Relationship between American College of Radiology In-training Examination Scores and American Board of Radiology Written Examination Scores

Bruce R. Baumgartner, MD, and Susan Brothers Peterman, MD, MPH

Rationale and Objectives. To retrospectively evaluate the association between resident performance on the American College of Radiology in-training examination and performance on the American Board of Radiology written examination.

Methods. Percentile scores from the in-training examination (low score, <20th percentile) and written board examination (low score, <25 percentile) were collected for 58 residents in a large, university-based program during a 6-year period. Mean in-training examination scores were compared for the high score and low score written board groups. In-training examination scores were correlated with the written board scores, and odds ratios were calculated for the association between in-training examination and written board scores. Adjusted in-training examination and written board odds ratios were calculated for Alpha Omega Alpha status and prior clinical training.

Results. The mean in-training examination scores were statistically significantly higher in the high score written board group ($P = .0001$). There was significant correlation between the in-training examination and the written board scores ($P = .05$). There was a significant association between a resident's average in-training examination score and written board score. Alpha Omega Alpha status was associated with high written board scores, and prior clinical training was associated with low written board scores (not significant).

Conclusion. The resident's average in-training examination score was a strong predictor of the written board score. The resident with a low in-training examination score is at risk for poor performance on the written board examination and may benefit from remedial training.

Key Words. American College of Radiology in-training examination; American Board of Radiology written examination; radiology residency; radiology education.

The American College of Radiology (ACR) Commission on Education and Committee on Residency Training have developed the In-training Examination for Diagnostic Radiology Residents. Its purpose is to provide the residents with information that is useful in the evaluation of their progress and to provide the program directors with data that are helpful in analyzing and evaluating their programs. The examination is intended to be a measure of general achievement in diagnostic radiology for use by residents and program directors. It is not intended for use in judging the performance of examinees for qualification to any postgraduate program or certification (ACR In-training Examination Brochure, 1995). However, this examination is often viewed as a practice test for the American Board of Radiology (ABR) written board examination. Many programs use the in-training examination to counsel residents regarding areas of deficiency as they prepare for

From the Department of Radiology, Emory University School of Medicine, Atlanta, GA.

Supported in part by the GE-AUR Radiology Research Academic Fellowship.

Address reprint requests to B.R. Baumgartner, MD, Department of Radiology, Emory University Hospital, 1364 Clifton Road NE, Atlanta, GA 30322.

Received February 29, 1996, and accepted for publication after revision June 12.

Acad Radiol 1996;3:873–878

the written board examination. The ABR encourages program directors to utilize in-training examinations to assess the progress of residents in training, to identify individual and program-related strengths and weaknesses, and to improve graduate radiologic education in general (ABR Certification Brochure, 1995).

The purpose of this study was to evaluate the relationship between the results of the in-training examination and the written board examination. We were especially interested in whether poor performance on the in-training examination was predictive of subsequent poor performance on the written board examination, given the probable difference in resident attitudes toward these two examinations, namely, that residents seem to study much more for the written board examination than for the in-training examination.

The primary null hypothesis was that the in-training examination score was not associated with the written board score. This study was designed to test this null hypothesis. To further evaluate the predictive value of the in-training examination score, the in-training examination score for each level of radiology residency was compared with the written board score.

MATERIALS AND METHODS

Study Design

This cohort study was performed retrospectively. The cohort comprised the six most recent residency groups at a large, university-based radiology residency program. Fifty-eight residents in the cohort were included; three were excluded because they did not complete all of their radiology residency in this program.

Predictor Variables

The overall score percentiles for the in-training examination from the 1st, 2nd, and 3rd years of residency were collected for each resident. Scores for the 1st year were not available for two residents, scores for the 2nd year were not available for four residents (one of whom did not have a 1st-year score), and scores for the 3rd year were not available for another two residents. An individual average in-training examination score was calculated for each resident.

All of the in-training examination scores were categorized *a priori* as either high or low. A score that was greater than or equal to the 20th percentile was considered "high," and a score that was less than the 20th percentile was considered "low." This distinction was chosen to identify those residents who might be at risk of failing the written board examination. The specific choice of the 20th percentile as the dividing point between high and low in-training examination scores was an arbitrary selection. Each resident's series of in-training examinations were also categorized as "all high" if all available scores were high or as "one low" if one or more scores were low.

