A multi-site study on the assessment of clinical psychology competencies by field supervisors: Should vignettes replace rating scales?

This initiative is supported by an Australian Learning & Teaching Council Grant.

Research & Clinical Investigators

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Gonsalvez</td>
<td>University of Wollongong</td>
</tr>
<tr>
<td>Roslyn Knight</td>
<td>Macquarie University</td>
</tr>
<tr>
<td>Kathryn Nicholson Perry</td>
<td>University of Western Sydney</td>
</tr>
<tr>
<td>Alice Shaw</td>
<td>Clinic Director, University of NSW</td>
</tr>
<tr>
<td>Christopher Allan</td>
<td>University of Wollongong</td>
</tr>
<tr>
<td>Russell Blackman</td>
<td>Project Officer</td>
</tr>
<tr>
<td>John Busnell</td>
<td>University of Wollongong</td>
</tr>
<tr>
<td>Judy Hyde</td>
<td>Sydney University</td>
</tr>
<tr>
<td>Yasmina Nasstasia</td>
<td>University of Newcastle</td>
</tr>
</tbody>
</table>

Competencies in Psychology

Recent Developments

- Several notable developments on the competency front
  Highlight: Documents by the APA commissioned Competency Benchmarks Working Group
- Several documents released recently (Fouad et al., 2009; Kaslow et al., 2009)

Competency Benchmarks

Fouad et al., 2009

- Assessment of Competency Benchmarks Work Group, APA Board of Educational Affairs project from 2005
- 15 core competency areas identified
- At 3 developmental levels of education and training
  - Readiness for practicum
  - Readiness for internship
  - Readiness for entry to practice
- Identifies essential components and behavioural indicators for each of these levels

Supervisor’s Evaluations: Face Validity

Credibility accorded to supervisor judgments is understandable.

- High academic qualifications (Norcross et al., 1986; Rubiner et al., 1997)
- Experienced clinicians with significant supervisory experience (Norcross et al., 1986; Gonsalvez et al., 2002)
- Ecological validity: able to observe trainee performance, across real-life situations over an extended period of time
- Directors of training ranked it first, above 36 other quality assurance indices of professional training (Norcross et al., 1986).
Despite advances, there are major problems with competency assessments that need to be addressed:

- What are the principal competency domains?
- Can they be assessed reliably in the field by supervisors and within programs by trained faculty?
- Do they have any predictive value?
- Do supervisor and clinical assessments suffer from systematic rating biases?

**Factor Structure?**

- Limited research in psychology
- Proliferation of domains and items
- A few studies have examined specific aspects of clinical work
  - Gonsalvez & Freestone (2007): 11 broad domains → 2 clusters
- Recent report (Fouad et al., 2009) is the outcome of expert consensus, but no corroborative PCA analyses as yet

**Clinical Skills Assessment Rating Form (CSA-RF, Tweed et al, 2010)**

- Five domains identified from assessment interview:
  - Demonstrating professional therapeutic engagement (13 items)
  - Creating a secure base (7 items)
  - Formulation (5)
  - Facilitating mutual understanding (4)
  - Session structure (3)

**Assessments: Agreement among Raters**

- Limited research
  - Field supervisors vs. independent trained raters (Borders & Fong, 1992)
  - Between-supervisor ratings of trainee videotapes (Tweed et al., 2010)
- Rater reliability for CTRS-Revised: Mixed results
  - Acceptable rater reliability for trained raters (Blackburn, et al., 2001), but negative results (Keen & Freeston, 2008)

**Between Supervisors: Reliability (Tweed et al., 2010)**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demonstrating professional therapeutic engagement</td>
<td>.12</td>
</tr>
<tr>
<td>Creating a secure base</td>
<td>.09</td>
</tr>
<tr>
<td>Formulation</td>
<td>.12</td>
</tr>
<tr>
<td>Facilitating mutual understanding</td>
<td>.07</td>
</tr>
<tr>
<td>Session structure</td>
<td>.04</td>
</tr>
</tbody>
</table>

**Predictive Value**

- To what extent does a supervisor’s assessment predict trainee performance in a subsequent placement?
- Predictive value:
  - Gonsalvez & Freestone (2007)
  - Two studies from social work, both yielded poor results (Bogo et al., Lazar & Mosok,)
**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 to 4 placements, respectively.
- Supervisors’ overall ratings on 11 broad domains were used
- Only examined comparisons between consecutive placements. Placement 1 vs. P2; P2 vs P3, etc.,

