STATISTICAL REPORT OF THE 2009 IBLCE EXAMINATION

Leon J. Gross, Ph.D.
Psychometric Consultant

In July 2009, the International Board of Lactation Consultant Examiners (IBLCE) administered its 25th annual credentialing examination in lactation consulting. The test was administered in 14 languages to 4,017 candidates in 159 locations across 40 countries and territories representing 5 continents.

This candidate population for the 2009 administration was the largest in the test's history, 21% higher than the prior record in 2008. This marks the third administration in the past four, inclusive, in which the candidate population exceeded all of the prior administrations. The population was also larger than the first nine test administrations combined; more than ten times the size of each of the first five administrations; more than twice the size of each of the first 16 administrations; the second consecutive administration with more than 3,000 candidates; and the first administration with more than 4,000 candidates. Across the 25 annual administrations, more than 40,000 candidates have sat for the test.

The 2009 candidate population continues a significant trend regarding its composition. For the first 14 years of the program (1985-1998), the United States alone accounted for the majority of the candidates. For eight of the nine most recent years, including 2009, the United States accounted for less than half the candidates. Similarly, for the first eight years of the program (1985-1992), candidates from countries *other* than the United States, Canada, and Australia accounted for less than 10% of the candidates. The 2009 administration is the tenth consecutive year in which candidates from these other countries have accounted for more than 30% of the candidates, and the third in which this constituency has reached or exceeded 40%.

These statistics underscore that the IBLCE credential is the global standard of competence assessment in lactation consulting. Table 1 displays the test centers by region and country, and includes the number of candidates in each.

As the program matures, a significant number of candidates take the test for recertification. For this administration, 829 of the candidates sat for recertification.

Table 1: Number of Test Centers and Candidates

Region and Country	*Number of Test Centers*	*Number of Candidates*
North America		
Canada	11	211
United States	63	1710
Central and South America		
Mexico	1	3
Argentina	1	4
Brazil	4	19
Peru	1	10
Europe		
Austria	2	59
Belgium	2	41
Croatia	1	17
Denmark	1	12
France	1	92
Germany	5	205
Greece	1	8
Ireland	1	29
Italy	3	33
Netherlands	1	61
Poland	1	20
Portugal	1	4
Slovenia	1	14
Spain	2	8
Sweden	1	4
Switzerland	1	87
United Kingdom	3	55
Middle East		
Egypt	1	71
Israel	1	42
Kuwait	1	4
Saudi Arabia	1	2
United Arab Emirates	3	98
Africa		
South Africa	3	13
Australasia		
Australia	20	305
China	1	33
India	1	23
Indonesia	1	27
Japan	5	221
Korea	4	357
Kyrgyzstan	1	1
Malaysia	1	11
New Zealand	4	45
Singapore	1	7
Taiwan	1	51

This was the fifth administration in which candidates sat for the examination at the 20-year recertification interval. Approximately 30% of the first four candidate cohorts have recertified by examination at the 20-year interval, a remarkable milestone for the individuals as well as for IBLCE and the profession. Overall, more than 20% of the candidates sat for the test to recertify. In short, the 2009 marked a continuation and acceleration of important trends in the candidate population composition.

The test was administered in English (both American and British), Croatian, Dutch, French, German, Indonesian, Italian, Japanese, Korean, Polish, Portuguese, Slovenian, Spanish, and Taiwanese. This was the first administration in which the test was translated into Croatian and Taiwanese.

A total of 1339 candidates sat for one of the 13 translated forms of the test. Although there were no culturally adapted versions of the test, the English version was linguistically adapted to British English for most English-speaking candidates in countries outside North America, and for candidates in countries in which English is a secondary language but the test was not translated into the primary language. The IBLCE examination has now been administered in 19 languages (Arabic, Croatian, Dutch, English, French, German, Hebrew, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Polish, Portuguese, Slovenian, Spanish, Swedish, and Taiwanese) in nearly 50 countries across all major continents, offering unparalleled geographical, cultural, and linguistic access.

