Purpose: There have been few investigations into score gains for medical students who repeat multiple-choice tests. Consequently, little is known about the magnitude and validity of such gains. This study evaluated the score changes for examinees who repeated USMLE Step 2 CK.

Methods: Participants included 64,512 examinees first-time completing Step 2 CK in 2011 or 2012; 2,767 examinees failed CK on their first attempt, later retested. Score increases were tabulated, and the impact of measurement error was evaluated. Regression analysis was used to predict Step 2 CK scores on the first attempt and second attempt from Step 1 scores, and the residuals from each attempt were compared.

Results: (a) The mean gain was 20 points (179 to 199), or .79 SD units; (b) Score gains were similar for US and international examinees; (c) Five points of the increase was explained by regression to the mean due to measurement error; (d) Step 1 scores predicted first-attempt and second-attempt CK scores with equal accuracy (R² = .579, .581).

Conclusion: Score gains were larger than those reported in other contexts, where increases typically range from .25 to .75 SD units. 4,5 These findings add to the validity evidence supporting the use of scores from repeat attempts for decision-making purposes.

Table 1: Step 1 vs. Step 2 Score Changes for Examinees with Unimpeded Progress and Step 2 Repeaters

<table>
<thead>
<tr>
<th>Examinee Group</th>
<th>Step 1 Mean</th>
<th>S.D.</th>
<th>Step 2 Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMG</td>
<td>229.33</td>
<td>18.39</td>
<td>199.38</td>
<td>10.19</td>
</tr>
<tr>
<td>USMG</td>
<td>221.98</td>
<td>17.74</td>
<td>195.30</td>
<td>13.74</td>
</tr>
<tr>
<td>Total</td>
<td>223.93</td>
<td>18.55</td>
<td>196.67</td>
<td>14.88</td>
</tr>
</tbody>
</table>

Table 2: Predicted Step 2 CK Scores* for Repeaters Based on Examinees with Unimpeded Progress and Residuals from Each Attempt

<table>
<thead>
<tr>
<th>Examinee Group</th>
<th>Observed Score</th>
<th>Mean</th>
<th>S.D.</th>
<th>Predicted Score</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMG</td>
<td>229.30</td>
<td>18.44</td>
<td>221.30</td>
<td>199.30</td>
<td>10.19</td>
<td></td>
</tr>
<tr>
<td>USMG</td>
<td>222.90</td>
<td>17.74</td>
<td>195.30</td>
<td>13.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>226.93</td>
<td>18.55</td>
<td>200.60</td>
<td>14.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Unimpeded progress” for an examinee is passing both Step 1 and Step 2 CK on the first take. The regression analysis was used to predict 2 CK scores from Step 1 last attempt scores (often, but not always, taken prior to Step 2 CK), and the residuals from the first- and second-attempt scores were compared for each examinee group (USMG, USIMG, and IMG).

Conclusions

Score gains were larger than those reported in other contexts, where increases typically range from .25 to .75 SD units. Only a small portion of the score increase could be attributed to measurement error, leaving other factors (e.g., additional preparation) to explain the increased gain. The fact that USMLE repeaters are administered different test forms on repeat attempts eliminates the risk of relearning the same material or rote memorization, adding to the validity evidence supporting the use of Step 2 CK scores from both initial and repeat attempts for the purpose of licensure.

References: Available on back of handout

Method

Data were analyzed for 64,512 examinees who took Step 2 CK for the first time in 2011 or 2012 administration; 2,767 of these examinees failed CK on their first attempt and later retook Step 2 CK. For the subset of repeaters, score increases were tabulated, and Kelley’s regression residual scores were computed to evaluate the impact of measurement error.

Results

- The mean score increase was 20 points (179 to 199), or .79 SD units for the total group. Score gains varied across examinee groups but the difference was negligible.
- Approximately 5 points of the score increase for the total group could be explained by regression to the mean due to measurement error. Table 1 displays the size of the score gains by examinee group.
- Step 1 scores predicted first-attempt and second-attempt Step 2 CK scores with equal accuracy (R² = .579, .581) for the total group. Figure 2 depicts the strong relationship between Step 1 and Step 2 score for repeaters and the unimpeded progress examinees.
- Table 2 displays the predicted Step 2 scores for repeaters in each examinee group, using a model based on examinees with unimpeded progress. The root mean squared residuals confirmed that the 2nd-attempt scores were closer to the predicted scores; however, some of this effect can be explained by regression to the mean.

Abstract

Purpose: To evaluate score changes for examinees who repeated USMLE Step 2 Clinical Knowledge (CK), which is part of the licensure process for physicians seeking licensure in the US. Step 2 CK is intended to address whether the examinee can apply medical knowledge, skills, and understanding of clinical science essential for the provision of patient care under supervision. Examinees were from one of three subgroups: those graduating from US/Canadian medical schools (US/MS); those graduating from international schools (IMGs); and those who were not US citizens graduating from international schools (IMGs).

Background & Purpose

Rationale: Credentialing programs across a wide variety of occupations and professions rely on the use of standardized exams as a way to determine which candidates are ready for practice, and which are not. On virtually all of these exams, candidates who fail are given the opportunity to retake the exam, and thus, pass to these tests is extremely high. Given this, one might expect large score gains on certification retest attempts. However, there have been few investigations into the validity and the magnitude of score gains specifically for medical students who repeat multiple-choice tests in medical licensing exams.

Consequently, little is known about the magnitude and validity of such gains. While these score gains might reflect true knowledge increase as the result of additional study, they might also reflect other factors such as measurement error, or the candidate having previously seen and memorized test items. Thus, it is important to validate and evaluate the score gains so as to separate the measurement error from other potential causes of repeat score gains.

Purpose: This study evaluated score changes for examinees who repeated USMLE Step 2 CK, which is part of the licensure process for physicians seeking licensure in the US. Step 2 CK is intended to address whether the examinee can apply medical knowledge, skills, and understanding of clinical science essential for the provision of patient care under supervision. Examinees were from one of three subgroups: those graduating from US/Canadian medical schools, international medical schools, and those who were not US citizens graduating from international schools.

References: Available on back of handout