

---

## Assessing ourselves: Is the Assessment of Performance in Clinical Psychology Field Placements due to Biased Raters or Defective Rating Instruments?

John Bushnell, Craig J. Gonsalvez, Russell Blackman, Alice Shires, Christopher Allan, Kathryn Nicholson Perry, Roslyn Knight, Yasmina Nasstasia, and Frank Deane

---

### Abstract

In common with the training for many health professions, there are serious weaknesses in normal practices for assessing the performance of clinical psychology students in field placements, and these are evident in the field placement data captured from five NSW postgraduate clinical psychology courses. Iterative improvements in assessment forms, including the introduction of electronic data capture, and a criterion-referenced basis for decisions about competence within a developmental framework applied to agreed domains were found to do little to reduce the evident leniency and halo bias in supervisor ratings. Whilst not totally devoid of value (the field placement ratings do show credible improvement between mid-placement and end-placement), the supervisors' ratings of student performance at end of placement show uniform reluctance to describe a student as "unsatisfactory" or "needs development". The implications of this finding are discussed with particular reference to patient safety and professional responsibility.

### Introduction

Supervised experience in workplaces remains a cornerstone of clinical psychology education. However, widely used methods of assessment of student performance in the workplace are fraught with shortcomings, both in terms of which aspects of clinical performance should be assessed and how competence in these dimensions should be assessed (Gonsalvez et al., 2013; Lichtenberg et al., 2007).

Determining what to assess poses many challenges, including difficulties defining foundation and functional competencies and their component parts, lack of appropriate tools, especially for addressing the integration of knowledge, skills, and attitudes, and relating those tools to agreed minimal standards of competence across the spectrum of professional development from beginner to seasoned clinician (Rodolfa et al., 2005). At a broader level, lack of consensus about the value of a competency-based model has inhibited commitment to work towards resolving the challenges of

effective implementation (Lichtenberg et al., 2007). A decade ago it was argued that psychology had lagged behind other professions in defining what professional psychologists know and can do, and that a result of this failure was that psychologists largely failed to communicate the nature of their competence to the public and to policymakers (Kaslow, 2004).

Common practice in assessing the adequacy of student performance involves a supervisor making a judgement that is recorded using a rating scale describing several dimensions of behaviour. Relying on the credibility of field supervisors and their opportunities to directly observe students' consultations, Competency Evaluation Rating Forms (CERFs) are widely used as a low-cost, versatile measure of end-of-placement performance of clinical psychology students (Gonsalvez et al., 2013; Gonsalvez & Freestone, 2007). This type of assessment of competence has been subject to the criticism that it is prone to leniency bias (reluctance to award low scores) and

*John Bushnell is an Associate Professor at the Graduate School of Medicine, University of Wollongong; Craig J. Gonsalvez, Russell Blackman, Christopher Allan and Frank Deane are at the School of Psychology and Illawarra Institute for Mental Health, University of Wollongong; Alice Shires is from the School of Psychology, University of New South Wales, Kathryn Nicholson Perry, the School of Psychology, University of Western Sydney, Roslyn Knight, the School of Psychology, Macquarie University and Yasmina Nasstasia, the School of Psychology, University of Newcastle.*

halo bias (performance in one dimension influences judgment of performance in other dimensions) (Gonsalvez & Freestone, 2007). Although there have been several reports, across disciplines, suggesting that leniency and halo biases may affect these ratings (Gonsalvez et al., 2013; Vinton & Wilke, 2011) the data upon which these criticisms are based have been somewhat preliminary: for example, data from the Gonsalvez and Freestone study were based on one training programme (Gonsalvez & Freestone, 2007), and the recent Gonsalvez et al., study had a small number of participants (57). However, the adequacy of assessment practices is an issue with potential for adverse consequences: excessively lenient ratings of student performance may undermine the development of accurate self-perception in students, or reduce their motivation to learn or address deficits in performance, and ultimately erode the confidence of the public in the competence of the profession (Gonsalvez et al., 2013; Robiner et al., 1997).

