range of situations. A recent survey of doctoral psychology students indicates that the vast majority (84%) report experiencing impostor-type symptoms (Tigranyan, et al., 2021). At later developmental stages, trainees are expected to grow both in competence and confidence, leading to more accurate self-appraisals.

In general, there is limited research that has systematically examined the level of agreement between self-assessment and other assessments of performance. A recent meta-analysis (Zell & Krizan, 2014) on self-evaluations and objective performance reports no more than modest positive correlations for both knowledge ($M = 0.27$) and skills ($M = 0.29$;). A few studies that have selectively examined competence in delivering cognitive behavior therapy demonstrated on video recordings of therapist–client sessions have reported similarly low correlations between self-rating and expert rating ($r = .10–.15$) and slightly higher correlations between self-rating and supervisor rating ($r = .21–.27$; Mathieson et al., 2008). Finally, converging multidisciplinary research suggests that some individuals are more vulnerable to distorted self-evaluations (Brosan et al., 2008). Paradoxically, trainees and professionals who are less competent also tend to have reduced self-insight, are less likely to be aware of their deficiencies (Ehrlinger et al., 2008), and also more likely to overestimate their competence (Carter & Dunning, 2008; Parker & Waller, 2015), leading to less responsive strategies for improving accuracy. It has been estimated that up to 10% of trainees in postgraduate mental health programs may have limited suitability for working with clients (Gaubatz & Vera, 2002) and that some of these are not identified or received remediation during their programs. A deeper understanding of self-assessment biases requires research to go beyond correlational findings to a scrutiny of the magnitude (size of discrepancies) and direction of the bias (under- and overestimation), and to also investigate the psychosocial mechanisms that may contribute to these effects (Karpen, 2018; Ward et al., 2002).

1.4 | The current study

In summary, reliable and valid assessment of competence is an important foundational pillar of competency-based approaches, and training students to reliably evaluate themselves is an important competency in its own right. Further, accurate self-assessment is essential to higher order meta-competence such as reflective practice. In contrast, inaccurate self-assessment is likely to lead to overconfident but incompetent practitioners and jeopardize the safe and effective treatment of clients (Kaslow et al., 2009; Mathieson et al., 2008). Preliminary research examining self-assessment versus supervisor assessment differences has yielded promising results, but several studies have methodological inadequacies including small sample sizes (e.g., Hitzeman et al., 2020) and different assessment measures for trainees and supervisors, making direct comparison problematic (Brosan et al., 2008). Moreover, it is important to evaluate trainee–supervisor rating differences across the range of competency domains and to track these differences across developmental stages. Little is known about the accuracy and discrepancies regarding trainee and assessor and importantly whether and how placement experience and training will affect discrepancy scores. An enhanced understanding of the developmental trajectory of competence has the potential to inform the development of more effective remediation plans, improve supervision effectiveness, and maximize client outcomes. The current study aims to chart perceived trajectories of competence development as assessed by supervisors and trainees for a range of discrete competencies and to study their similarities and differences. A specific prediction was that trainees would be influenced by features of the impostor syndrome and underestimate their competence early in training with this bias reducing across developmental stages.

2 | MATERIALS AND METHODS

2.1 | Participants

The data set comprised competency ratings provided at the end of placements by trainees (T) and supervisors (S) during a 2-year period (2019–2020). The data were sourced from a multisite study on competency assessments.

Only placements that had both trainees' self-ratings and supervisor ratings of trainee competence were included. From an initial pool of 178 possible data matches, 37 ratings were excluded from further analyses because they failed to meet inclusion criteria, leaving a final data set that included 141 T–S dyads. Specifically, ratings were excluded when trainees and supervisors failed to provide ratings within the designated 2-week time frame, when beginning and end-placement dates provided by members of the dyad did not match, or when names provided for the supervisor and trainee for the placement were insufficient to identify a match (e.g., names were truncated to "Jo" that identified more than one possible match).

2.1.1 | Trainees

The trainees ($N = 141$; females 77%; males 21%; other 2%) were enrolled in a clinical psychology degree (Master of Clinical Psychology, PsyD/DPsyc, or PhD) in Australia and were undertaking their placements (clinical psychology rotations) as part of their training. The high proportion of female trainees in the sample is representative of clinical psychology enrollments in Australia and other developed countries.

Trainees in Australia typically follow a two-step process to become a registered psychologist with a clinical psychology endorsement in Australia. In Step 1, in addition to coursework and research requirements, trainees complete a minimum of 1000 placement hours including a minimum of 400 direct, face-to-face client-contact hours. Whilst the prescribed number of total placement and face-to-face hours are mandated, trainees are allowed to accrue these hours in a flexible manner (e.g., through 3–5 individual placements). In Step 2, after completion of requirements for generic registration, trainees complete a 2-year full-time clinical psychology registrar training to meet practicum requirements to achieve a clinical psychology endorsement. For the current research, only trainees undertaking Step 1 of their training were included. Each placement involved working under supervision for 1–2 days a week for a 3–5-month period, involving a minimum of 200 placement hours including 80 hours of direct, client contact. The typical sequence included an initial placement within a university psychology clinic followed by two or three externally supervised placements/rotations in accredited agencies (i.e., hospitals, out-patient clinics, agencies providing mental health services). To track the developmental trajectory of competency development, we separated the 141 trainees into three developmental levels based on the number of practicum hours completed at the time of their competency assessments: Level 1 ($n = 67$; up to 299 placement hours); Level 2 ($n = 43$; 300–599), and Level 3 ($n = 31$; 600 and above).

