Investigating Score Gains for Repeat Examinees on USMLE® Step 2 Clinical Knowledge (CK)

Yu Ling MPA, Kimberly A. Swygert PhD, and Mark R. Raymond PhD
National Board of Medical Examiners

Abstract

Purpose: There have been few investigations into score gains for medical students who repeat multiple-choice tests.[1,2] 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[3] 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 nearly identical 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 (R^2 = .579; .581).

Conclusion: Score gains were larger than have been 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.

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 provided the opportunity to retest, and thus, motivation to pass 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 licensure tests in medicine.[1,2]

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 out the measurement error from other various potential causes of repeat score gains.

Purpose. This study evaluated 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 content 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[3]. Examinees were from one of three subgroups: those graduating from US/Canadian medical schools (USMGs); those where US citizens graduating from international schools (USIMGs); and, those who were not US citizens graduating from international schools (IMGs).

Table 1: Step 1 Scores for Examinees with Unimpeded Progress[1] and Step 2 Repeaters

Examinee Group	Examinees with Unimpeded Progress			Step 2 Repeaters		
	N	Mean	S.D.	N	Mean	S.D.
IMG	12,500	223.93	18.39	1,028	195.30	17.37
USIMG	7,351	220.85	17.21	868	196.18	14.50
USMG	34,782	229.30	18.44	673	199.38	10.19
Total[2]	54,633	226.93	18.55	2,569	196.67	14.88

[1] "Unimpeded progress" for an examinee is passing both Step 1 and Step 2 CK on the first take.
[2] Steps 1 and 2 CK can be taken in any order, so not all of the 64,512 examinees with Step 2 CK scores have Step 1 scores.

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

	Examinee with Unimpeded Progress (n=54,633)		Step 2 Repeaters (n=2,569)			
			1st-attempt		2nd-attempt	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Observed Score						
IMG	231.00	18.14	178.08	17.41	199.08	21.10
USIMG	228.97	17.79	179.40	16.32	197.46	21.32
USMG	241.28	17.45	185.00	9.69	204.68	15.56
Total	237.28	18.45	180.34	15.61	200.00	20.08
Predicted Score						
IMG	231.00	12.60	211.88	11.59	211.88	11.59
USIMG	228.97	12.19	212.49	9.67	212.49	9.67
USMG	241.28	12.04	221.30	6.80	221.30	6.80
Total	237.28	12.92	216.13	10.36	216.13	10.36
Residuals						
IMG	--	--	-33.80	14.78	-12.80	16.93
USIMG	--	--	-33.10	13.07	-15.04	17.59
USMG	--	--	-36.31	10.88	-16.62	15.16
Total	--	--	-35.79	13.44	-16.13	16.80
Root Mean Squared Residuals						
IMG	--	--	36.89	36.72	21.22	28.16
USIMG	--	--	35.58	33.17	23.14	30.19
USMG	--	--	37.90	30.12	22.49	26.28
Total	--	--	38.23	35.06	23.29	29.36

