Rating Clinical Competencies in Clinical Placements: Can we enhance training outcomes?

This initiative is supported by an Australian Learning & Teaching Council (ALTC) Grant: PP10-1624

Reference group & other contributors

<table>
<thead>
<tr>
<th>Names</th>
<th>Affiliation/Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nancy Pachana</td>
<td>University of Queensland</td>
</tr>
<tr>
<td>Derek Milne</td>
<td>Curriculum (International)</td>
</tr>
<tr>
<td>Christine Brown</td>
<td>Education Expert</td>
</tr>
<tr>
<td>Vida Bitkas</td>
<td>Prim Clin Psychologist, Area Health</td>
</tr>
<tr>
<td>Frank Dianne</td>
<td>Expertise in Research and Evaluation</td>
</tr>
<tr>
<td>Nigel George</td>
<td>Clinic Director, Uni of Auckland</td>
</tr>
<tr>
<td>Grace Couchman</td>
<td>Monash Uni, Clinic Director</td>
</tr>
<tr>
<td>Moira Lewis</td>
<td>Nursing Expert</td>
</tr>
</tbody>
</table>

Objectives

- Provoke scrutiny and reflection about assessment of competencies in field placements
- Enhance awareness of biases that may affect supervisor ratings of students on placements
- Discuss issues and solutions
- Outline a multi-site, ALTC-funded initiative that seeks to address problems

For clinical psychology training

- Mid-term and end-placement evaluations of trainees by supervisors are mandated by accreditation bodies (e.g., APAC in Australia)
- What competencies, how they should be measured, when and by whom: Mostly unspecified in past, although more specifications in recent revisions
- Many important competencies relate to integration of theory with practice of clinical psychology => best evaluated at end-placements?
Summative evaluations: Inherent responsibilities

<table>
<thead>
<tr>
<th>Responsibility</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee</td>
<td>to provide effective and accurate feedback</td>
</tr>
<tr>
<td>Public</td>
<td>protect the public by maintaining standards of care</td>
</tr>
<tr>
<td>Agency-institution</td>
<td>to assess whether the institution’s goals and standards are met</td>
</tr>
<tr>
<td>Profession</td>
<td>Gate-keeping function</td>
</tr>
</tbody>
</table>

Summative ratings: practices in the field

- End-placement evaluation by supervisors vary widely among practicum and internship sites (Norcross, Stevenson, & Nash, 1986; Tyler et al., 2000).
- Written feedback (92%)
- Structured and unstructured (41%) evaluation
- Structured rating scales are becoming popular (Bernard & Goodyear, 2004).
- “nearly as many evaluation instruments as there are training programs,” (Bernard and Goodyear, p 161-162)

Summative ratings: implications

- Supervisors’ assessments are given serious consideration
  - Directors of training ranked it first, above 36 other quality assurance indices of professional training (Norcross et al., 1986).
  - In psychology, on the basis of a supervisor’s report, a trainee’s enrolment may be
    - Terminated
    - An entire placement (or part) may be repeated

Supervisor’s evaluations have face validity

- The importance and credibility accorded to supervisor judgments is understandable.
  - High academic qualifications (Norcross et al., 1986; Robiner, Saltzman, Hoberman, & Schryve, 1997)
  - Experienced clinicians (Norcross et al., 1986; Gonzalez et al., 2002)
  - Significant supervisory experience (Norcross et al., 1986)
  - Ecological validity:
    - able to observe trainee performance
    - across a wide variety of real-life situations
    - over an extended period of time

Factorial Structure Of Competencies?

- Proliferation of scales and items, with more and more competencies being added
  - Is it efficient or necessary to measure everything that looks or smells or sounds like a competency?
- Three published studies examining supervisor evaluations
  - Ralf Dohrenbusch & Lipka, 2006
  - Gonsalvez & Freestone, 2007
  - Fordham et al, 1988

II. Empirical Research

Supervisor Evaluations

**Inter-rater reliability**

- Evaluations by supervisor vs. independent raters
  - As with other aspects of research within the evaluation domain, there is a significant lacuna in this area
- Inter-rater agreement: 2 studies
  - Borders and Fong (1992)
  - Lazar and Mosek (1993)
- Predictive validity: 1 study
  - Gonsalvez & Freestone, 2007

Ratings of Cq skills: Supervisor vs. independent raters (Borders & Fong, 1992)

