

AMERICAN BOARD OF SURGERY 2009 IN-TRAINING EXAMINATION EXPLANATION & INTERPRETATION OF SCORE REPORTS

Attached are the performance reports and analyses for participants from your surgery program on the 2009 American Board of Surgery In-Training Examination. The materials include: four separate reports for 2009, as described below; a norms table showing national averages for the total group and for each training year level; score distribution graphs for each examination, by year level; charts that show total test percentiles for each year level; and, a table that shows the relationship between scores on the 2008 In-Training Examination (Senior-Level) and 2008 ABS Qualifying Examination.

Examinee names are listed on the reports **exactly** as coded on their answer sheets. If names were not provided, they will either be blank on the score reports or identification numbers will be inserted. In cases where social security numbers were not provided, “dummy” ID numbers were assigned. Where year levels were not provided, percentile scores could not be computed, but these can be determined from the enclosed percentile charts. Likewise, percentiles for other year levels can be determined from these charts in cases where residents listed incorrect year levels on their answer sheets. As noted in all previous correspondence, **the ABS will not prepare corrected reports.**

Purpose and Structure of the In-Training Examination

As noted in all documentation for the ABS ITE over the years, *the primary goals of the In-Training Examinations* are: 1) to describe a content domain of basic knowledge important for successful clinical practice; 2) to assess relative strengths and weaknesses of individual residents at a time early enough in their training so that deficits can be corrected; and 3) to help directors of residency programs detect areas of relative strengths and weaknesses in their program. Specifically, the American Board of Surgery Booklet of Information indicates that the In-Training Examinations are designed to test the general level of knowledge attained by residents regarding the fundamentals of the basic sciences and the management of clinical problems related to Surgery

In order to reemphasize the primary goals of the ITE, the Board moved to two versions of the In-Training Examination in 2006 – the Junior- and Senior-level examinations (IJE & ISE) – in order to incorporate a greater number of questions that are relevant to each group’s level of clinical activity. The Junior-level examination incorporated increased emphasis on basic science and the clinical management items focused on the diagnosis and evaluation of common surgical problems. On the Senior-level examination, the clinical management items had greater depth and breadth than before and the basic science items focused on areas that are deemed important but apparently not well understood.

Approximately 60% of the Junior Level Examination items are designated as “basic science” and 40% are designated as “clinical management”. For the Senior Level Examination, 20% are designated as “basic science” and 80% are designated as “clinical management”. In addition, each item was also categorized into one of the five content categories traditionally reported for the In-Training Examination: 1) body as a whole; 2) gastrointestinal system; 3) cardiovascular and respiratory systems; 4) genitourinary, head and neck, musculoskeletal, and nervous systems; and 5) endocrine, hematic, breast, and lymphatic systems. The relative emphasis on each category is reflected in the number of items in each subtest (see Report “A/B” for the percentage breakdown). Surgical basic science items specifically address anatomy, physiology, bacteriology, biochemistry, pathology, and immunology.

ITE Scores

In addition to the above changes in structure and content, there was also a change in the way scores are presented. As a result of these changes, the meaning of scores on the 2009 examination may differ from those on ABS In-Training examinations prior to 2006. Percentile scores, where examinee performance is compared to others in their national peer group (i.e., training-year level), are calculated for the Total Test but are no longer calculated for examination subtests. Standard scores for the Total Test and for the Clinical Management and Basic Science subtests were again calculated in 2009 (these scores were not prepared in 2006 or 2007). Score reports are designed to emphasize content mastery and an absolute level of knowledge (given the primary intent of the exam) rather than emphasize one's performance relative to their peers (although percentiles and standard scores are normative measures). Scores related to specific subtest content focus on number of items correct and content descriptions of items that were missed. Total test percentiles for some residents may differ from prior years due to the changes in test content. In particular, scores may change for residents who took the Junior-Level Examination in 2008 but the Senior-Level Examination in 2009. Neither the percentiles nor standard scores are comparable across the two versions of the ITE due to different norm groups and different test composition.