2.1.2 | Supervisors

All supervisors (male 25%, female 67%, and other/missing data 8%) met the Psychology Board's requirements to provide clinical supervision for clinical psychology trainees. These requirements included the prescribed educational qualifications, full and current registration to practice as a clinical psychologist, postqualification experience in delivering clinical psychology services (minimum 2 years, full-time), and board-approved supervisor training. Preferred therapeutic orientations of supervisors included: cognitive behavioral therapy (60.6%), eclectic therapy (18.2%), and other orientations (21.3%, i.e., acceptance and commitment, schema, and family therapy). Supervisors were experienced both as practitioners (mean years of clinical experience = 9.58; SD = 1.34) and as clinical supervisors (M = 7.48, SD = 2.86).

2.2 | Measures

2.2.1 | Clinical Psychology Practicum Competencies Rating Scale (CΨPRS)

The current study used a 40-item version of the CΨPRS that assesses 40 specific competencies across 11 distinct competency domains. A copy of the instrument is available in the Supporting Information: Material. Each of the

40 CΨPRS competencies was rated on a 1.0–4.9-point visual analog scale, ranging from beginner (1.0–1.9, stage 1) to competent (4.0–4.9, stage 4), with stage 2 (2.0–2.9) and stage 3 (3.0–3.9) representing intermediate anchors. The 40 items are presented online in a random sequence and yield 40 discrete competency scores (from 1.0 to 4.9). The 11 domain scores are obtained by averaging scores across the specific items within each domain and a grand mean score is obtained by averaging the domain scores for each rater. The CΨPRS has good psychometric properties including high internal consistency (Cronbach's α = .91; Gonsalvez et al., 2015) and good convergent validity (Gonsalvez et al., 2021). The 11-domain structure used by the current study is supported by results from a stepwise hierarchical cluster analysis indicating a close relationship between items included within each domain and a more distant relationship between items across domains (Gonsalvez et al., 2020). For the current study, data from 10 domains are reported. D1-Diagnosis and Clinical Assessment: Knowledge and Knowledge Application; D2-Case Conceptualization and Interventions: Knowledge and Knowledge Application; D3-Individual and Cultural Diversity; D4-Effective and Skillful Assessment and Intervention; D5-Relational and Communication; D6-Reflective Practice; D7-Scientist-Practitioner: Attitudes and Values; D8-Ethical Practice: Knowledge and Knowledge Application; D9-Professionalism: Organized and Disciplined Practice; D10-Professionalism: Attitudes and Values. The 11th domain, *Psychological Testing was* omitted from analyses because several supervisors indicated that the placement did not focus on psychological testing and provided "Not-applicable" ratings.

2.3 | Procedure

Supervisors and trainees completed the CΨPRS at the end of the trainee's placement within 2-weeks of one another and, at the time of completion, supervisors and trainees were each blind to the other's ratings. In accordance with policy, supervisors' ratings were communicated to trainees, and trainees' self-ratings were provided to supervisors to enable supervisors to initiate a discussion of and feedback on T–S rating differences should supervisors choose to do so. Participants provided informed consent and the research was approved by the Ethics Committees of each of the institutions involved (approval number H10828).

2.4 | Analyses

Mean scores for each of the 10 domains (from the 40 items) were computed separately for each supervisor's and trainee's ratings within each dyad. The statistical analyses (SPSS version 29) were conducted in two steps. First, trainees' self-ratings were analyzed using a 3 Levels × 10 Domains analysis of variance (ANOVA). The data from the three levels were derived from different groups of trainees, so levels are a between-subjects factor. The data from the 10 domains constitute a within-subjects factor and were analyzed using a repeated-measures ANOVA. For the domain factor, 10 a priori, planned contrasts were performed, comparing mean scores for each of the 10 domains against the grand mean (mean score from 10 domains). Two contrasts (Level 1 vs. Level 2 and Level 2 vs. Level 3) were tested for the developmental stage.

For the second analysis, difference scores (δ-score) were computed for each domain by subtracting the supervisor's score from the trainee's score. Therefore, positive and negative δ-scores, respectively, indicate trainees' overestimation and underestimation of their competence. The statistical analyses used in Step 1 (including the between- and within-subjects contrasts) were replicated in the second analysis, using the δ-scores as the dependent measure. Similar analyses have been used by previous researchers (e.g., Gonsalvez et al., 2015) and these analyses were repeated to facilitate comparisons across studies. Because only a limited number of a priori planned contrasts were computed and analyzed, Bonferroni-type corrections for α were not required (Tabachnick & Fidell, 2007).