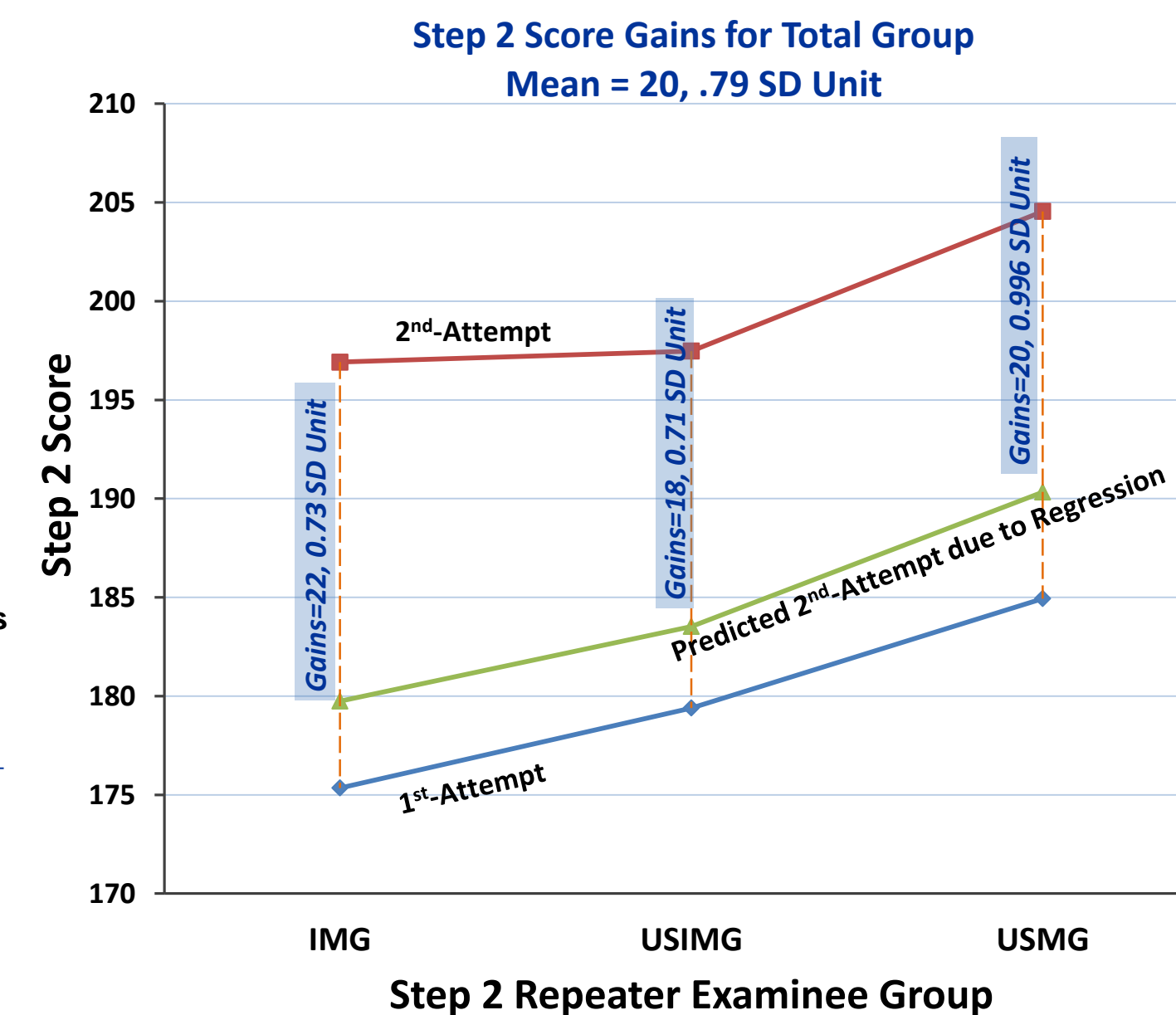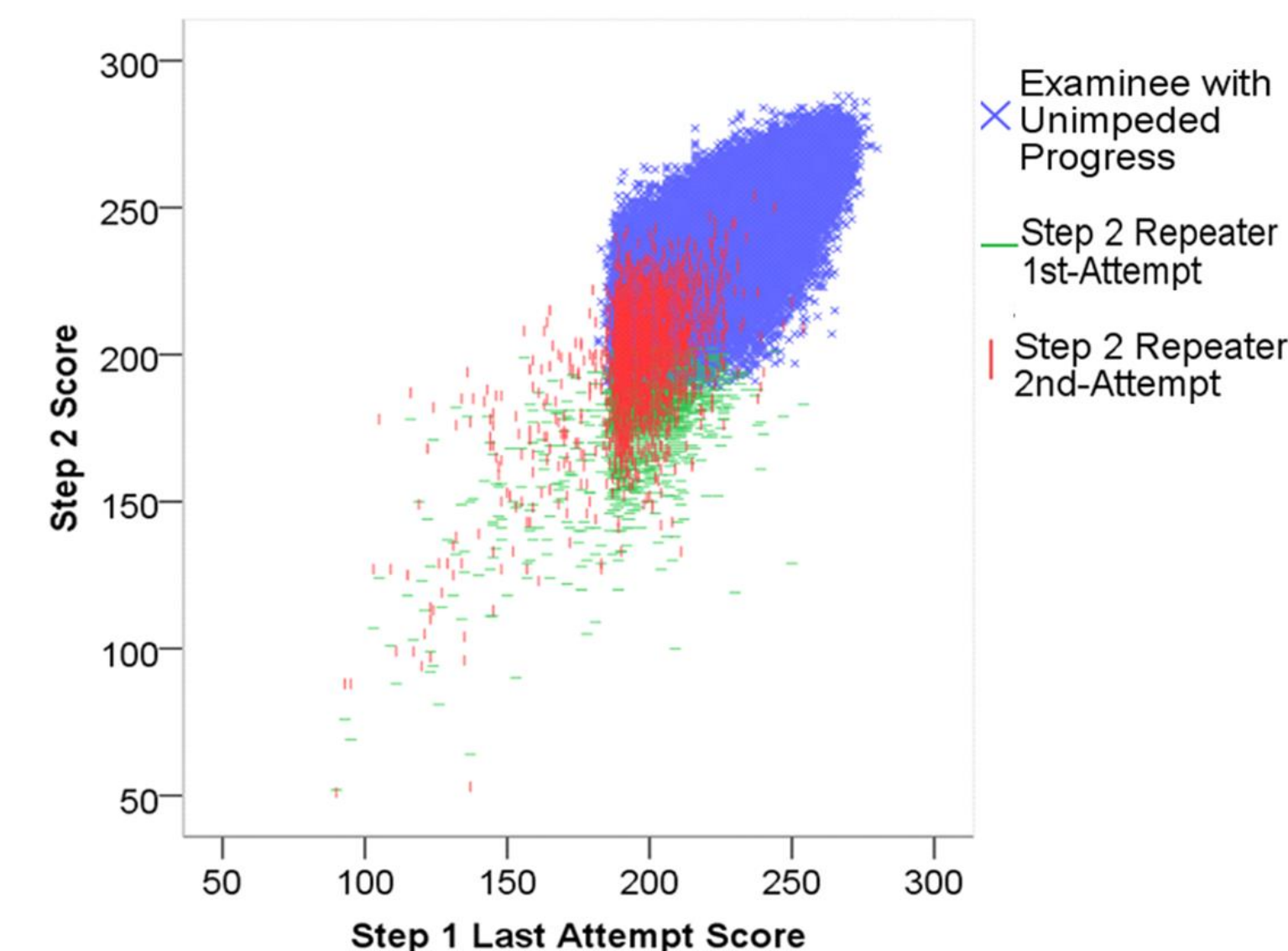### Figure 1: Step 2 Repeater Score Gains by Examinee Group

Figure 2: Relationship between Step 1 and Step 2 Score by Examinee Group

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 regressed true scores were computed to evaluate the impact of measurement error.[4] Regression analysis was done based on examinees with unimpeded progress, who passed Step 1 and Step 2 CK on the first take. The regression analysis was used to predict Step 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 were compared for each examinee group (USMG, USIMG, and IMG).

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. Figure 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 (R^2 = .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.

Conclusions

Score gains were larger than have been reported in other cognitive assessment contexts, where increases typically ranged from .25 to .75 SD units.[5,6] Only a small portion of the score increase could be attributed to measurement error, leaving other factors (e.g., additional preparation) to explain the increase. The fact that USMLE repeaters are administered different test forms on repeat attempts eliminates the possibility of them benefiting from memorizing questions seen during their first attempt. These findings add 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