---

<table>
<thead>
<tr>
<th>Competency Domains (11)</th>
<th>CP1 vs CP2</th>
<th>CP2-CP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Understanding agency work and function</td>
<td>.10</td>
<td>.16</td>
</tr>
<tr>
<td>2. Professional conduct towards staff and clients</td>
<td>.03</td>
<td>.40*</td>
</tr>
<tr>
<td>3. Conducting psychological assessments</td>
<td>.23</td>
<td>.14</td>
</tr>
<tr>
<td>4. Knowledge of the aetiology and treatment approaches appropriate to the placement</td>
<td>.24</td>
<td>.10</td>
</tr>
<tr>
<td>5. Understanding important theoretical concepts in clinical psychology</td>
<td>.06</td>
<td>.26</td>
</tr>
<tr>
<td>6. Self awareness and insight into one’s behaviour</td>
<td>.23</td>
<td>.50**</td>
</tr>
<tr>
<td>7. Clinical intervention: interviewing skills</td>
<td>.27</td>
<td>.08</td>
</tr>
<tr>
<td>8. Selection and application of interventions</td>
<td>.27</td>
<td>.06</td>
</tr>
<tr>
<td>9. Clinical intervention: skills in conducting intervention</td>
<td>-.27</td>
<td>-.08</td>
</tr>
<tr>
<td>10. Case preparation and oral presentation</td>
<td>.35*</td>
<td>-.09</td>
</tr>
<tr>
<td>11. Written case and psychological reports</td>
<td>.45*</td>
<td>.61**</td>
</tr>
</tbody>
</table>

---

**Supervisor Assessments**

**Predictive Value**
- Two studies from social work yielded disappointing results
- Which factors best predicted the placement grade of social work trainees? (Lazar and Mosek, 1993)
  - Field supervisors (n=54) and trainees (n=70)
  - Relationship: How well the supervisor and trainee got along during their placement was the best predictor, not prac grades given by previous supervisor or academic ability.

---

**PCA of Social Work Competencies** (Bogo et al, 2002)
- Examined competencies (80 items) rated in Year 1 and Year 2 completed by social work students.
- Correlations between Year 1 and Year 2 supervisors was .12, which was statistically significant but unacceptably low

---

**Are Supervisors’ Ratings Biased?**
- Broad consensus from most researchers that their data are consistent with systematic rater biases.
- The two biases that stand out are leniency and halo.
- Leniency bias (Bogo et al., 2002, 2004; Borders & Fong, Gonsalvez & Freestone, 2007; Lazar & Mosek, 1993)
  - Placement grades: Above 80% of students received D or above
  - Coursework: 60% received Distinction grades or above
  - Changed the scaling to a 6-point scale; no changes

---

**Results of Supervisor Ratings**

End placement ratings from 58 placement across 3 universities in NSW

<table>
<thead>
<tr>
<th></th>
<th>Unsatis (%)</th>
<th>Needs Dev (%)</th>
<th>Dev Well (%)</th>
<th>Competent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>0</td>
<td>0.63</td>
<td>47</td>
<td>52.3</td>
</tr>
</tbody>
</table>
### Are Supervisor Ratings Biased? Beliefs Among Supervisors

- Study by Robiner et al., 1987
- Sample: supervisors in APA-accredited clinical psychology in USA
- Examined
  - Acknowledgement of bias in own ratings
  - Beliefs that other supervisor ratings were biased
  - Beliefs about bias in letters of reference

<table>
<thead>
<tr>
<th>Item</th>
<th>Yes (%)</th>
<th>Not Sure (%)</th>
<th>No (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief that other supervisors are biased in rating interns</td>
<td>58%</td>
<td>31%</td>
<td>11%</td>
</tr>
<tr>
<td>Supervisors who believed that their own ratings of interns were biased</td>
<td>58%</td>
<td>32%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Robiner et al

### Leniency Bias


### Are Supervisors’ Ratings Influenced by the Halo Effect?

- Several researchers indicate the possibility of halo bias (e.g., Bogo et al., 2002; Borders & Fong, 1995; Dohrenbusch & Lipka, 2006)
- Halo-type bias is consistent with our data (Gonsalvez and Freestone, 2007)
  - Strong, relationships observed among the 11 domains when rated by the same supervisor, AND
  - Relatively weak between-supervisor agreement

### Rating 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: Leniency may foster inflated self perceptions? Prevents appropriate self-assessment and remediation strategies?