Faced with these challenges, recognising the weakness of a pivotal component of student assessment and possible gains from a more systematic criterion-referenced approach to assessment (Wilkinson et al., 2007), five of the eight universities within the state of New South Wales that offered postgraduate training in clinical psychology participated in a collaborative process of review and iterative improvement of the Competency Evaluation Rating Forms (CERF), which became known as the Clinical Psychology Practicum Rating Scale (CΨPRS) (Gonsalvez et al., 2013). Domains were agreed, refined, trialled, and improved, and a developmental framework was applied to agreed domains using judgments related to explicit criteria. Improvements introduced in 2011 included refinement of the domains, and introduction of electronic data capture of supervisor judgments. This article describes the evidence of the impact on rater bias of this process of iterative improvements in a new, developmental, “standards-based” method of assessing

relevant domains of student performance in clinical placement settings.

## Methods

Ratings of student performance in clinical placement settings were collected retrospectively over a period of two years (2009 and 2010) by five participating universities ( $n = 187$  ratings) using a CERF-type rating scale known as the CΨPRS, and prospectively using electronic data capture in 2011 to obtain 111 matched mid-placement and end-of-placement ratings, as part of a larger study exploring alternative methods of assessment involving vignettes. The iterative development of the CΨPRS has been described previously (Gonsalvez et al., 2013). It was formulated from analysis of relevant CERFs already in use after extensive discussion over several sessions by the second author (C.G.) and representatives of the universities, who were usually clinic director or placement coordinator. The discussion was informed by developments in the international literature including the list of practicum competencies identified by Hatcher and Lassiter (2007).

The CΨPRS consisted of nine dimensions. Iterative changes designed to foster greater use of lower end of scale included use of a developmental model and developmental terms; and most significantly, the competency assessment shifted from a relative reference point (*peers at the same developmental level*) to a notionally absolute anchor (*readiness to practice*) from 2009 onwards, and clearer definition of stages (2011 version). CΨPRS domains assessed are Relational Skills, Clinical Assessment Skills, Case formulation and Intervention Skills (Combined), Ethical Practice, Professional Skills, Scientist-Practitioner Approach, Psychometric Skills, Personal capacities, and Use of Supervision.

The CΨPRS was used for placement evaluations of students in the five

participating universities in NSW, Australia. Supervisors completed the CΨPRS (60 items plus 9 overall items) at the end of each placement, separately for each student. Supervisors assigned an overall rating for each of the nine domains before rating each of the items within the domain. Data from 187 completed CΨPRS ratings were available for analysis from 2009-10, and form data set 1.

In 2011, using electronically-captured ratings, there were 204 completed CΨPRS forms, of which 111 also had mid-placement ratings, and this constitutes dataset 2. In data set 1, capture was by hardcopy, each university used different mid-placement assessment formats, and the ratings were categorical (“unsatisfactory”, “needs development”, “developing well”, “competent”). Dataset 2 was based on a more systematic online data capture for both mid-placement and end-of-placement, using the revised and updated scale, with

developmental stages defined, visual analogue scale used providing an interval rather than categorical scale, and supervisors were actively encouraged to use lower end of scale (see Appendix 1). This study, which was part of a larger project exploring vignette-based assessments as an alternative to CERF ratings, was approved by the ethics committees of each of the five partner universities.

## Results

There was extremely little use of any grades other than “Developing Well” or “Competent”. Less than 1% of 187 student performance ratings made during 2009-2010 fell into either the “unsatisfactory or “needs development” stage, a result that is barely credible, and suggests that some aspect of either the rating process or the behaviour of examiners may fail to identify those students whose performance is suboptimal.

**Table 1: Percentage of Student Performance Ratings Assigned to the Four Levels of Competence using CΨPRS (all items) for Data Set 1 (n=187)**

	Stage 1	Stage 2	Stage 3	Stage 4
	Unsatisfactory	Needs Development	Developing Well	Competent
Relational Skills	0.0	0.5	40.8	58.8
Clinical Assessment Skills	0.0	2.1	54.7	42.9
Formulation and Intervention Skills	0.0	3.0	55.7	41.0
Intervention Skills <sup>1</sup>	0.0	0.0	30.8	69.2
Psychometric Skills	0.0	0.5	55.5	43.6
Scientist Practitioner Approach	0.0	0.3	45.7	54.0
Ethical Practice	0.0	0.3	45.7	54.0
Professional Skills	0.7	1.2	46.4	52.5
Personal Capacities <sup>2</sup>	0.0	0.8	41.7	57.5
Supervision <sup>2</sup>	0.0	0.8	31.8	67.5
Progress During Placement	0.0	0.0	15.6	84.4

<sup>1</sup> Dimension assessed separately only in 2009 CΨPRS,  $n = 47$ .