- Cq trainees (43) enrolled in a field-based practicum or internship
- Supervisors (n=27)
  - Counselor education faculty members (10), site supervisors (8) + others (adjunct faculty & doctoral students)
- Independent judges
  - Trained raters (advanced doctoral student + PhD counselor)

Ratings of Cq skills: Supervisor vs. independent raters (Borders & Fong, 1992)

- Counselor evaluation rating scale (CERS) was used
- External judges used the Vanderbilt Psychotherapy Process Scales (VPSS; O’Malley et al, 1983)
  - 3 subscales used: therapist warmth and friendliness
  - Negative therapist attitude
  - Therapist exploration
- Acceptable rater-agreement between trained raters: (r=.86)

Ratings of Cq skills: Supervisor vs. independent raters (Borders & Fong, 1992)

- Supervisors ratings and external judges (r=.12)
- External judges ratings increased with experience of student rated (first prac, second prac, internshp).
- Supervisor ratings of counseling performance showed a quadratic trend (V-shaped; similar ratings for first prac and internship students)

Ratings of Cq skills: Supervisor vs. independent raters (Borders & Fong, 1992)

- Several possibilities raised in discussion
- The results were, “consistent with the notion that supervisors’ evaluations may be influenced by their interactions with supervisees during weekly supervision sessions.”

Supervisor Ratings: Inter-rater reliability

- Lazar and Mosek (1993) attempted to determine which factors best predicted the grade social work trainees obtained for a field placement from their field instructors.
  - Hyp 1: Trainee’s abilities and past clinical skills would be best predictor
  - Hyp 2: How well the supervisor and trainee got along during their placement.
- Field supervisors (n=54) and trainees (n=70)
  - 38 supervisors supervised 1 student each and 16 supervisors supervised 2 students each

**Supervisor ratings: Inter-rater agreements**

- Independent variables:
  - Ability and skills:
    - Knowledge about the therapy and practice area: The student’s grade on a therapy and practice seminar was the index
  - Skills: The placement grade obtained by the trainee during the previous placement was included as an index of the trainee’s clinical skills
  - Relationship: The supervisor’s perception of the supervisor-trainee relationship: The Barrett-dyadic relationship unit was used to give measures of empathy, regard and congruence

A stepwise multiple regression analysis was conducted

- Relationship variables were the dominant predictors
  - The supervisor’s perception that there was an empathic understanding between the dyad accounted for 72% of the variance.
  - The two ability factors had a somewhat less significant influence
  - The authors conclude, “the present findings appear to indicate that the influence of the relationship on grading is such that the evaluation is invalid.”

**Is there a psychology practitioner aptitude g-factor?**

- There are indications that this is assumed.
- The interview is often meant to “evaluate” this
- We terminate students from programs inferring that they are unsuited/untrainable for clinical practice
- Strong correlations among dimensions rated (by same individual)
- If this is the case, this should be evident across placement ratings

There are indications that this is assumed.

- The interview is often meant to “evaluate” this
- We terminate students from programs inferring that they are unsuited/untrainable for clinical practice
- Strong correlations among dimensions rated (by same individual)
- If this is the case, this should be evident across placement ratings

**What’s the predictive value of supervisors’ ratings of trainees? (Gonsalvez & Freestone, 2007)**

- Data: 291 end-placement reports by supervisors over a 12 year period (1993-2004)
  - Data involved reports on a total of 131 clinical trainees by 160 supervisors
  - Numbers: 47, 38, 29, and 15 trainees completed 1 through to 4 placements, respectively.
  - Because we did not want to confound developmental levels, we separately examined correlations for CP1 and CP2, and for placements that occurred later (CP2 vs. CP3; CP3 vs. CP4). In sum, we examined the extent to which consecutive supervisors’ ratings on the same student matched.