3 | RESULTS

3.1 | Self-assessment by trainees

The results of the 3 Levels × 10 Domains ANOVA are presented in Table 1, and mean scores from both trainees and supervisors are presented in Figure 1.

3.1.1 | Competence across Levels

Competence scores systematically increased from Level 1 through to Level 3, $F(2, 138) = 60.81$, $p < .001$, $\eta\rho^2 = .47$. Level 2 trainees rated themselves higher than Level 1 trainees ($p < .001$), and Level 3 trainees rated themselves higher than at Level 2 trainees ($p < .001$). To test for a plateau effect between Levels 2 and 3, we also subjected the data to linear and quadratic contrasts. The linear contrast was significant ($p < .001$) but the quadratic contrast was not significant ($p = .68$), indicating that whilst there was a progressive increase in competence from Level 1 to Level 3, there was no evidence of a larger initial growth spurt from Level 1 to Level 2.

TABLE 1 Results of the 3 Levels × 10 Domains ANOVA for trainees' ratings of their competence.

Factor	Contrast	Sum of squares	df	Mean square	F	p	Partial η^2
			Between-subjects effects				
Level		38.80	2,138	19.40	60.81	<.001	0.47
			Within-subjects effects				
Domain	D1 vs. mean	2.46	1,138	2.46	60.29	<.001	0.30
	D2 vs. mean	0.51	1,138	0.51	14.69	<.001	0.10
	D3 vs. mean	0.13	1,138	0.13	4.06	.046	0.03
	D4 vs. mean	0.23	1,138	0.23	3.30	.071	0.02
	D5 vs. mean	0.18	1,138	0.18	8.79	.004	0.06
	D6 vs. mean	0.15	1,138	0.15	13.18	<.001	0.09
	D7 vs. mean	1.24	1,138	1.24	41.86	<.001	0.23
	D8 vs. mean	0.01	1,138	0.01	0.31	.580	0.00
	D9 vs. mean	0.39	1,138	0.39	4.69	.032	0.03
	D10 vs. mean	3.02	1,138	3.02	108.24	<.001	0.44
Domain × Level	D3 vs. mean	0.29	2, 137	0.14	4.50	.013	0.06
	D4 vs. mean	0.78	2, 137	0.39	5.65	.004	0.08
	D5 vs. mean	0.12	2, 137	0.05	2.58	.080	0.03

Note: Only statistically significant results are provided for Domain × Level interactions. Level = Developmental levels; D1 = Diagnosis and Clinical Assessment: Knowledge and Knowledge Application; D2 = Case Conceptualization and Interventions: Knowledge and Knowledge Application; D3 = Individual and Cultural Diversity, D4 = Effective and Skillful Assessment and Intervention; D5 = Relational and Communication; D6 = Reflective Practice; D7 = Scientist-Practitioner: Attitudes and Values; D8 = Ethical Practice; D9 = Professionalism: Organized and Disciplined Practice; D10 = Professionalism: Attitudes and Values.

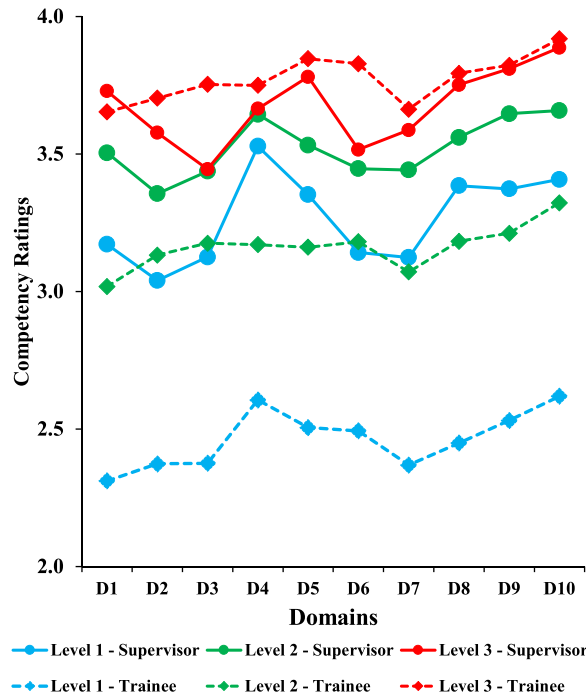Abbreviation: ANOVA, analysis of variance.

FIGURE 1 Assessment of trainee competence by trainees and their supervisors at three developmental levels for 10 domains. D1 = Diagnosis and Clinical Assessment: Knowledge and Knowledge Application; D2 = Case Conceptualization and Interventions: Knowledge and Knowledge Application; D3 = Individual and Cultural Diversity; D4 = Effective and Skillful Assessment and Intervention; D5 = Relational and Communication; D6 = Reflective Practice; D7 = Scientist-Practitioner: Attitudes and Values; D8 = Ethical Practice; D9 = Professionalism: Organized and Disciplined Practice; D10 = Professionalism: Attitudes and Values.

