ITERs, OSCEs, MCQs, & SAQs: How well do commonly used resident-evaluation formats correlate with each other?

Dr. Lacey D. Pitre
Young-In Kim
Dr. K. Dore
Dr. P. Wasi

McMaster University
Department of Medicine
The knowledge base of McMaster Internal Medicine (IM) residents is evaluated using many different formats, which include the:

- Internal Medicine American In-Training Exam (AITE)
- Short-Answer Exam (SAE)
- Objective Structured Clinical Exam (OSCE)
- In-Training Evaluation Report (ITER)
AITE:
- Mean scores on the AITE correlated well with scores on the Royal College IM written exam for all post graduate years: $r=0.62$, $r=0.55$, and $r=0.65$ for years 1,2, and 3 respectively (Brill-Edwards et al., 2001).

SAE

OSCE
- Shown in one study to correlate with AITE scores: $r=0.30$ (Dupras and Li, 1995).

ITER
- Overall accuracy has been questioned in prior studies (Noel et al., 1992).
- 74.5% of clinical faculty surveyed indicated they felt an ‘unwillingness to record negative evaluations’ (Dudek et al., 2005).
Our objectives included:

- To determine the degree of overall correlation between the exams produced by the McMaster IM program (SAE, OSCE) and the nationally-standardized AITE with respect to resident knowledge-base.

- To determine how predictive the faculty-based ITER is of a resident’s knowledge-base as measured by the AITE and OSCE.
The evaluation data for the 84 IM residents (PGY-1/2/3) at McMaster in 2009 was compiled for all four evaluation formats.

- No OSCE score was available for PGY-1s.
- ITER score is for CTU rotations only.
- The ITER score used was the specific prompt ‘knowledge base’.

Each score was expressed as a single percentage.

Pearson correlations were determined using SPSS.
Results

- AITE
- OSCE
- SAE
- ITER

PGY1, PGY2, PGY3

n=84
## Results

<table>
<thead>
<tr>
<th></th>
<th>SAE</th>
<th>ITER (Medical Knowledge)</th>
<th>OSCE</th>
<th>AITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAE</td>
<td></td>
<td></td>
<td>r=.536</td>
<td>r=.771</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>p=0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>ITER (Medical Knowledge)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>r=-.060</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p=.596</td>
</tr>
<tr>
<td>OSCE</td>
<td>r=.536</td>
<td>r=-.066</td>
<td>-</td>
<td>r=.670</td>
</tr>
<tr>
<td></td>
<td>p=0.001</td>
<td>p=.699</td>
<td></td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>AITE</td>
<td>SAE and AITE r=.771</td>
<td>r=-.060</td>
<td>OSCE and AITE r=.670</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>p=.596</td>
<td>p&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>
Our ‘home-grown’ OSCE has a higher correlation with the AITE ($r=0.67$) in comparison to prior studies ($r=0.30$).

Both the OSCE and SAE strongly correlate with AITE scores indicating that the in-program developed exams can predict resident performance on a nationally standardized exam.

There is no correlation between the faculty rating of ‘knowledge base’ on the ITER and the quantitative exams used to determine knowledge base (AITE, OSCE).
Only 2009 data currently analyzed. Further data available for future studies.

No OSCE results for PGY-1s.

The knowledge-base prompt is only one small component of the ITER.

Results are compiled from a single institution only.
In light of faculty reluctance to record negative evaluations – how would making the ITER anonymous affect its correlation with knowledge-base? (future addition of companion-ITER).

How does the ability of the SMR/CMR ITER rating correlate with more objective measures of medical knowledge in comparison to staff?

Long-term outcomes:
- Does faculty development affect outcomes?
- Do early interventions/remediation strategies help residents in need?
Many Thanks!

- Many thanks to our team:
  - Young In Kim
  - Mary-Beth Ribble
  - Dr. Kelly Dore
- Lastly to Dr. P. Wasi for her guidance and overall direction of the project.

Internal Medicine In-Training Examination®. Available at: http://www.acponline.org/education_recertification/education/in_training/.


ITER (overall) correlated to a small degree with the SAE (r=.303 p=0.007), and did not correlate with the OSCE (r=0.114) or the AITE (0.075).

Noel et al. – using a randomized controlled trial evaluating a clinical scenario evaluated by the ITER – 48% of staff internists assessed the residents’ clinical skills as unsatisfactory or marginal, whereas 52% assessed them as satisfactory or superior.

How do the objective tests correlate with future performance?
- Independent predictors of disciplinary action: Low professionalism rating on year-end ITER.
- Independent predictor of decreased risk for disciplinary action: high performance on the ABIM.
<table>
<thead>
<tr>
<th>n/a</th>
<th>1: Many major deficiencies</th>
<th>2: Several important deficiencies</th>
<th>3: Satisfactory performance (at appropriate level)</th>
<th>4: Often exceeds level of training</th>
<th>5: Outstanding: exceeds level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>