**Between-placement correlations for the 11 Performance Dimensions from Data Set A**

<table>
<thead>
<tr>
<th>Item descriptions</th>
<th>CP1 vs CP2</th>
<th>CP2/CP3</th>
<th>CP2/CP4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) 1. Understanding agency work and functions</td>
<td>.10</td>
<td>.10</td>
<td>.10</td>
</tr>
<tr>
<td>2) 2. Professional conduct towards staff and clients</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>3) 3. Conducting psychological assessments</td>
<td>.23</td>
<td>.23</td>
<td>.23</td>
</tr>
<tr>
<td>4) 4. Knowledge of the aetiology and treatment approaches appropriate to the placement</td>
<td>.24</td>
<td>.24</td>
<td>.24</td>
</tr>
<tr>
<td>5) 5. Understanding important theoretical concepts in clinical psychology</td>
<td>.07</td>
<td>.07</td>
<td>.07</td>
</tr>
<tr>
<td>6) 6. Self awareness and insight into one’s behaviour</td>
<td>.23</td>
<td>.23</td>
<td>.23</td>
</tr>
<tr>
<td>7) 7. Clinical intervention: interviewing skills</td>
<td>.27</td>
<td>.27</td>
<td>.27</td>
</tr>
<tr>
<td>8) 8. Selection and application of interventions</td>
<td>.27</td>
<td>.27</td>
<td>.27</td>
</tr>
<tr>
<td>9) 9. Clinical intervention: skills in conducting intervention</td>
<td>-.27</td>
<td>-.27</td>
<td>-.27</td>
</tr>
<tr>
<td>10) 10 Case preparation and oral presentation</td>
<td>.35*</td>
<td>.35*</td>
<td>.35*</td>
</tr>
<tr>
<td>11) 11. Written case and psychological reports</td>
<td>.45*</td>
<td>.45*</td>
<td>.45*</td>
</tr>
</tbody>
</table>

**Paper presented at “Taking Clinical Psychology Postgraduate Training into the Next Decade” Conference, Brisbane, May 20-21.**
Inter-rater agreements:
Summary
- Research on inter-rater agreement between supervisors and independent judges is preliminary but disappointing. There is much less agreement than professional psychology assumes.
- Evidence to suggest that supervisor ratings of trainees have the expected predictive validity is weak.
- Research from other disciplines emphasise concern about field supervisor ratings
- More research is clearly needed
- Need to consider possibility that supervisor ratings are biased

III. Are Supervisor Ratings Biased?

Common types of rating biases
- Extensive psychological literature on rating biases
  - Especially in 1970’s and 1980’s
- Halo effect
- Leniency/Stringency bias
- Central tendency bias
- Contrast error

Are supervisor ratings biased?
Beliefs among supervisors
- Robiner et al., 1987
- Supervisors in APA-accredited clinical psychology
- Rated biases among supervisors in general
- Own bias
- Evaluations and in letters of reference

Rating bias by supervisors

<table>
<thead>
<tr>
<th>Supervisors’ belief whether other psychology supervisors are biased in rating interns</th>
<th>Yes</th>
<th>Not Sure</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>58%</td>
<td>31%</td>
<td>11%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage of supervisors who believed that their own ratings of interns were biased</th>
<th>Yes</th>
<th>Not Sure</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>58%</td>
<td>32%</td>
<td>10%</td>
<td></td>
</tr>
</tbody>
</table>

Types of biases by supervisors

<table>
<thead>
<tr>
<th>Types of biases</th>
<th>Others</th>
<th>Own</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leniency</td>
<td>39%</td>
<td>40%</td>
</tr>
<tr>
<td>Central Tendency</td>
<td>43%</td>
<td>45%</td>
</tr>
<tr>
<td>Strictness</td>
<td>16%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Robiner et al
What does the research say?

- Strong evidence that supervisors are lenient raters
  - Within psychology and across other professions
  - Gonzalvez & Freestone (2007) compared placement and academic coursework grades of a cohort of 58 students who had 123 placements between them:
    - About 80% of students received Distinction grades or above in placements
    - 60% of same group received Distinction grades or above in coursework.

More recently we examined placement ratings received by 58 students across 3 universities in NSW. Supervisors rated students on a 4-point scale (unsatisfactory, needs development, developing well, and competent).

Results of ratings

End placement ratings are from 1st placement (58%); 2nd placement (22%), 3rd placement (9%); Unspecified (11%).

Competence was defined as competence on par with a CP working in their first job upon qualification.

<table>
<thead>
<tr>
<th></th>
<th>Unsat (%)</th>
<th>Needs Dev (%)</th>
<th>Dev Well (%)</th>
<th>Competent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observed</td>
<td>0</td>
<td>0.63</td>
<td>47</td>
<td>92.3</td>
</tr>
</tbody>
</table>

Reasons for leniency bias

- Guilt or fear of damaging supervisee’s career (60%)
- Awareness of subjectivity inherent in evaluation (49%)
- Difficulty providing negative feedback (50%)
- Fear of potentially diminished rapport (48%)
- Personal identification with Se’s problems (32%)
- Legal and administrative issues (10%)
  - Being lenient may be an attractive option if process requires Sr assessments to demonstrate transparency, due-process and objectivity
  - Robiner et al, 1987