Outcome Variables

The percentile scores for the diagnostic and physics portions of the written board examination were collected for each resident. These scores were categorized as either high or low. A score greater than or equal to the 25th percentile was considered "high," and a score less than the 25th percentile was considered "low." This *a priori* distinction was chosen because one resident actually failed a portion of the written board examination with a score in the 24th percentile. Any resident scoring in this range is at risk for failing the written board examination. Each resident's written board scores were also categorized as "both high" if both were high or as "one low" if at least one score was low. Seven residents were in the one low category, three of whom had a low score on the diagnostic portion and four of whom had a low score on the physics portion. Of these seven residents, only three actually failed that portion of the written board examination. Because of the low number of residents who did not pass, failure was not used as the outcome variable. No resident scored low in both portions. Only the initial written board results were considered.

Confounding Variables

Before analysis was begun, we proposed two potential confounding variables that also may have influenced the written board scores. These variables were the resident's Alpha Omega Alpha (AOA) Honor Medical Society status and whether the resident had clinical training before the radiology residency. These data were collected from the residents' application files.

Data Analysis

Data analysis was performed with the SAS/STAT statistical package (SAS Institute, Cary, NC) .

Comparability of the Written Board Score Groups

A comparison was made of AOA status; clinical training before radiology residency; overall score percen-

TABLE 1: Resident and Examination Characteristics According to Written Board Scores

Variable	Written Board Scores: Both High (n = 51)	Written Board Scores: One Low (n = 7)	P Value
AOA			
Yes	16	0	.173
No	35	7	
Prior clinical training			
Yes	30	7	.041*
No	21	0	
R1 ITE			
Mean percentile (SD)	56.67 (26.91)	35.86 (19.47)	.0543
No. with high score	43	6	1.000
No. with low score	6	1	
R2 ITE			
Mean percentile (SD)	55.73 (24.17)	25.50 (23.23)	.0055*
No. with high score	45	3	.013*
No. with low score	3	3	
R3 ITE			
Mean percentile (SD)	48.40 (29.52)	13.50 (12.93)	.0063*
No. with high score	40	2	.014*
No. with low score	10	4	
Average ITE			
Mean percentile (SD)	53.27 (23.37)	24.67 (8.92)	.0001*
No. with high score	46	4	.048*
No. with low score	5	3	
All ITE			
No. with all high scores	36	2	.041*
No. with one low score	15	5	

Note.—ITE = ACR in-training examination, R1 = 1st-year resident, R2 = 2nd-year resident, R3 = 3rd-year resident, SD = standard deviation.
*Statistical significance at P = .05.

TABLE 2: Correlation between In-training Examination Percentile Scores and Written Board Percentile Scores

In-training Examination Percentile	Written Board Percentile: Diagnostic	Written Board Percentile: Physics
First-year residents	r = .59801 (P = .0001)*	r = .42820 (P = .0010)*
Second-year residents	r = .66367 (P = .0001)*	r = .52438 (P = .0001)*
Third-year residents	r = .67888 (P = .0001)*	r = .47183 (P = .0002)*
Average	r = .75467 (P = .0001)*	r = .55434 (P = .0001)*

*Statistical significance at P = .05.

tiles for the in-training examination score from the 1st, 2nd, and 3rd years of residency; each resident's individual average in-training examination score; and all of each resident's in-training examination, according to both written board scores (Table 1). Continuous variables were compared by using an unpaired t-test comparison of the means. The categoric variables were compared by using two-tailed P values calculated with the Fisher exact method. The χ^2 test was not used; each comparison had at least one cell with an expected value less than five.

Correlation between In-training Examination Scores and Written Board Scores

For each resident, the overall score percentiles for the in-training examination from the 1st, 2nd, and 3rd years and the average in-training examination score were correlated with the diagnostic and physics written board scores. Pearson correlation coefficients and P values were calculated for each correlation (Table 2).