<sup>2</sup> Dimension assessed only in 2010 CΨPRS,  $n = 140$ .

The data were explored further using the 2011 data for 111 students who provided both mid-placement and end-of-placement data, which demonstrate a statistically significant change in matched ratings of student performance between mid-placement and end-of-placement, with greater use of the “needs development” stage for students at mid-placement. Mean ratings across the 60 items were computed and mean differences

Bushnell JA, Gonsalvez CG, Blackman R, Deane F, Bliokas V, Nicholson-Perry K, Shires A, Nasstasia Y, Allan C, and Knight R, (2013) . Assessing ourselves: Is the assessment of performance in Clinical Psychology field placements due to biased raters or defective rating instruments? *Journal of the New Zealand College of Clinical Psychologists* 23(3), 4-11.

subjected to *t*-test. Pair-wise comparisons (mid- vs. end-placement) on all dimensions reveal differences significant at  $p < 0.001$ . However, by the end-of-placement ratings, less than 1% of students were found to be either “unsatisfactory” or “needs development”.

**Table 2: Percentage of Student Performance Ratings at Mid-Placement and End-of-Placement Within Four Stages of Competence for Data Set 2 (n = 111).**

Stage	Mid-Placement				End-of-Placement				<i>t</i> <sup>2</sup>	df	Sig. (2-tailed)
	1	2	3	4	1	2	3	4			
	Unsatisfactory	Needs Development	Developing Well	Competent	Unsatisfactory	Needs Development	Developing Well	Competent			
Relational Skills	0	4.5	34.2	52.3	0	0.9	12.6	84.7	-10.02	110	0.000
Clinical Assessment	0.9	6.3	40.5	37	0	0	18.9	81.1	-11.302	110	0.00
Case Formulation/ Intervention Skills	1	3.9	41.2	43.1	0	0.9	19.5	77.8	-9.915	100	0.00
Psychometrics	0	6.5	45.1	35.4	0	0	29	71	-6.613	55	0.000
Scientist-Practitioner	0.9	5.4	30.6	49.6	0	0	13.5	86.5	-8.451	110	0.000
Personal Capacities	0	3.6	27.9	61.3	0	0.9	10.8	86.5	-6.965	110	0.000
Ethical Practice	0	0.9	18	79.3	0	0	5.4	94.6	-5.856	110	0.000
Professional Skills	0	0.9	24.3	73	0	0	10.8	89.2	-5.824	110	0.000
Supervision	0	1.8	23.4	71.2	0	0	5.4	94.6	-7.774	110	0.000

100

Bushnell JA, Gonsalvez CG, Blackman R, Deane F, Bliokas V, Nicholson-Perry K, Shires A, Nasstasia Y, Allan C, and Knight R, (2013) . Assessing ourselves: Is the assessment of performance in Clinical Psychology field placements due to biased raters or defective rating instruments? *Journal of the New Zealand College of Clinical Psychologists* 23(3), 4-11.

<sup>1</sup>  $n = 111$  matched pairs mid–end placement, except for domain 4 (psychometrics) which was  $n = 55$ , and case formulation/intervention which had  $n = 100$  matching ratings, 2011 data.

<sup>2</sup>  $t$ -values are based on mean differences derived from overall scores.

## Discussion

Previous research has elucidated some of the reasons that clinical field supervisors are reluctant to give supervisees anything other than good news (Robiner et al., 1997). Common reasons included fear of damaging the supervisee's career, awareness of subjectivity inherent in assessment, experiencing difficulty providing negative feedback, fear of potentially diminished rapport, and personal identification with supervisee's problems. Although the students undertaking postgraduate training in clinical psychology are high academic achievers selected for good interpersonal skills, these data suggest that fewer than 1 in 100 occasions of supervisor ratings identified a level of performance that was either "unsatisfactory" or "needing development", a finding more likely to be attributable to bias and the inadequacy of CERF-type rating scales than truly reflecting levels of competency.