3.1.2 | Developmental trajectories across the 10 competency domains

Compared to the grand mean (see Table 1 and Figure 1), trainees rated themselves significantly lower in four domains: D1-Diagnosis and Clinical Assessment: Knowledge Competence; D2-Case Conceptualization and Intervention: Knowledge Competence; D3-Individual and Cultural Diversity; and D7-Scientist-Practitioner Competencies, and significantly higher on four domains: D5-Relational Communication, D6-Reflective Practice, D9-Professionalism, and D10-Professionalism: Attitudes and Values. The ratings on the other competencies: D4-Effective and Skillful Intervention and D8-Ethical Practice were comparable to the grand mean ($p > .05$ in each instance).

3.1.3 | Differential rates of competence development: Level × Domain Effects

We examined whether growth trajectories (across the four placements) for the 10 competencies were different from the trajectory of the overall mean (across domains). The results indicated that, in an overall sense, most competencies followed a comparable trajectory across developmental levels. The trajectories of two competencies, Individual and Cultural Diversity and Effective and Skillful Interventions, tracked differently. At Level 1, trainees rated themselves much less competent on Individual and Cultural Diversity ($M = 2.38$ vs. 2.46; $SE = 0.07$, $\eta\rho^2 = 0.31$), with this difference disappearing at Level 2 ($M = 3.17$ vs. 3.16) and at Level 3 ($M = 3.75$ vs. 3.77). Conversely, at Level 1, trainees rated Effective and Skilful Interventions as a strength ($M = 2.61$ vs. 2.46; $SE = 0.07$, $\eta\rho^2 = 0.31$), with this difference disappearing at Level 2 ($M = 3.17$ vs. 3.16) and at Level 3 ($M = 3.75$ vs. 3.77).

TABLE 2 Mean and standard deviation difference scores of trainees versus supervisors in developmental stages 1, 2, and 3.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	Grand mean
Level 1, mean (SD)	−0.86 (1.01)	−0.67 (1.05)	−0.75 (0.91)	−0.92 (0.80)	−0.85 (1.03)	−0.65 (1.01)	−0.75 (0.96)	−0.94 (1.09)	−0.84 (1.07)	−0.79 (1.02)	−0.80 (0.88)
Level 2, mean (SD)	−0.49 (1.10)	−0.22 (1.08)	−0.26 (1.07)	−0.47 (0.87)	−0.37 (1.13)	−0.27 (1.03)	−0.37 (0.94)	−0.38 (1.12)	−0.43 (1.08)	−0.34 (1.06)	−0.36 (0.93)
Level 3, mean (SD)	−0.08 (1.28)	0.13 (1.14)	0.31 (1.27)	0.09 (0.88)	0.07 (1.21)	0.31 (0.97)	0.08 (1.18)	0.04 (1.13)	0.01 (1.07)	0.03 (1.08)	0.10 (0.99)

Note: Dev = Developmental, D1 = Diagnosis and Clinical Assessment: Knowledge and Knowledge Application, D2 = Case Conceptualization and Interventions: Knowledge and Knowledge Application, D3 = Individual and Cultural Diversity, D4 = Effective and Skillful Assessment and Intervention, D5 = Relational and Communication, D6 = Reflective Practice, D7 = Scientist-Practitioner: Attitudes and Values, D8 = Ethical Practice, D9 = Professionalism: Organized and Disciplined Practice, D10 = Professionalism: Attitudes and Values. Negative scores indicate underestimation and positive scores overestimation of competence.

3.2 | Trainees' underestimation and overestimation of their competence (δ-scores)

Mean *d*-scores for the 3 Levels × 10 Domains are presented in Table 2 and the ANOVA results are presented in Table 3.

3.2.1 | Competence across levels

The 3 Levels × 10 Domains ANOVA yielded significant main effects for the Levels and Domain factors on δ-scores, but no Level × Domain interaction effects (see Table 2 for *F* values). In an overall sense, the results indicated that trainees systematically underrated their competence (grand mean = −0.32, SE = 0.10). However, T−S discrepancies varied across developmental levels, with negative scores (underestimation) being most pronounced early in training (at Level 1, $M = -0.80$; SE = 0.12), reducing at Level 2 ($M = -0.36$; SE = 0.15; $p < .05$) and drifting into positive δ-scores (overestimation) at Level 3 ($M = 0.10$; SE = 0.17; $p < .05$).

3.2.2 | Domain effects

To determine if the profile of δ-scores varied across the different competency domains, each of the 10 domains was compared to the grand mean. Relative to the overall trend, larger negative δ-scores (i.e., greater levels of underestimation) occurred for one competency, D1-Diagnosis and Clinical Assessment: Knowledge and Knowledge Application. Smaller δ-scores (deviations from the mean) were obtained for three competencies: D2-Case Conceptualization and Interventions; D3-Individual and Cultural Diversity; and D6-Reflective Practice.

4 | DISCUSSION

The current study provides new empirical data and valuable insights into an important but poorly researched topic, the degree of agreement between supervisor–trainee assessments of trainee competence. The examination of competence assessments for a range of competencies across developmental levels is also of value.

TABLE 3 Results of 3 Developmental Levels × 10 Domains ANOVA for trainee–supervisor difference (δ) scores.