Association between In-training Examination Scores and Written Board Scores

Odds ratios with test-based 95% confidence intervals were calculated (Table 3). The in-training examination scores were represented by two variables, the average in-training examination score and all in-training examination scores. The average in-training examination score was chosen to represent a resident's scores for all in-training examinations taken. All in-training examination scores were analyzed to determine whether poor performance on at least one in-training examination was predictive of poor written board scores. The written board scores were considered together. This outcome was chosen because of the importance to the resident of passing both portions of the written board examination.

Individual Subgroup Analysis

Subgroup analysis was performed on two variables, the resident's AOA status and whether the resident had clinical training before the radiology residency. Associations of each subgroup with written board results were calculated (Table 3).

The associations between in-training examination scores and written board scores were adjusted for each subgroup (Table 4). Mantel-Haenszel summary odds ratios were calculated along with the respective test-based 95% confidence intervals. Logistic regression analysis could not be performed for a combined subgroup analysis because many of the subgroup categories contained zeros.

RESULTS

Comparability of the Written Board Score Groups

Table 1 shows the comparison of in-training examination scores and confounding variables between those residents with both high written board scores and those with at least one low written board score. The *t* test is a method to determine whether the difference in the mean or average of a variable (eg, in-training examination percentile) for two groups (eg, written board scores) is real or due to chance alone. The Fisher exact test is a method to determine whether the difference in percentages of a variable (eg, high 2nd-year score percentile for the in-training examination) in two groups (eg, written board scores) is real or due to chance alone.

Except for the 1st-year score percentiles for the in-training examination, which approached statistical significance, the mean percentile in-training examination scores were statistically significantly lower for the residents with one low written board score (P = .0055/.0063/.0001). When the in-training examination scores were divided into high and low categories, a greater percentage of residents with one low written board score had low in-training examination scores. This finding was statistically significant except for the 1st-year score percentiles for the in-training examination (P = .013/.014./.048/041).

For the confounding variables, a lower percentage of residents with one low written board score were AOA members. A higher percentage of residents with one low written board score had undergone prior clinical training. The clinical year variable was statistically significant (P = .041).

Correlation between In-training Examination Scores and Written Board Scores

Table 2 lists the Pearson correlation coefficients (r) and P values. The Pearson correlation coefficient is a measure of the degree of linear correlation between two variables. There is no correlation at $r = 0$; $r = 1$ is a measure of perfect correlation. The associated P value tests whether the linear correlation is statistically significant.

There was high correlation between the in-training examination scores and the diagnostic written board scores (P = .0001). There was a trend toward increasing correlation as the resident advanced in training. The best correlation was between the resident's average in-training examination score and the diagnostic written board scores (P = .0001). Although still statistically significant, the correlations were not so high between the in-training examination scores and the physics written board scores; again, the resident's average in-training examination score had the best correlation (P = .0001).

Association between In-training Examination Scores and Written Board Scores

Table 3 lists the odds ratios for the two different ways of classifying the in-training examination scores. The odds ratio is a measure of the strength of the association between a predictor variable (eg, high in-training examination score) and an outcome variable (eg, high written board score). There is no association when the odds ratio is 1. If the outcome variable (eg, high written board score) is more likely with a predictor variable (eg, high in-training examination score), then the odds ratio is greater than 1. The association becomes stronger as the

TABLE 3: Association with Written Board Examination Scores according to Odds Ratio

Variable	Written Board Scores: Both High (n = 51)	Written Board Scores: One Low (n = 7)	Odds Ratio	95% CI
AOA				
Yes	16	0	6.972	0.375, 129.482
No	35	7		
Prior clinical training				
Yes	30	7	0.095	0.005, 1.746
No	21	0		
Average ITE				
No. with high score	46	4	6.900	1.385, 34.375*
No. with low score	5	3		
All ITE				
No. with all high scores	36	2	6.000	1.193, 30.175*
No. with one low score	15	5		

Note.—CI = confidence interval, ITE = ACR in-training examination.
*Statistical significance at P = .05.