One must be cautious in comparing residents' percent correct scores from year to year since the difficulty of the examination can change each year. No direct statistical comparison of changes in test difficulty can be made with examinations from the past few years since test questions are not repeated for several years. For example, the average total scores on the IJE from 2006 through 2009 were 66%, 70%, 69%, and 68% correct, respectively; averages for the ISE in the same years were 64%, 67%, 72%, and 66% correct, respectively. One cannot say whether these differences are due to changes in performance of residents or changes in examination difficulty.

Despite the inability to statistically equate current with previous examinations (due to lack of common, anchor items), the 2009 Senior-Level examination is certainly much more difficult than last year, but similar to difficulty values in 2006 and 2007. At face value, the the 2009 Junior-Level examination seems slightly more difficult than examinations in the past two years. Numerous difficult items, deemed by the Test Committee as representing important content material, were retained in final scoring on both examinations; several items were correctly answered by relatively small percentages of residents (7% of IJE items and 12% of ISE items). This year, only 1% of examinees scored 80% correct or higher on the ISE (compared to 15% in 2008 but only 3% in 2007); clearly, there was a ceiling effect on scores on the 2009 ISE. On the IJE, 12% scored 80% correct or higher (compared to 18% in 2007 and 15% in 2008). **In conclusion, both 2009 examinations seemed more difficult than the 2008 examinations (particularly the ISE).**

Given the stated purposes of the ITE, the changes in examination structure and score reporting are designed to reemphasize the importance of the keyword feedback provided with examination results. Keywords illustrate potential strengths and weaknesses of programs and residents, indicating areas that may not have been mastered. Keywords are intended to provide generic descriptors of item content. They are intended to be very general, relating to a topic area rather than narrowly focusing on more specific content. Hopefully, the listing of a keyword for an item answered incorrectly would motivate the resident to review resource material that discusses the general topic.

Relationship between ITE and the ABS Qualifying Examination

Data showing the relationship between the 2008 Senior-Level In-Training Examination and the 2008 ABS Qualifying Examination are presented in a separate section at the end of this document (Predicting Success on the Qualifying Examination from In-Training Examination Scores). The correlation between scores on the two examinations was 0.58 (corrected correlation of 0.66). The relationship between ITE percentiles and Pass/Fail on the QE was relatively strong (correlation of .92). Approximately 92% of those examinees who scored above the 30th percentile on the In-Training Examination passed the Qualifying Examination on their first attempt, compared to only 72% below the 30th percentile. (However, many who scored very low on the 2008 ITE did not take the 2008 QE).

Important Note for Program Directors

The In-Training Examinations are designed to assess a resident's knowledge base. An adequate knowledge base is considered to be a prerequisite for clinical expertise or competence. While written examination scores may be related to clinical competence, Program Directors need to keep scores in perspective. For example, high clinical management scores may not assure that residents are clinically competent if they cannot apply their knowledge in clinical situations. Low scores may not match Program Directors' ratings of clinical competence if the ratings, themselves, are not reliable or valid measures. Also, one must recognize that written examinations, in general, have some inherent limitations in the extent to which they measure higher order skills. In summary, the Board has continually maintained that In-Training Examination scores reflect medical knowledge and, as such, represent only one aspect of overall "competence". Other competencies to be considered are patient care, interpersonal and communication skills, professionalism, practiced-based learning and improvement, and systems-based practice. These competencies represent a summative assessment of residents' knowledge, clinical skills, attitudes, habits, and technical expertise. Such assessments are obviously very complex and require the skilled judgment of Program Directors.

2009 ITE Score Reports

While the content and structure of the examinations and specific examination scores have changed as noted above, the Board has prepared score reports that are analogous in structure to previous ABS In-Training Examination reports. Reports prepared for each program are described below. There are separate Junior- and Senior-Level Examination versions of each report.