Contrast	Sum of squares	df	Mean square	F	p	Partial η^2
			Between-subjects effects			
Dev Level	17.91	2, 138	8.95	9.81	.0001	0.13
			Within-subjects effects			
D1 vs. mean	1.83	1,138	1.83	10.72	.001	0.07
D2 vs. mean	1.28	1,138	1.28	8.06	.005	0.06
D3 vs. mean	1.84	1,138	1.84	9.36	.003	0.06
D4 vs. mean	0.87	1,138	0.87	2.99	.086	0.02
D5 vs. mean	0.11	1,138	0.11	0.84	.362	0.01
D6 vs. mean	3.03	1,138	3.03	34.28	.001	0.20
D7 vs. mean	0.003	1,138	0.00	0.02	.902	0.00
D8 vs. mean	0.62	1,138	0.60	3.28	.072	0.02
D9 vs. mean	0.57	1,138	0.57	2.76	.099	0.02
D10 vs. mean	0.01	1,138	0.01	0.08	.773	0.00

Note: Dev = Developmental; D1 = Diagnosis and Clinical Assessment: Knowledge and Knowledge Application; D2 = Case Conceptualization and Interventions: Knowledge and Knowledge Application; D3 = Individual and Cultural Diversity; D4 = Effective and Skillful Assessment and Intervention; D5 = Relational and Communication; D6 = Reflective Practice; D7 = Scientist-Practitioner: Attitudes and Values; D8 = Ethical Practice; D9 = Professionalism: Organized and Disciplined Practice; D10 = Professionalism-Attitudes and Values.

Abbreviation: ANOVA, analysis of variance.

4.1 | Self-assessment of competence across three levels of development

As may be expected, trainees assessed themselves as gaining competence as they advanced through their 2-year training courses in clinical psychology. Whilst there is little research on developmental trajectories, a previous study examining progress across four separate placements indicates an early spurt, an intermediate plateau, and a modest end-stage competence gain (Deane et al., 2018). Our results confirm the overall growth trajectory but found no evidence of a mid-stage plateau. The discrepancies between the two studies may be related to how developmental milestones are defined. In fact, we found a trend similar to that reported by Deane et al. (2018) when we tracked competence as a function of the number of placements. However, because not all placements are equal in terms of work experience (they may range from 200 to 400 hours and some students commence a new placement whilst the previous placement is ongoing), differentiating developmental milestones based on the number of placement hours completed at the time of end-placement assessments (regardless of whether these hours are accrued from one or more placements), as done in our study, may be a more accurate and reliable way to track competency development. Using our methods, we found no evidence of an early growth spurt between 300 and 600 hours of placement, followed by a developmental lag after this milestone.

4.2 | Trainees' evaluation of competency strengths and needs

The study also reveals how trainees perceive their relative strengths and needs. Compared to their average performance level, trainees assessed themselves as *stronger* in four domains (D5-Relational Competence;

D6-Reflective Practice; D9-Professionalism: Disciplined Practice; and D10-Professionalism: Attitudes and Values), and weaker in four others, including D1-Diagnosis and Clinical Assessment; D2-Case Conceptualization and Interventions; D3-Individual and Cultural Diversity; and D7-Scientist-Practitioner competencies. These results closely replicate the preliminary results of Hitzeman et al. (2020), in which domains falling within the assessment-and-intervention cluster were perceived as areas of relative weakness, whereas professionalism, reflective practice, and ethical practice were evaluated as relative strengths. Viewed from the foundational/functional classification, trainees gave themselves higher ratings on foundational competencies (all four domains listed). The set of competencies rated lower include both functional (D1 and D2) and foundational competencies (D3 and D7). When competencies are conceptualized in terms of knowledge, knowledge application, skills, and attitude values, trainees rate themselves lower on competencies where knowledge and knowledge application are central (D1, D2, D3, and D7), and higher on domains where relational skills, attitudes, and values (D5, D6, D9, and D10) are pivotal. It is also possible that trainees are more aware and/or willing to acknowledge inadequacies in knowledge and knowledge-application domains, whilst being less aware of and/or less willing to accept inadequacies in relationship skills and attitude-value domains.

These findings have important implications for training. For instance, it may be useful for supervisors to design methods and techniques that would facilitate a better understanding of effective practitioner attributes and skills, foster a deeper awareness of such biases, and formulate strategies to mitigate them when appropriate. Trainees' appraisal that they need greater improvement in cultural competence at Level 1 is reassuring because this area is an obvious need for growth and development. Their appraisal of their competence in assessment and intervention is interesting and of relevance. Whilst they readily acknowledge inadequacies in knowledge and knowledge-application early in training (Level 1), they perceive Effective and Skillful Assessment and Intervention as a relative strength. This perception is of some concern because effective practitioner skills may require substantive effort and systematic and prolonged training to acquire. Consequently, supervisors may need to address these naïve perceptions early in training.