The CYPRS rating scale is not completely without value. Supervisors do show greater willingness to give lower grades at mid-placement, with approximately 1 student in 20 having one or more dimensions endorsed as either unsatisfactory or needing development. The dimensions on which relatively poorer student performance was most commonly found include Relational Skills, Clinical Assessment, Case Formulation /Intervention Skills, Scientist-Practitioner and the more technical dimension of Psychometrics. These dimensions include those that require higher order synthesis of knowledge and application of skills, which would be expected to be most challenging for students to learn and are different from more generic professional attributes and capacities embodied by other dimensions such as professional skills, ethical practice and personal capacities (see percentages in the "developing well" and "competent" columns in Table 2). However, the fact that less than 5% of students were identified as performing less than either "competent" or

"developing well" at the middle of their final year placements does bear close scrutiny.

In spite of our improvements to the CYPRS over time, using clearer definition of stages, using developmental terms, making the reference-point a criterion-referenced decision about "readiness to practice" and introducing electronic data capture, we found no discernible improvement in the distribution of ratings at the end-of-placement. This is a matter of considerable concern, as it suggests that any Likert-type scale, regardless of consistency or concordance with international opinion about what should be measured and how, may consistently expose students learning to become clinical psychologists to unrealistically positive feedback about their performance, leading them to believe that they are more competent than they are in fact. Such overconfidence may ultimately damage trainee psychologists, failing to direct them towards appropriate remedial learning. It may also damage the standing and public perception of the profession, and more importantly the health status of patients/clients who fail to receive an optimal standard of care (Gonsalvez et al., 2013; Robiner et al., 1997).

The underlying cognitive and affective processes that lead raters to avoid negative and critical feedback on student performance may go beyond the issues identified by Robiner et al. (1997), resulting from, or being influenced by, characteristics of the rating scale itself. Poor conceptual anchoring of the terminal and intermediate points of a Likert scale may make discrimination between levels of achievement difficult, hence encouraging use of only one or two columns. The visibility of all items on one page may encourage halo bias as each new item is compared with completed items. Alternative approaches to assessment using vignettes or "standardised narratives" have been the subject of recent research that shows considerable promise in reducing the bias of supervisor ratings of clinical

Bushnell JA, Gonsalvez CG, Blackman R, Deane F, Bliokas V, Nicholson-Perry K, Shires A, Nasstasia Y, Allan C, and Knight R, (2013) . Assessing ourselves: Is the assessment of performance in Clinical Psychology field placements due to biased raters or defective rating instruments? *Journal of the New Zealand College of Clinical Psychologists* 23(3), 4-11.

performance in both social work and clinical psychology (Gonsalvez et al., 2013; Regehr et al., 2007; Regehr et al., 2012). There is little evidence to suggest that continued reliance upon either the CΨPRS or other CERF-type ratings for assessment of student performance in clinical placements is justifiable, and strategies to implement multi-trait and multimodal assessments (Gonsalvez et al., 2013; Kaslow et al., 2007) should be pursued as a matter of urgency.

#### Acknowledgments

This project was supported by an Australian Learning & Teaching Council Grant PP-1624