Program Summary - Report A/B (Junior- and Senior-Level Examinations)

This report presents a list of examinee scores on the Total Test, on the Clinical Management and Surgical Basic Science subtests, and on the five major Body System subtests included in the examination. Examination structure and content is described at the top of each report. Examinees are listed according to the training-year level marked on their answer sheet. **Percent correct scores** are reported for each subtest. These scores represent the percentage of subtest content mastered by each resident. In addition, each examinee's **percentile score** on the Total Test is calculated within each training-year level. This score indicates the percentage of examinees that scored below a particular percent correct score within that year level. For example, a percentile of 60 indicates that the examinee's score was higher than 60% of other examinees at the same training-year level. A percentile score of 0 indicates that the examinee scored in the lowest 0.5% of all examinees at that particular training-year level. Percentiles are not available for cases where residents did not code a year level on their answer sheet or for those designated as "Staff". **Standard (scaled) scores** are presented for the Total Test as well as for the Clinical Management and Basic Science subtests. These scores are calculated from the total examinee group (regardless of year level). They are linear transformations of raw scores that arbitrarily set the mean/average of the total group equal to 500 and the standard deviation equal to 100. The scores show each resident's performance level relative to the entire examinee group for the specific examination (different norm groups for Junior and Senior exams). Finally, **national average** percent correct and standard scores are also shown for each training-year level. These averages are also summarized in the attached "National Norms" table.

Examinee Performance - Report C (Junior- and Senior-Level Examinations)

An individual performance report is provided for each examinee in your program. All subtest scores described above are included in the individual examinee report. Also, the keywords/phrases of all items that the examinee answered incorrectly are listed according to subtest. In keeping with the "diagnostic/prescriptive" purpose of the In-Training Examination, it is intended that residents follow-up on content of questions that they missed.

Item Summary - Report D (Junior- and Senior-Level Examinations)

As Program Director, you are provided with a summary which groups items according to content categories and lists the keywords/phrases of the questions answered incorrectly by your residents. The number of residents in each year level of training in your program who incorrectly answered the questions is shown. As noted above, keyword answered correctly/incorrectly by your residents can indicate particular areas of strength/weakness in your program.

Items Not Included in Final Scoring - Report E (Junior- and Senior-Level Examinations)

This is a listing of the keywords describing those items not included in final scoring of the 2009 In-Training Examination. For the most part, these reflect questions that were very difficult or questions that may have had other possible correct answers. Program Directors have a listing of all examination keywords in Reports D and E. It is the intent of the ITE Examination Committee that the total list of keywords reflects important topics for study.

AMERICAN BOARD OF SURGERY

2009 IN-TRAINING/SURGICAL BASIC SCIENCE EXAMINATION

NATIONAL NORMS: AVERAGE SUBTEST SCORES BY TRAINING YEAR LEVEL

Training Year Level: (N/% of Residents*)	<u>Junior Examination (IJE)</u>			<u>Senior Examination (ISE)</u>			
	Level I (2224/56%)	Level II (1715/43%)	IJE Total Group* (3965)	Level III (1509/40%)	Level IV (1117/30%)	Level V (1034/28%)	ISE Total Group** (3749)
<u>Standard Scores</u>							
Total Test	470	539	500	466	512	539	500
Clinical Management	467	542	500	462	512	544	500
Basic Science	474	534	500	493	506	507	500
<u>Percent Correct Scores</u>							
Total Test	65	72	68	64	67	69	66
Clinical Management	64	72	67	63	66	69	66
Basic Science	66	72	69	68	69	69	68
Body as a Whole	64	71	68	62	64	65	63
Gastrointestinal	68	77	72	65	69	71	68
CV, Respiratory	58	66	61	60	64	66	63
GU, Head & Neck, etc.	72	77	74	71	73	74	73
Endocrine, etc.	67	73	70	64	67	70	67

* “N” for the Junior Examination total group includes 11 Level III examinees, 12 staff, and 3 examinees who did not specify year level.

