Revising the In-Training Evaluation Report (ITER) to Improve its Utility

Clarissa Agusto, MD, FRCPC, PGY-5
Vijay Daniels MD, MHPE, FRCPC, Associate Professor
Division of General Internal Medicine
University of Alberta
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

• Formative utility
  ➢ Shift to CBME – Assessment for Learning

• ITER is one of the most commonly used assessment tools
  ➢ yet is widely criticized for lack of validity

• Summative utility
  ➢ ITER Evidence of Validity
    » Scores reliable across multiple rotations and predictive validity for PGY3 Internal Medicine performance - Ginsburg et al
    » Within scale alpha too high (Kassam et al, locally)

• ? Halo effect with rater fatigue reducing comment quality
The purpose of this study is to evaluate the effect of reducing length of the ITER used for an Internal Medicine residency program on one aspect of validity, reliability, and on the quality of feedback.
Methods

• Setting

➢ Old ITER contained 35-items, based on the CanMEDS framework
➢ Consensus approach with Division of General Internal Medicine (GIM), reduced to 14-items
  » 13 items capturing broader competencies
  » 1 overall global rating scale
Methods

• ITER implementation
  ➢ 2015-16 academic year, PGY1 residents rotate through 4 GIM hospital sites
  ➢ Shorter (14-item) ITER at 2 GIM sites
  ➢ Longer (35-item) ITER at 2 other GIM sites

• Faculty Development
Analysis

• ITER reliability (G-studies)
  ➢ Within scale (Phi)
  ➢ Across two rotations with same ITER (Phi)
  ➢ Decision study to estimate reliability across 6, and 12 rotations

• ITER formative quality
  ➢ Completed Clinical Evaluation Report Rating (CCERR), Dudek et al
    » Each ITER rated by 3 GIM subspecialty residents using CCERR
    » CCERR reliability - overall G, internal consistency, inter-rater reliability
    » T-tests for differences in mean CCERR scores of long and short ITER to compare quality
Results

• ITER reliability

<table>
<thead>
<tr>
<th>Phi (reliability)</th>
<th>Long (35-item)</th>
<th>Short (14-item)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within scale</td>
<td>0.96-0.98</td>
<td>0.95</td>
</tr>
<tr>
<td>Across 2 Rotations</td>
<td>0.54</td>
<td>0.47</td>
</tr>
<tr>
<td>Across 6 Rotations</td>
<td>0.78</td>
<td>0.73</td>
</tr>
<tr>
<td>Across 12 Rotations</td>
<td>0.88</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Results

• ITER Formative Quality
  ▶ 33 PGY1 residents, 131 ITERs, 393 CCERRs scores
  ▶ CCERR reliability
    » Overall reliability G coefficient = 0.75
    » High internal consistency, Cronbach alpha 0.88
    » Very strong interrater reliability, ICC 0.86
    » ICC for each item at least moderate agreement, except item #1 (ICC -0.07)
Results

• ITER Formative Quality
  » Average CCERR score (with and without Item #1)
    » Long ITER 2.76 vs Short ITER 2.92 (p-value 0.076)
    » Long ITER 2.68 vs Short ITER 2.85 (p-value 0.051)
Discussion

• Shortened ITER scores slightly less reliable
  ➢ But still acceptable for use
    » to determine who can progress to more advanced rotations earlier

• Shortened ITER likely provides better quality of feedback
  ➢ Although p=0.076 (0.051 without Item 1)
    » Would meet standard p<0.05 with one-tailed t-test
  ➢ Anecdotally (need to study)
    » Improved feasibility of use and timeliness of completion
Discussion

• “Re-evaluation” of the CCERR
  › Very good reliability of CCERR scores
  › Feasible: on average 3 to 4 mins/ITER
Limitations

• Broader applicability
  ➢ PGY1 residents, GIM ward rotation
  ➢ Internal Medicine Residency Program at one institution

• Unusually high within scale reliability coefficients
  ➢ Rater errors
  ➢ Scale issues
  ➢ Items can be further reduced
Future Directions

• Implementation of the shortened ITER to all rotations in our program
• Further reduced ITER to 10-items
• Rotation-specific item content
• Further analysis of data across subspecialty rotations
References

References

- Kassam A, Donnon T, Rigby I. Validity and reliability of an in-training evaluation report to measure the CanMEDS roles in emergency medicine residents. CJEM 2014 Mar; 16(2):144-150.
Thanks for your attention!

• Questions??
Help us improve.
Your input matters.

• Download the ICRE App,

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

• Go to: http://www.royalcollege.ca/icre-evaluations to complete the session evaluation.

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 http://www.collegeroyal.ca/evaluations-cifr afin de remplir une évaluation de la séance.