### Summary

- Converging evidence from several studies within psychology and across disciplines
- Case of practicum assessment

WE HAVE A PROBLEM
Current Project: Three Main Objectives

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

Key assumption: The problem is the measurement instrument not the raters.

Problems With Previous Scales

- Supervisors rate students relative to performance of peers, but have no normative reference point.
- Use of relative “anchor” has other disadvantages:
  - Allows for an individual to be rated as adequate even if no progress is achieved (e.g., slow progress of group)
  - Precludes efforts to benchmark training
  - Difficult to track an individual’s developmental progress
  - Difficult to examine differential trajectories among competencies

The CPYPRS

- Adopted a developmental model of competency training
- Discarded the “relative criterion” as anchor for scaling
- Adopted the “ready-to-practice” benchmark as a reference point. Specifically, competence was defined as “abilities and skills demonstrated by a clinical psychologist working in their first job upon qualification”
  - Field supervisors would better understand this anchor

- Adopted a 4-stage model from Beginner (Stage 1) to Competent (Stage 4)
  - Stages 2 and 3: “not labelled”
  - Supervisors observed to use stage achieved and rate of progress to help determine their ratings, so we differentiated these factors:
    - Stage achieved (milestone reached)
    - Pace of progress (horse-power)

The CPYPRS: Section 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 of development
- Use of a visual analogue scale for ratings

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/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.

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
The visual-analogue scale

To record your rating, move the slider to the point that represents the trainee’s current level of performance

Stage 1 Beginner Stage 2 Stage 3 Stage 4 Competent

The CYPRS: D1 Relational Skills

Will Changes to the CYPRS Fix the Rating Problem?
- Answer: We don’t know.
- Word of caution from Regehr et al, 2007. Can we build a better mouse-trap?
  "Based on a several-year program of research...finally believe the growing body of evidence which would seem to indicate that trying to improve field evaluation scales may be the academic equivalent of rearranging the deck chairs on the Titanic. We now believe that the problem is not with the particular manifestations of the scales, but with the existence of the scales themselves." (Regehr et al., p 338)

The Vignette-Matching Procedure
The Vignette-Matching Procedure

- Pioneered by team from the Uni of Toronto Team (Regehr, Bogo and others)
- Designed a catalogue of 20 vignettes representing competency profiles
- Supervisors were asked to read all 20 and to pick out the vignettes that best matched their trainee
- Then narrowed to one or two that best fit trainee
- Preliminary evidence indicated better results with the vignette procedure; better distribution across performance levels
  - Reduced leniency bias

Current Study

- Used 9 domains X 4 developmental stages from CYPRS. Allows independent scores on each domain
- Adopted a more complex model that better simulates better identification of persons of forensic interest
- Disadvantage: larger number of vignettes needed to be crafted and standardised
- Derived a calibration score for vignettes
- More rigorous process involved in designing and standardisation

Standardisation of Vignettes

- More difficult than anticipated
- Vignettes 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 by experts (n=7; n=12)
    - Good results for 50% of vignettes
    - Modest results for 50% of vignettes
- Main study (just commenced)

Pilot Data from Vignette Project

<table>
<thead>
<tr>
<th>Domains</th>
<th>N</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>20</td>
<td>5%</td>
<td>20%</td>
<td>55%</td>
<td>20%</td>
</tr>
<tr>
<td>D2</td>
<td>20</td>
<td>5%</td>
<td>15%</td>
<td>60%</td>
<td>20%</td>
</tr>
<tr>
<td>D3</td>
<td>19</td>
<td>0</td>
<td>32%</td>
<td>58%</td>
<td>10%</td>
</tr>
<tr>
<td>D4</td>
<td>20</td>
<td>0</td>
<td>35%</td>
<td>45%</td>
<td>20%</td>
</tr>
<tr>
<td>D5</td>
<td>20</td>
<td>0</td>
<td>15%</td>
<td>25%</td>
<td>55%</td>
</tr>
</tbody>
</table>

Targets: Next 12 months

- 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 trainees
  - Factor structure
  - Examine effectiveness of briefer versions
Goals: Next 12 months

- Compare results from the CPPRS and the Vignette-Matching Procedure
- Within-trainee data across several placements: Chart developmental trajectory over time for group
  - Compare domains and competencies
- Between supervisor data for some placements:
  - Inter-rater agreement for CPPRS and the V-Procedure
- Plan to continue research next year