4.3 | Trainee–supervisor differences (δ-scores) in competence assessments

Current empirical findings on trainee–expert discrepancies are inconsistent and hard to reconcile as these studies use heterogeneous samples, employ different instruments for trainees and experts, and report different results. Some studies indicate that practitioners overestimate their competence whilst others report the opposite effect. Our results provide new data that suggest that under- and overestimation may be influenced by developmental level, thereby offering a more nuanced explanation that reconciles these findings. Trainees in our study changed from underestimating their competence at Level 1 to marginally overestimating their competence at Level 3, a time point close to their graduation. If supervisors' assessments are taken to be more reliable and accurate measures of trainee competence, trainees show a pronounced underestimation of competence early in training ($M = -0.80$), with this bias persisting to a lesser degree at Level 2 ($M = -0.36$) and reversing to a slight overestimation at Level 3 ($M = 0.10$). Our results are largely consistent with outcomes from the study by Hitzeman et al. (2020), who also reported that their small sample of trainees (largely novices) underrated their competence. Our study extends these results by demonstrating that this bias reduces with training and is no longer evident by Level 3. The underestimation observed at Level 1 in our study is consistent with self-appraisal biases attributed to the impostor phenomenon found to be highly prevalent among psychology doctoral trainees (Tigranyan et al., 2021) and associated with features of anxiety, self-doubt, and self-criticism (Brosan et al., 2008; Kamen et al., 2010).

Alternative explanations for T–S discrepancies only partially explain our findings. For instance, a lack of familiarity with the rating instrument could make reliable self-assessments difficult for trainees at Level 1, with these discrepancies reduced at later stages after trainees receive "benchmark" scores from their supervisors at Level 1 (Al-Kadri et al., 2012; Boud et al., 2013). Whilst this explanation predicts improved accuracy later in training,

it does not predict the direction of our results (negative δ-scores at Level 1 and positive scores at Level 3). Inflated self-ratings of academic performance have sometimes been attributed to academic entitlement among millennial students who view themselves as customers engaged in an economic transaction with their degrees being a "product"' they purchase (e.g., Finney & Finney, 2010; Keener, 2020). However, a sense of entitlement would be expected to lead to inflated self-ratings across all levels of training, a pattern not observed in the present study. Because Level 3 ratings occurred near the point of completion of postgraduate training, it is possible that trainees' ratings were influenced by the need to meet social expectations and standards, especially at the final threshold to full registration. This may be especially pertinent to trainees who receive employment offers from agencies where they do their final placements.

In summary, it is unclear whether this inflation bias close to the final competency checkpoint is driven by a conscious strategy to influence assessors and potential employers, an attempt to correct an underestimation bias noted in earlier developmental levels, a nonconscious, implicit bias driven by cognitive dissonance or other mechanisms, or a combination of several factors. The magnitude of the inflation bias is admittedly small and, of itself bears little significance. However, it would be important to examine the trajectory of this inflationary trend as sustained or increasing inflation postlicensure would be of serious concern because it could lead to neglect of required professional development and adverse client consequences. Admittedly, large discrepancies early in training (negative δ-scores at Level 1) may be a consequence of multiple influences: trainees' underrating themselves and supervisors' tendency to leniency, a bias well-documented in the literature (e.g., Gonsalvez et al., 2021). It is of interest to determine if supervisors' concerns about client care and their gate-keeping responsibilities temper their default tendency to be lenient when they assess trainees on the threshold of gaining registration.

The diverging trajectories generated by supervisors and trainee assessments are informative, intriguing, and have useful applications for supervision and certification. Specifically, supervisors and trainees differed in their evaluations of the *pace* of competency development, with supervisors' assessments yielding a much flatter trajectory. In contrast, trainees commenced with underestimating their competence at the beginning of training and overestimating their competence at training completion. Our research also adds to the growing literature suggesting that self-assessments of competence early in training may be more variable, more amenable to relationship and developmental dynamics, and therefore less reliable as measures of trainee competence.

4.4 | Limitations and future directions

Although our study has several merits, the systematic charting of competency trajectories and the reliable and valid assessment of competence are challenging tasks and together comprise a substantive and ongoing endeavor. From that perspective, our results are no more than preliminary observations and interesting trends that warrant further investigation. For instance, the current study attempted to examine competency trajectories by a between-groups experimental design. A within-subjects, repeated-measures design that systematically monitors competence from the commencement of training to readiness to practice would be extremely useful. Although assessments were completed blind, supervisors and trainees were aware that their assessments would be communicated to each other, possibly influencing their ratings. It is possible that the pattern of supervisor-trainee differences observed in our study would change if these ratings were not divulged to each other. Future consideration should also be given to qualitative research to better understand competency assessment from the different supervisee and supervisor perspectives. There is also a need to examine whether and how supervisor competence, supervisory methods and techniques, placement variables, supervisor–trainee processes, and dynamics (e.g., demographic, culture, theoretical orientation) affect competency trajectories and assessment.

It is also worth noting that CΨPRS assessments capture supervisors' overall, subjective ratings that may be vulnerable to leniency and halo biases (e.g., Gonsalvez et al., 2021). As with other ratings, CΨPRS scores may be less able to reliably capture complex processes including meta-competencies (Kaslow et al., 2009). The reliability of such

assessments would be substantively improved by supplementing supervisors' judgments with objective assessments (e.g., objective, structured clinical examinations). Future research should also aim to build on our findings by examining factors that influence the trends reported in our results. For instance, are there specific factors within individuals that amplify or reduce over-estimation and under-estimation biases? An understanding of the psychological mechanisms that contribute to biased self-assessment is essential for creating and implementing effective mitigation strategies (Karpen, 2018).