** “N” for the Senior Examination total group includes 23 Level I & II examinees, 22 Level VI examinees, and 41 staff.

AMERICAN BOARD OF SURGERY IN-TRAINING EXAMINATION
PREDICTING SUCCESS ON THE QUALIFYING EXAMINATION FROM
IN-TRAINING EXAMINATION SCORES

For years, the Board has analyzed the relationship between performance on the ABS In-Training Examination and the ABS Qualifying Examination. This report illustrates the relationship between the 2008 In-Training Examination and the 2008 Qualifying Examination. Given the radical modifications to the In-Training Examination in 2006, where the examination structure split the examination into Junior and Senior resident versions, results and correlations are not necessarily comparable to prior years. One might expect the correlation between the two examinations to improve since examination content has shifted from a 50/50 Basic Science/Clinical Management split to an 80/20 Basic Science/Clinical Management split. With increased emphasis on Clinical Management, the Senior-Level In-Training Examination structure more closely resembles the Qualifying Examination structure.

Over the years, In-Training Examination-Qualifying Examination correlations have ranged from .54 to .66. A total of 894 examinees took both the 2008 In-Training Examination and the 2008 Qualifying Examination. The correlation between scores on the two examinations was .58 in 2008, within the range of values seen in recent years. The corrected coefficient in 2008 (accounting for examination unreliability) was .66, again, similar to values observed over the past few years. Several residents who scored low on the ITE did not take the QE in 2008, restricting the range of scores and lowering the magnitude of the correlation. While the relationship is moderately high, one would expect some differences between performance on one examination when compared to performance on the other.

An expectancy chart illustrating the relationship between success on the 2008 Qualifying Examination (first-takers) and percentile scores on the 2008 In-Training Examination is provided in Table 1. The relationship is graphically portrayed in Figure 1. The correlation between ITE percentiles and Pass/Fail on the 2008 QE is high at .92. Table 2 combines results for 2007 and 2008; these data are graphically portrayed in Figure 2; the correlation between ITE percentiles and passing or failing the QE is very high for the combined group.

While it is generally true to say that a resident should score well on the Qualifying Examination if he/she scored well on the In-Training Examination, one cannot predict with absolute certainty. This year, there were only eight examinees above the 65th percentile on the 2008 In-Training Examination who failed the 2008 Qualifying Examination (compared to nine examinees last year and fewer in prior years). Several examinees below the 10th percentile on the 2008 In-Training Examination passed the 2008 Qualifying Examination (54%, compared to 58% last year and 42% in 2006). However, as noted above, many examinees with low In-Training Examination scores do not take the Qualifying Examination immediately after completion of their residency.

In summary, the relationship between In-Training Examination (Senior Level) and the Qualifying Examination scores is similar to recent years. The correlation in 2008 was .58 (.66 corrected value). The correlation between ITE percentile scores and Pass/Fail on the 2007 QE was high at .92. Approximately 92% of those examinees that scored above the 30th percentile on the In-Training Examination will pass the Qualifying Examination on their first attempt, compared to only 72% below the 30th percentile, similar to totals in prior years. As usual, a few exceptions were observed where those who scored high on the In-Training Examination failed the Qualifying Examination, while several of those who scored low passed.

TABLE 1
AMERICAN BOARD OF SURGERY
THE RELATIONSHIP BETWEEN IN-TRAINING EXAMINATION
TOTAL SCORES AND QUALIFYING EXAMINATION PERFORMANCE
(2008 EXAMINATION RESULTS)

