An Elaborate Scheme to Investigate the Psychometric Properties of the Clinical-Decision Making Exam Questions

Isaac Li PhD1, Qiongqiong Liu MS1, Yi Wang MS1, Edward Tsai PhD1

1National Board of Osteopathic Medical Examiners, Chicago, Illinois

With a focus on competency-based outcomes1-3 in graduate medical education and assessment, a novel way to assess clinical decision-making (CDM) skills has emerged.2,3 CDM cases center on key features (KFs), critical steps required to diagnose and treat patients in clinical case scenarios.2,3 NBOME supplements multiple choice questions (MCQs) with CDM cases in the COMLEX-USA Level 3 exam to assess residents’ ability to choose appropriate tests, procedures, and treatments and ensure patient safety. (Online practice can be found at http://www.nbome.org/cdm.asp.)

CDM/KF cases include a clinical scenario comprised of a patient presentation, vital signs, and laboratory or instrument findings, followed by Short Answer (SA) or Extended Multiple Choice (EMC) items. The illustration here is based on response data from 54 CDM/KF cases administered in 2016, including 74 SA and 101 EMC items. 2,327 candidates responded to these cases. Overall item performance is given in the table below.

<table>
<thead>
<tr>
<th>P-Val_MN</th>
<th>P-Val_SD</th>
<th>Rpbis_MN</th>
<th>Rpbis_SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMC</td>
<td>0.613</td>
<td>0.214</td>
<td>0.241</td>
</tr>
<tr>
<td>SA</td>
<td>0.210</td>
<td>0.231</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Analyses at the case level involve difficulty and discrimination indices averaged over all KFs embedded in the case. In addition, factors like response time, speededness, question length, and position effect are taken into account because they are instructive to test forms assembled at the item (KF) level, indices like p-value, point-biserial correlation (r-value), and response time, as well as their relationship, are examined. For partial credit (polytomous) items, polyserial correlation is used to evaluate the discrimination power. The Rasch model is fitted to the data to yield item measures, different item fit statistics, and item response function (not shown here due to limited space). Analysis of options, including “killer” and “exceed the limit”, is the most informative for item reviewers to reveal the proportion of candidates picking each response and their correlation with the sum score.

In general, more competent candidates do better on a partial-credit item. We expect candidates with the highest points on the item obtain the highest total score. The resulting positive polyserial correlation implies r-values of each category increasing monotonously. For example, the 6 items (1-4 points) graphed on the left hold this trend except for two items at point 4, indicating potential collapsing of point 3 and 4 during item revision.

This presentation illustrates a comprehensive scheme of psychometric analyses devised to collect pertinent statistics from different sources inherent in the computer-based CDM/KF questions: cases, key features, options, and in particular, killer options and response limits. Test development will benefit from the feedback in revising and improving CDM/KF cases to help this innovative examination realize its full potential in evaluating clinical competence and patient safety management.

Conclusion