5 | CONCLUSIONS AND PRACTICE IMPLICATIONS

This study contributes to the training and supervision of psychologists in several important ways. It offers much needed empirical data on self-assessment and supervisor assessment of competence. It highlights key differences between competence trajectories: supervisors' assessments indicate a flatter growth trajectory, whereas trainees' self-assessments suggest a steeper trajectory characterized by an underestimation of competence at the commencement of training, and an overestimation of competence by the end of training/at entry to independent practice. Trainees' evaluation of their competence, regardless of their accuracy, has important implications for supervision, training, and client outcomes (Creed et al., 2016; Mathieson et al., 2008). Consequently, our results highlight the importance of monitoring and reflecting on T–S discrepancies within supervision and a more careful scrutiny of these differences by training programs. It may also be important for educators and supervisors to ensure that their well-intentioned efforts to mitigate the effects of the impostor syndrome do not lead to trainee self-appraisals that are inflated. Moreover, additional training and supervisory resources are required in cultural competence and reflective practice either during or immediately postqualification to ensure safe and effective practice standards are met. Given the responsibility placed on supervisors as "gate-keepers" for the profession as they shape a trainee's practice, having a better understanding of the reliability of competency assessment and competency attainment throughout clinical training can only benefit trainee development and add to the integrity of the profession.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID
Craig J. Gonsalvez http://orcid.org/0000-0002-3294-6314
Tanya Riebel http://orcid.org/0000-0002-6424-0404
Laura J. Nolan http://orcid.org/0000-0001-5264-1238
Sonja Pohlman http://orcid.org/0000-0001-5719-246X
Warren Bartik http://orcid.org/0000-0001-8344-3306

PEER REVIEW
The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/jclp.23590.

REFERENCES

Al-Kadri, H., Al-Moamary, M. S., Al-Takroni, H., Roberts, C., & van der Vleuten, C. P. M. (2012). Self-assessment and students' study strategies in a community of clinical practice: A qualitative study. *Medical Education Online*, *17*(1), 11204. https://doi.org/10.3402/meo.v17i0.11204

APA. (2014). *APA guidelines for clinical supervision in health services psychology*. https://www.apa.org/about/policy/guidelines-supervision.pdf

Barrett, J., Gonsalvez, C. J., & Shires, A. (2020). Evidence-based practice within supervision during psychology practitioner training: A systematic review. *Clinical Psychologist*, *24*(1), 3–17. https://doi.org/10.1111/cp.12196

Bennett-Levy, J., Thwaites, R., Chaddock, A., & Davis, M. (2009). Reflective practice in cognitive behavioural therapy: The engine of lifelong learning. *Reflective Practice in Psychotherapy and Counselling*, *2007*, 115–135.

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, *38*(8), 941–956. https://doi.org/10.1080/02602938.2013.769198

Brosan, L., Reynolds, S., & Moore, R. G. (2008). Self-evaluation of cognitive therapy performance: Do therapists know how competent they are? *Behavioural and Cognitive Psychotherapy*, *36*(5), 581–587. https://doi.org/10.1017/S1352465808004438

Carter, T. J., & Dunning, D. (2008). Faulty self-assessment: Why evaluating one's own competence is an intrinsically difficult task: Faulty self-assessment. *Social and Personality Psychology Compass*, *2*(1), 346–360. https://doi.org/10.1111/j.1751-9004.2007.00031.x

Creed, T. A., Wolk, C. B., Feinberg, B., Evans, A. C., & Beck, A. T. (2016). Beyond the label: Relationship between community therapists' self-report of a cognitive behavioral therapy orientation and observed skills. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(1), 36–43. https://doi.org/10.1007/s10488-014-0618-5

Deane, F. P., Gonsalvez, C., Joyce, C., & Britt, E. (2018). Developmental trajectories of competency attainment amongst clinical psychology trainees across field placements. *Journal of Clinical Psychology*, *74*(9), 1641–1652. https://doi.org/10.1002/jclp.22619

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, *105*(1), 98–121. https://doi.org/10.1016/j.obhdp.2007.05.002

Falender, C. A., & Shafranske, E. P. (2017). Competency-based clinical supervision: Status, opportunities, tensions, and the future. *Australian Psychologist*, *52*(2), 86–93. https://doi.org/10.1111/ap.12265

Finney, G. T., & Finney, Z. R. (2010). Are students their universities' customers? An exploratory study. *Education + Training*, *52*(4), 276–291. https://doi.org/10.1108/00400911011050954

Fisher, P., Chew, K., & Leow, Y. J. (2015). Clinical psychologists' use of reflection and reflective practice within clinical work. *Reflective Practice*, *16*(6), 731–743. https://doi.org/10.1080/14623943.2015.1095724

Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B., Collins, F. L., & Crossman, R. E. (2009). Competency benchmarks: A model for understanding and measuring competence in professional psychology across training levels. *Training and Education in Professional Psychology*, *3*(4, Suppl.), S5–S26. https://doi.org/10.1037/a0015832

Gaubatz, M. D., & Vera, E. M. (2002). Do formalized gatekeeping procedures increase programs' follow-up with deficient trainees? *Counselor Education and Supervision*, *41*(4), 294–305. https://doi.org/10.1002/j.1556-6978.2002.tb01292.x

Gonsalvez, C. J., & Crowe, T. P. (2014). Evaluation of psychology practitioner competence in clinical supervision. *American Journal of Psychotherapy*, *68*(2), 177–193. https://doi.org/10.1176/appi.psychotherapy.2014.68.2.177

Gonsalvez, C. J., Deane, F. P., Blackman, R., Matthias, M., Knight, R., Nasstasia, Y., Shires, A., Nicholson Perry, K., Allan, C., & Bliokas, V. (2015). The hierarchical clustering of clinical psychology practicum competencies: A multisite study of supervisor ratings. *Clinical Psychology: Science and Practice*, *22*(4), 390–403. https://doi.org/10.1111/cpsp.12123

Gonsalvez, C. J., Deane, F. P., Terry, J., & Hoong Gooi, C. (2020). Anatomy of psychology practitioner competence: Clustering approach reveals the hierarchical organization of competencies. *Clinical Psychology: Science and Practice*, *27*, e12347. https://doi.org/10.1037/h0101779

Gonsalvez, C. J., Terry, J., Deane, F. P., Nasstasia, Y., Knight, R., & Gooi, C. H. (2021). End-of-placement failure rates among clinical psychology trainees: Exceptional training and outstanding trainees or poor gate-keeping? *Clinical Psychologist*, *25*(3), 294–305. https://doi.org/10.1080/13284207.2021.1927692

Hitzeman, C., Gonsalvez, C. J., Britt, E., & Moses, K. (2020). Clinical psychology trainees' self versus supervisor assessments of practitioner competencies. *Clinical Psychologist*, *24*(1), 18–29. https://doi.org/10.1111/cp.12183

Kamen, C., Veilleux, J. C., Bangen, K. J., VanderVeen, J. W., & Klonoff, E. A. (2010). Climbing the stairway to competency: Trainee perspectives on competency development. *Training and Education in Professional Psychology*, *4*(4), 227–234. https://doi.org/10.1037/a0021092

Karpen, S. C. (2018). The social psychology of biased self-assessment. *American Journal of Pharmaceutical Education*, *82*(5), Article 6299.

Kaslow, N. J., & Ammirati, R. J. (2020). Reflective practice: Finding synergy between psychological science and competency-based frameworks. *Clinical Psychology: Science and Practice*, *27*(4), 1–6. https://doi.org/10.1111/cpsp.12373

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. https://doi.org/10.1037/a0015833

Keener, A. (2020). An examination of psychological characteristics and their relationship to academic entitlement among millennial and nonmillennial college students. *Psychology in the Schools*, *57*, 572–582. https://doi.org/10.1002/pits.22338

Lilienfeld, S. O., & Basterfield, C. (2020). Reflective practice in clinical psychology: Reflections from basic psychological science. *Clinical Psychology: Science and Practice*, *27*(4), e12352. https://doi.org/10.1111/cpsp.12352

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*(1), 43–50. https://doi.org/10.1017/S1754470X08000081

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: Harnessing anxiety as motivation. *Journal of Personality and Social Psychology*, *51*(6), 1208–1217.

Parker, Z. J., & Waller, G. (2015). Factors related to psychotherapists' self-assessment when treating anxiety and other disorders. *Behaviour Research and Therapy*, *66*, 1–7.

Rodolfa, E., Greenberg, S., Hunsley, J., Smith-Zoeller, M., Cox, D., Sammons, M., Caro, C., & Spivak, H. (2013). A competency model for the practice of psychology. *Training and Education in Professional Psychology*, *7*(2), 71–83.

Stoltenberg, C. D., Bailey, K. C., Cruzan, C. B., Hart, J. T., & Ukuku, U. (2014). *The integrative developmental model of supervision*. Wiley. https://doi.org/10.1002/9781118846360.ch28

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon/Pearson Education.

Tice, D. M., Butler, J. L., Muraven, M. B., & Stillwell, A. M. (1995). When modesty prevails: Differential favorability of self-presentation to friends and strangers. *Journal of Personality and Social Psychology*, *69*(6), 1120–1138. https://doi.org/10.1037/0022-3514.69.6.1120

Tigranyan, S., Byington, D. R., Liupakorn, D., Hicks, A., Lombardi, S., Mathis, M., & Rodolfa, E. (2021). Factors related to the impostor phenomenon in psychology doctoral students. *Training and Education in Professional Psychology*, *15*, 298–305. https://doi.org/10.1037/tep0000321

Ward, M., Gruppen, L., & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, *7*(1), 63–80. https://doi.org/10.1023/A:1014585522084

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.