IT/SBSE Percentile Range: Level V	# Pass (%)	# Fail (%)	Total
96 - 100	36 (100%)	0 (0%)	36
91 - 95	56 (98%)	1 (2%)	57
86 - 90	35 (100%)	0 (0%)	35
81 - 85	55 (93%)	4 (7%)	59
76 - 80	43 (100%)	0 (0%)	43
71 - 75	47 (98%)	1 (2%)	48
66 - 70	22 (92%)	2 (8%)	24
61 - 65	49 (93%)	4 (8%)	53
56 - 60	57 (97%)	2 (3%)	59
51 - 55	50 (91%)	5 (9%)	55
46 - 50	29 (88%)	4 (12%)	33
41 - 45	41 (87%)	6 (13%)	47
36 - 40	43 (83%)	9 (17%)	52
31 - 35	30 (73%)	11 (27%)	41
26 - 30	37 (86%)	6 (14%)	43
21 - 25	37 (88%)	5 (12%)	42
16 - 20	36 (78%)	10 (22%)	46
11 - 15	31 (66%)	16 (34%)	47
6 - 10	26 (63%)	15 (37%)	41
0 - 5	14 (42%)	19 (58%)	33
TOTAL	774 (87%)	120 (13%)	894 (100%)

TABLE 2

AMERICAN BOARD OF SURGERY

THE RELATIONSHIP BETWEEN IN-TRAINING/SURGICAL BASIC SCIENCE
EXAMINATION AND QUALIFYING EXAMINATION PERFORMANCE
(2007-2008 EXAMINATION RESULTS)

IT/SBSE Percentile Range: Level V	# Pass (%)	# Fail (%)	Total
96 - 100	80 (100%)	0 (0%)	80
91 - 95	98 (99%)	1 (1%)	99
86 - 90	88 (100%)	0 (2%)	88
81 - 85	91 (96%)	4 (2%)	95
76 - 80	81 (100%)	0 (3%)	81
71 - 75	87 (99%)	1 (7%)	88
66 - 70	54 (96%)	2 (4%)	56
61 - 65	78 (94%)	5 (7%)	83
56 - 60	120 (94%)	8 (9%)	128
51 - 55	95 (90%)	10 (8%)	105
46 - 50	75 (89%)	9 (11%)	84
41 - 45	82 (82%)	18 (12%)	100
36 - 40	95 (85%)	17 (14%)	112
31 - 35	72 (79%)	19 (14%)	91
26 - 30	82 (81%)	19 (19%)	101
21 - 25	60 (81%)	14 (29%)	74
16 - 20	56 (76%)	18 (27%)	74
11 - 15	65 (68%)	30 (31%)	95
6 - 10	39 (61%)	25 (43%)	64
0 - 5	43 (53%)	38 (58%)	81
TOTAL	1541 (87%)	240 (13%)	1781 (100%)

