Beyond Rater Cognition
The Impact of Supervisor Continuity on the Quality of Work-based Assessments

Warren J. Cheung MD MMEd FRCPC
Nancy Dudek MD MEd FRCPC
Timothy J. Wood PhD
Jason R. Frank MD MA(Ed) FRCPC

September 30, 2016
I do not have an affiliation (financial or otherwise) with a pharmaceutical, medical device or communications organization.

Je n’ai aucune affiliation (financière ou autre) avec une entreprise pharmaceutique, un fabricant d’appareils médicaux ou un cabinet de communication.
Background

- Work-based assessments – an increasingly important means of reporting expert judgments of trainee performance

- Concerns in the literature about how these forms are being completed by supervisors
  - Do not always reflect the assessor’s actual judgments
  - Do not always contain meaningful assessments
Background

- Interactions between supervisors and trainees are often:
  - Limited, episodic, fragmented

- Impact on the quality of assessments:
  - Insufficient evidence to make judgments
  - Reluctance to provide discriminating assessments
  - Mediocre ratings and comments
Supervisor Continuity

• Supervisors and trainees report:
  ➢ Greater engagement
  ➢ More frequent observations
  ➢ Assessments are more fair and accurate
  ➢ Meaningful feedback given more frequently

• Objective:
  ➢ To determine the effect of increased supervisor-trainee continuity on the quality of completion of WBAs
Methods

• **Setting:** Department of Emergency Medicine, University of Ottawa

• **WBA form =** Daily Encounter Card (DEC)

• **Measure of DEC quality:**
  - Completed Clinical Evaluation Report Rating (CCERR)
    - 9 item instrument
    - Ability to discriminate DECs quality and produce reliable scores†

Methods

- Three naturally occurring supervisor-resident encounters:
Methods

• DECs retrieved from our electronic database (2013-2014)
• 20 staff involved in CTTs

CTT
All DECs (n=183)

Non-CTT
5 DECs/supervisor (n=100)

Off-Service
5 DECs/supervisor (n=100)
Methods

• 2 blinded raters scored the quality of each DEC using the CCERR

• Means between the 3 groups were compared using a univariate ANOVA

• Conducted a repeated measures ANOVA to determine if there was a difference in DEC quality with increasing interactions
Results
# Results

<table>
<thead>
<tr>
<th></th>
<th>CTT vs. Non-CTT</th>
<th>CTT vs. Off-service</th>
<th>Non-CTT vs. Off-service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean CCERR Score</td>
<td>21.0</td>
<td>21.0</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td>21.9</td>
<td>20.7</td>
<td>21.9</td>
</tr>
<tr>
<td>* statistically significant (p&lt;0.05)</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Discussion

• Supervisors are not completing DECs to an adequate standard
  > Consistent with existing literature

  > Studies of ITER quality demonstrate similar CCERR scores

  > Reinforces the need to determine better ways of documenting expert assessments of trainees
Discussion

• Greater supervisor continuity did not result in higher quality DECs

• DEC quality did not improve with increasing number of interactions
  › Time may not be the solution
  › The nature of the supervisor-trainee relationship ("Educational Alliance\(\textsuperscript{†}\)) may play a more important role

Discussion
Discussion

CCERR Score
Resident Performance
On-Service
Off-Service
Conclusions

• As a cohort supervisors were not completing DECs to high standard

• Increasing supervisor-trainee continuity alone did not result in higher quality assessments

• Additional studies are needed to explore the nature of supervisor-trainee relationships and the impact on the assessment and feedback process
Help us improve.
Your input matters.

• Download the ICRE App,

• Visit the evaluation area in the Main Lobby, near Registration, or

• Go to: 

You could be entered to win 1 of 3 $100 gift cards.

Aidez-nous à nous améliorer.
Votre opinion compte!

• Téléchargez l’application de la CIFR

• Visitez la zone d’évaluation dans le hall principal, près du comptoir d’inscription, ou

• Visitez le 
1. The checklist/numeric ratings show **sufficient variability** to allow identification of relative strengths and weaknesses of the trainee

2. Comments are **balanced** providing both strengths and areas for improvement

3. The trainee’s **response to feedback** and/or remediation during the rotation is described in the comments

4. Comments **justify** the ratings provided

5. Clearly explained **examples of strengths** using specific descriptions (not generalizations) are provided in the comments

6. Clearly explained **examples of weaknesses** using specific descriptions (not generalizations) are provided in the comments

7. Concrete **recommendations** for the trainee to attain a higher level of performance are provided

8. Comments are provided in a **supportive manner**

9. Overall, this DEC provides **enough detail** for an independent reviewer to clearly understand the trainee’s performance on the rotation
# Generalizability Analysis

<table>
<thead>
<tr>
<th>Facet</th>
<th>VC</th>
<th>% Variance</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>0.146</td>
<td>15</td>
<td>The variance attributable to differences between supervisors</td>
</tr>
<tr>
<td>d:s</td>
<td>0.130</td>
<td>13</td>
<td>The variance attributable to differences between DECs within particular supervisors</td>
</tr>
<tr>
<td>r</td>
<td>0.000</td>
<td>0</td>
<td>The variance attributable to differences between raters</td>
</tr>
<tr>
<td>i</td>
<td>0.189</td>
<td>19</td>
<td>The variance attributable to differences between items</td>
</tr>
<tr>
<td>sr</td>
<td>0.000</td>
<td>0</td>
<td>The variance attributable to the supervisor-rater interaction (i.e. did some raters rate supervisors differently than other raters?)</td>
</tr>
<tr>
<td>si</td>
<td>0.082</td>
<td>8</td>
<td>The variance attributable to the supervisor-item interaction (i.e. did some items have different scores depending on the supervisor?)</td>
</tr>
<tr>
<td>dr:s</td>
<td>0.008</td>
<td>1</td>
<td>The variance attributable to the DEC-rater interaction (i.e. did some raters rate supervisors’ DECs differently than other raters?)</td>
</tr>
<tr>
<td>di:s</td>
<td>0.176</td>
<td>18</td>
<td>The variance attributable to the DEC-item interaction (i.e. did some items have different scores depending on the DECs a supervisor completed?)</td>
</tr>
<tr>
<td>ri</td>
<td>0.073</td>
<td>7</td>
<td>The variance attributable to the item-rater interaction (i.e. did some raters score items differently than other raters?)</td>
</tr>
<tr>
<td>sri</td>
<td>0.011</td>
<td>1</td>
<td>The variance attributable to the supervisor-item-rater interaction (i.e. did some raters score items and supervisors differently than other raters?)</td>
</tr>
<tr>
<td>dri:s</td>
<td>0.165</td>
<td>17</td>
<td>The variance attributable to the DEC-item-rater interaction plus random error</td>
</tr>
</tbody>
</table>

s (supervisor); r (rater); i (item); d (DEC); VC (variance component)
Discussion

![Graph with CTT, Non-CTT, and Off-Service trends]

- **CCERR Score**: Performance metric across different services.
- **Resident Performance**: Range from 0 to 5.
- **CTT**: On-Service performance.
- **Non-CTT**: Performance for non-CTT residents.
- **Off-Service**: Performance for off-service residents.