Other factors possibly contributing to supervisor biases

- It could be argued that several contextual factors collude to prime and accentuate rating biases in prac settings
- Biases that stem from the close relationship between supervisor and supervisee
  - Parallel processes that may affect ratings
  - Supervisors are trained to be “facilitative, supportive, and constructive” in their feedback to supervisees BUT are required to switch to an assessor’s role at end-placement

Leniency Bias: Implications

- “It may not be an exaggeration to consider the existence and extent of supervisory bias to be the most critical quality assurance issue confronting clinical psychology…” (Robiner et al, 1987, p 62)
- Inaccurate ratings may be a disservice to the training program, profession, public.
- Effect on Se: May foster inflated self perceptions? Prevents appropriate self-assessment and remediation strategies?
Are Supervisors' ratings influenced by the halo effect?

- Leniency bias may stem from halo effects
- Halo-type bias is consistent with our data (Gonsalvez and Freestone, 2007)
- the pattern of strong, relationships observed among the 11 performance dimensions when rated by the same supervisor, relatively weak between-supervisor agreement is the pattern you would predict for the halo-effect.
- Dohrenbusch & Lipka (2006) also raises this possibility as being an influence contributing to factor 1.

IV. The Vignette-Project: An Innovative Method To Assess Practicum Competencies

Three Main Objectives

- Improve the current rating scale used
- Design and standardise vignettes
- Compare outcomes from the two assessment methods

Problems With Previous Scales

- Adoption of a “relative anchor” for scaling. Supervisors are required to rate a student relative to performance of peers at the same stage of training
  - Inconsistent with theoretical framework and definition of competency
  - Field supervisors don’t have a “normative” anchor (esp new supervisors).
  - Precludes efforts to benchmark training
  - Difficult to track an individual’s developmental progress
  - Difficult to establish developmental trajectories among competencies

Progress?

- Adopt a developmental model of competency training
- Discrimination between stage (milestones) and pace of progress (horse-power)
- Replaced the relative anchor with an ‘absolute’ definition of competence: “clinical psychologist working in their first job upon qualification”
  - Field supervisors would better understand this anchor
- Defined the extremes of a continuum and left the intermediate stages “untagged”
- The evolution of the Clinical Psychology Practicum Competencies Rating Scale (CYPRS)

The CYPRS

- Section A: Nine competency domains incorporating 60 items
- Section B: Self-evaluation by trainees
- Section C: Qualitative comments by supervisor
- Section D: Overall rate of progress
- Section E: Supervisor’s overall evaluation (e.g., pass, fail, further assessment)
The CYPRS Rating Competencies (Sec A)

- Nine competency domains
- 3-10 items per domain
- 60 items across 9 domains
  - Guided by international literature
  - Consensus among 6 NSW universities
- Identification and definition of 4 stages

Competency Domains

1. Relational skills
2. Clinical assessment skills
3. Formulation and Intervention skills
4. Psychometric skills
5. Scientist practitioner approach
6. Personal capacities
7. Ethical practice
8. Professional skills
9. Response to supervision

Description of Stages

Stage 1 (Beginner)

- Knowledge and skills are at an early stage or yet to be developed. Inadequate knowledge and 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.

Stage 4 (Competent)

- 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.

Rating Competencies

- Ratings are made in reference to a notional absolute standard of competent professional practice (Stage 4).
- Supervisors are instructed it is anticipated that ratings across placements should reflect progression towards competency and that performance levels during earlier placements are likely to match Stages 1 and 2 and, as training progresses, move towards Stages 3 and 4.
- Srs are encouraged to mention concerns and to mention uncertainties if applicable
- Pace of progress is rated on a separate scale

Rating competencies
The visual-analogue scale
To record your rating, move the slider to the point that represents the trainee’s current level of performance

Section D: Progress levels

<table>
<thead>
<tr>
<th>Progress Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsatisfactory Progress</td>
<td>Progress is considerably slower than the pace expected at this stage of training. Consequentially, little or no change has been observed in the trainee’s capabilities. Major deficits in one or more areas that are of serious concern.</td>
</tr>
<tr>
<td>Slow progress</td>
<td>Some progress has been made, but progress has been uneven and inconsistent following above-average investments of staff resources. Rate of progress is below the standard expected at this stage of training.</td>
</tr>
<tr>
<td>Inconsistent progress</td>
<td>Progress has been inconsistent or patchy across time and/or domains, with satisfactory progress achieved some of the time in some domains but not all the time across all domains.</td>
</tr>
<tr>
<td>Developing Well</td>
<td>Consistent and good progress has been achieved. The rate of progress matches expectations for trainees at this stage of training.</td>
</tr>
<tr>
<td>Excellent progress</td>
<td>The trainee has made accelerated progress during the placement, much above the rate expected at this stage of training.</td>
</tr>
</tbody>
</table>