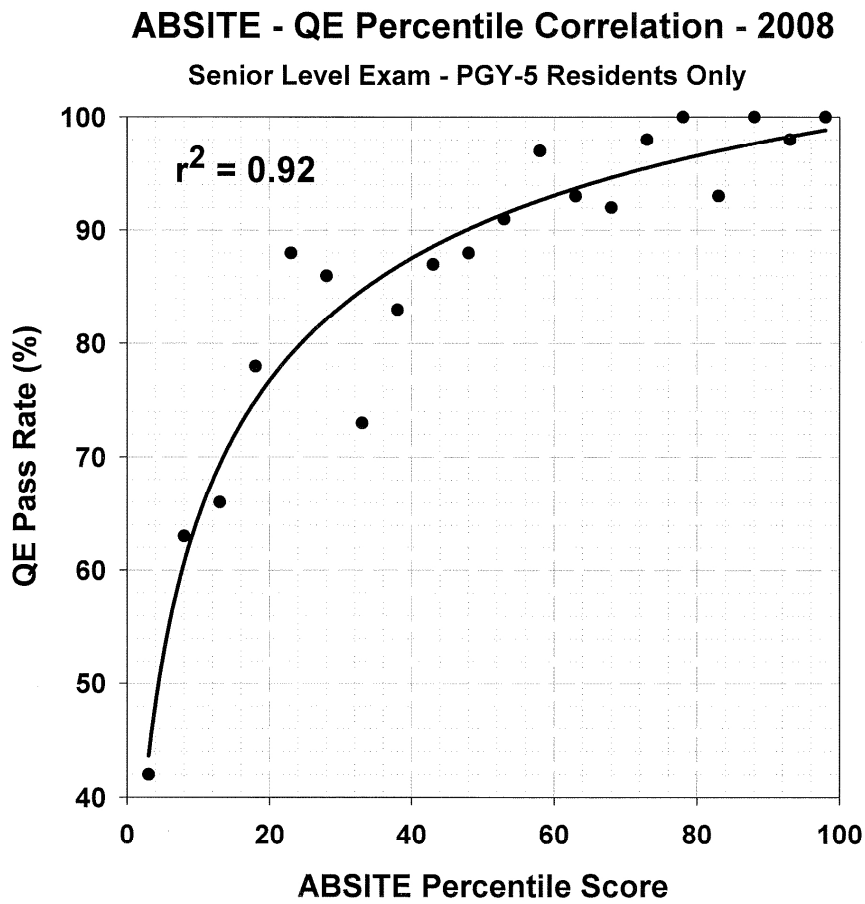


Figure 1 : Relationship between 2008 In-Training Examination percentiles (Senior Level) and Pass/Fail on the 2008 ABS Qualifying Examination

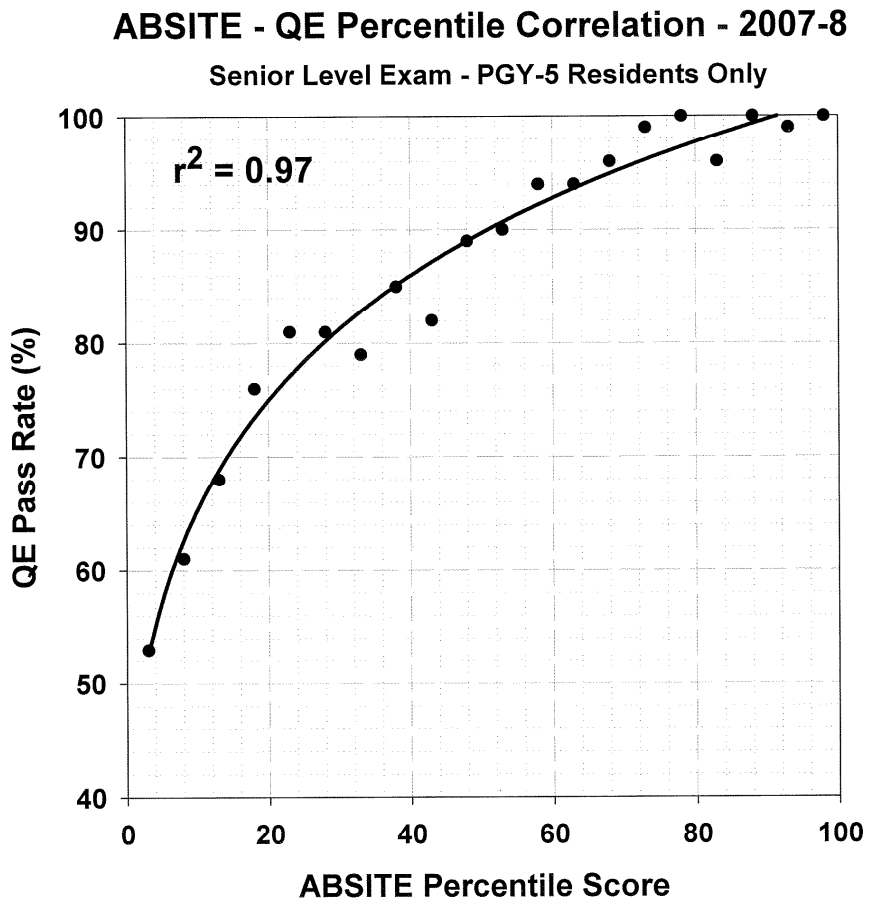


Figure 2 : Relationship between 2007 & 2008 In-Training Examination percentiles (Senior Level) and Pass/Fail on the 2007 & 2008 ABS Qualifying Examination; combined data files