#### References

- Gonsalvez, C., Bushnell, J. A., Blackman, R., Knight, R., Allan, C., Nasstasia, Y., Shires A., Nicholson Perry, K., Deane, F., & Bliokas, V. (2013). Assessment of psychology competencies in field placements: Standardized vignettes reduce rater bias. *Training and Education in Professional Psychology*, 7, 99–111.
- Gonsalvez, C., & Freestone J. (2007). Field supervisors' assessments of trainee performance: Are they reliable and valid? *Australian Psychologist*, 42, 23–32.
- 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.
- Kaslow, N. J. (2004). Competencies in professional psychology. *The American Psychologist*, 59, 774–781.
- Kaslow, N. J., & Rubin, N. J. (2007). Guiding principles and recommendations for the assessment of competence. *Professional Psychology: Research and Practice*, 38, 441–451.
- Lichtenberg, J., Portnoy, S., Bebeau, M., Leigh, I. W., Nelson, P. D., Rubin, N. J., et al. (2007). Challenges to the assessment of competence and competencies. *Professional Psychology: Research and Practice*, 38, 474–478.
- Regehr, G., Bogo, M., Regehr, C., & Power, R. (2007). Can we build a better mousetrap? Improving the measures of practice performance in the field. *Journal of Social Work Education*, 43, 327–343.
- Regehr, G., Ginsburg, S., Herold, J., Hatala, R., Eva, K., Oulanova, O. (2012). Using “standardized narratives” to explore new ways to represent faculty opinions of resident performance. *Academic Medicine: Journal of the Association of American Medical Colleges*, 87, 419–427.
- Robiner, W. N., Saltzman, S. R., Hoberman, H. M., Semrud-Clikeman, M., & Schirvar, J. A. (1997). Psychology supervisors' bias in evaluations and letters of recommendation. *The Clinical Supervisor*, 16, 49–72.
- Rodolfa, E. R., Bent, R., Eisman, E., Nelson, P., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice*, 36, 347–354.
- Vinton, L., & Wilke, D. J. (2011). Leniency bias in evaluating clinical social work student interns. *Clinical Social Work Journal* 39, 288–295.
- Wilkinson, T. J., Wells, J. E., & Bushnell, J. A. (2007). What is the educational impact of standards-based assessment in a medical degree? *Medical Education*, 41, 565–572.



**Appendix 1. An Example of the CΨPRS Framework and Sample Items**

<b>Clinical Psychology Practicum Competencies Rating Scale End Placement Review Form (CΨPRS-EP)</b>	
<b>Stages</b>	<b>Description of Stages</b>
<b>Stage 1. Beginner</b>	Knowledge and skills are at an early stage or yet to be developed. Inadequate knowledge and/or difficulty applying knowledge to practice. Several problems or inadequacies occur during sessions. There may be an absence of key features, inability to prioritise issues or to make appropriate judgements. Little awareness of process issues. On par with trainees commencing training without any practicum experience. Regular and intensive supervision required.
<b>Stage 2.</b>	Some basic competencies in assessment and intervention, manages narrow range of clients with low levels of severity, using structured therapeutic activities. Performance is variable; major problems may occur occasionally; regular supervision required.
<b>Stage 3.</b>	Moderate repertoire of basic competencies in both assessment and intervention leading to management of a wider range of clients. Demonstrates understanding of underlying principles and a moderate ability to generalise these to new cases/situations. Performance can be improved in minor ways; less frequent supervision required.
<b>Stage 4. Competent</b>	Large repertoire of basic to advanced competencies in both assessment and intervention, applied across range of clients and severity levels. Performance has reached competency levels on a par with a clinical psychologist working in their first job upon qualification.

<b>1. Relational skills</b> Includes ability for empathic understanding, application of basic counselling techniques, and collaborative goal formulation with clients.	<b>Overall Rating</b>			
	Stage 1	Stage 2	Stage 3	Stage 4
	●	●	●	●
a) Ability to form and communicate an empathic understanding to clients, carers, and significant others.	●	●	●	●
b) Ability to apply basic counselling techniques appropriately, including clarification, paraphrase, and summarisation responses.	●	●	●	●
c) Ability to use active and responsive listening skills.	●	●	●	●
d) Ability to formulate client goals in a collaborative manner.	●	●	●	●

**2. Clinical Assessment Skills.**

Includes ability to perform adequate assessments in a time efficient and in a personally/socio-culturally sensitive manner. Ability to demonstrate appropriate diagnostic skills, prioritise issues, and assess risk.

- a) Efficiency in conducting an adequate assessment.
- b) Ability to apply appropriate breadth of questioning to cover important issues including mental state examination.
- c) Ability to apply appropriate depth of questioning to ensure adequate understanding of key issues.
- d) Ability to use a hypothesis testing framework effectively.
- e) Ability and skill to make correct diagnoses and differential diagnoses.
- f) Ability to undertake assessments in a socio-culturally sensitive manner.