SECTION E
SUPervisor’s overall evaluation

Unsatisfactory
- Serious concerns about intern’s competencies and rate of progress. Among other possibilities, further actions could be recommendations for remedial action that includes repetition of part or full placements.

Uncertain or partially satisfactory
- Some concerns about intern’s competencies or variable/inconsistent performance BEHAVIOUR by intern.
- Recommendations could include brief and specific remedial assistance for intern, or further assessment to be organised by the training program.

Satisfactory (Pass)
- Intern has demonstrated competencies at or exceeding expected standards at this stage of training.

The vignette-matching procedure
- Pioneered by Marion Bogo and her colleagues (social work)
- Designed a catalogue of 20 vignettes representing competency profiles
- Supervisors were asked to read all 20 and to pick out the vignette that best matched their trainee
- Preliminary evidence indicated better results with the vignette procedure
- Less prone to rating biases

Standardization of Vignettes
- More difficult than anticipated
- Vign for 9 domains x 4 levels
- V1: 6 Clinic Directors drafted vignettes; at least 2 sets of vignettes for each domain
- V2: Final draft of vignettes by subcommittee
- Blinded peer review …discussion within a subcommittee

**Standardization of Vignettes**
- Calibration exercise, revised
- V3: Pilot tested by field supervisors (n=22)
- Vignette calibration (7 experts)
  - Good results for 50% of vignettes
  - Modest results for 50% of vignettes
- Main study (just commenced)

**Counselling Skills**
(Completion Score = 47)
- Trainee O relates to clients effectively in most simple client situations but experiences difficulties in more complex cases. She/he experiences difficulty in maintaining a comfortable, warm, respectful and confident demeanour due to a focus on self-performance or other factors. She/he demonstrates genuine reflective listening skills and makes appropriate emotional and meaningful responses in some cases. However, she/he may sometimes reinforce poor coping strategies by confusing empathy with sympathy. She/he may have difficulties in appropriately directing and guiding client focus.

**CBT Skills**
(Completion Score = 8)
- Trainer XB demonstrates a good knowledge of rationales for and good skills to conduct a fairly large range of CBT techniques. She/he efficiently identifies unhelpful cognitions and beliefs, and poses relevant socratic questions after appropriate preparation. A collaborative style ensures that modest gains within and across sessions are typically achieved. When this does not occur it is because of client resistance or because an implicit or more subtle belief was not targeted for change. Despite demonstrating good CBT skills, aspects that could improve include fluency, timing, and improved consolidation of high impact moments during sessions.

**Response to Supervision**
(Completion Score = 9.8)
- Trainee A has a mature, open, and positive attitude towards supervision, perceiving it as an opportunity to acquire new ideas, to consolidate learning, to discuss one’s approach to clients, and one’s positive and negative feelings and reactions to the placement. She/he has a high level of motivation and prepares well for supervision and other practicum activities. The trainee is reflective and self-aware, and has a relatively accurate appraisal of one’s capabilities. Supervisory sessions are pleasant, collaborative, professional, and effective. Overall, the trainee has made accelerated progress during the placement, much above the rate of progress expected of peers at a similar stage of training.

**Plan Ahead**
- N=200 placement assessments using both CYPRS and vignettes across 6 universities
- N=300 placements assessed using CYPRS across 6 universities
- Mid- and end-placement data for large number of students
- Within subject data across several placements
- Between supervisor data for same placements

**As Resource**
- No commercialisation or financial incentives
- Free resource to training programs within Australia with research participation
- Flexibility for individual programs
- Core items + option for additional items
- Web-based administration and scoring
- Automated reports
- Secure data collection and storage
- Privacy of institutions protected

Where to from here?

- Need for Research Collaboration across training programmes
- Interested in CYPRS hardcopy: craigc@uow.edu.au
- Research collaboration options
  - Ethics approval is required, but should be easy to obtain
  - Research involvement: invitation to supervisors and trainees to volunteer use of de-identified data

Key References