TABLE 4: Adjusted Associations between In-Training Examinations with Both Written Board Scores

Stratified Variable	Average ITE: Odds Ratio	Average ITE: 95% CI	All ITE: Odds Ratio	All ITE: 95% CI
AOA	8.000	1.400, 45.722*	4.792	0.875, 26.248
Prior clinical training	6.750	1.131, 40.294*	6.875	1.237, 38.198*

Note.—CI = confidence interval, ITE = ACR in-training examination.
*Statistical significance at P = .05.

odds ratio increases. If the outcome variable (eg, high written board score) is less likely with a predictor variable (eg, prior clinical training), then the odds ratio is between 0 and 1. When the 95% confidence interval for the odds ratio does not include 1, then the association is statistically significant (P = .05).

Both ways of classifying the in-training examination scores showed strong associations between low in-training examination scores and one low written board score. These associations were statistically significant (P = .05).

Individual Subgroup Analysis

Individual subgroup analysis was performed to determine whether the primary association (eg, average in-training examination score and written board score) was actually the result of other confounding variables (eg, AOA status). First, the association between the confounding variable (eg, AOA status) and the outcome variable (eg, written board score) was evaluated. Then the primary association (eg, average in-training examination score and written board score) was adjusted to remove any possible effects of the confounding variable (eg, AOA status).

AOA status. AOA membership was associated with both high written board scores; however, the effect was not statistically significant (Table 3). When adjusted for AOA status (Table 4), there was an even greater association of high average in-training examination scores with both high written board scores (P = .05). The adjusted association when using the all high in-training examination scores was lower and not statistically significant.

Clinical training before radiology. Clinical training before the radiology residency was associated with one low written board score (Table 3); however, the effect was not statistically significant. When adjusted for prior clinical training (Table 4), there was little change in the associations of high in-training examination scores with both high written board scores.

DISCUSSION

There has been widespread use of the in-training examination among radiology residency programs. In

1994, it was administered to 3,843 residents in 237 diagnostic radiology programs (Mettler FA, final report of ACR in-training examination scores, 1994). Each resident and each program develop individual uses for the in-training examination results, from self-evaluation to predicting performance on the written board examination to initiation of remedial measures for residents [1]. Passing the written board examination is very important because it is one of the requirements for ABR certification.

The in-training examination was not designed to predict written board examination performance; however, many people have wondered how well the two examinations correlate [2]. A 1993 survey addressing this issue produced unclear results [3]. In our study, the in-training examination score was a strong predictor of the written board score. The mean in-training examination percentiles were significantly lower for those residents who had low written board scores. There was good, statistically significant correlation between in-training examination scores and written board scores. There was a significant association between low in-training examination scores and low written board scores as measured by the odds ratio. Our results suggest that an in-training examination score of less than the 20th percentile identifies a resident at risk for poor performance or failure on the written board examination.

A resident's average in-training examination score was the best predictor of the written board score in our study. This finding was not surprising given the variability in scores expected from a resident after taking the in-training examination several times. The average score should be a better estimate of the resident's actual level of knowledge. Thus, the average score would be expected to be a better predictor of the written board score.

The in-training examination score was better correlated with the diagnostic portion of the written board examination than with the physics portion. This difference may be explained by the in-training examination format, which includes basic medical and clinical radiologic questions as its major component. The in-training examination correlation was still good for the physics portion of the written board examination.

Results of the analysis of the confounding variables and the written board scores were interesting. The relationship of lower written board scores to prior clinical training may have been related to the resident selection process. The radiology residency positions that begin immediately after medical school are fewer in number and potentially more competitive in many programs. Thus, more highly qualified residents may have been in the group without clinical training before radiology residency.

None of the residents who were AOA members failed the written board examination, regardless of in-training examination scores. This effect was not statistically significant, but the lack of significance may be a function of small sample size. It may be that AOA residents developed better study habits and test-taking strategies to succeed in medical school, and they may have applied these skills to preparing for and taking the written board examination. Our study suggests that the in-training examination scores may be less predictive of written board scores in those residents who are AOA members.

REFERENCES

1. Edeiken BS. Remedial program for diagnostic radiology residents. *Invest Radiol* **1993**;28:269–274.
2. Mettler FA, Briggs J, Bresch JP, Schaufeld B. Survey of the American College of Radiology Diagnostic Radiology ITE. *Invest Radiol* **1994**;29:726–728.
3. Talner LB. Editorial. *Invest Radiol* **1994**;29:729.