The 2009 administration was the 19th in which IBCLCs (International Board Certified Lactation Consultants) chose to recertify by examination. Recertification is required every five years. By IBLCE policy, the first 5-year recertification requirement may be satisfied by either continuing education recognition points (CERPs) or by examination. When certificants are recertified by CERPs, the next 5-year recertification must be fulfilled by examination.

Historically, the vast majority of certificants has recertified at the 5-year point by CERPs, with small numbers of certificants recertifying initially by examination. As noted earlier, of the 4017 candidates, 829 sat for the test for recertification. An analysis of the recertification candidate performance is given in Table 5 and Figure 2 later in this report.

Examination Development and Structure

The IBLCE examination is based on a 3-dimensional content outline, or test blueprint. This document was derived from a practice analysis by the Board, in conjunction with its Examination Committee members and a Representative Panel of Experts (RPE). On the basis of this study, the Board arranged the examination content according to scientific disciplines, developmental stages, and taxonomy levels. The latter category indicates whether an item measures recall of knowledge (level 1) or application of knowledge (level 2).

As the end of its first ten years approached, the Board commissioned a study of the blueprint for possible updating. A panel of experts was assembled for this purpose, and it developed a survey of certificants to provide an empirical basis for any modifications deemed to be necessary. On the basis of this study, the underlying structure of the blueprint remained the same, but numerous secondary revisions were recommended. These included changing the two major classification dimensions from scientific disciplines and developmental stages to disciplines and chronological periods, respectively. In addition, greater detail was provided for each of the disciplines, and the chronological periods were restructured. The relative emphasis (i.e., range of items) of most of these categories also changed, but most of these changes were relatively slight. Finally, the number of disciplines increased from 12 to 13. The examination blueprint appears in its entirety on the IBLCE website (www.iblce.org).

The 2009 examination was developed by a 7-person examination committee that prepares, reviews, edits, and selects test items. The Examination Committee includes broad representation, with the objective of including the following types of subject matter experts.

- at least one physician who is a pediatrician
- the highest scorer from the previous year's examination
- a hospital-based IBCLC
- an IBCLC in private practice
- an educator in lactation management
- an IBCLC who received her training primarily through the mother-support system
- an IBCLC who received her training primarily through the traditional health professions
- at least two members from outside the US
- a midwife
- a dietician
- a Ph.D.- level researcher in lactation
- at least one member from the Australasian region
- at least one member from Europe
- a Canadian

The Committee also includes the Executive Director and one of the two consulting psychometricians. Additional staff participate in the meeting to monitor test item revisions and check item references. Since there are fewer committee members than constituencies, one member usually represents more than one constituency. If a second physician serves on the Committee, he or she is typically an obstetrician.

As noted earlier, the test was translated into 13 languages. This marks the 23rd consecutive year in which the examination has been translated. The opportunities provided by the translated test were made possible by the development of policies and standard operating procedures to govern the translation process. These policies and procedures were adopted and evaluated for the administration and analysis in 1986 of a Spanish "mini-test," consisting of a 25% sample representative of the content and difficulty of the overall 200-item test. A full explanation of the procedures used in developing this test, and an analysis of the comparative results, are contained in a published journal article and available on request.

On the basis of the procedures developed and implemented for this pilot test, IBLCE began offering a complete translated version of the examination in any language based on the Roman alphabet in 1987. Translations using non-Roman alphabets began in 1994. The criterion for translation is the documentation of sufficient number of candidates to support the requisite effort and costs.

Of the 200 multiple-choice test items, 75 contain an image. This stimulus is usually in the form of a color photograph depicting an aspect of breastfeeding, or breast anatomy or pathology that the candidates must resolve. These test items have a particularly high degree of clinical relevance, and because most of the question is contained in the graphic, potential linguistic issues are minimized.

Examination Results

The results of the 25th administration of the IBLCE examination are displayed in Table 2. These statistics are based on the total population of 4017 certification and recertification candidates. The data in Table 2 are presented in percentages only, as the multiple linguistic versions of the test differed slightly in the number of scored items. These differences preclude any meaningful raw score comparisons.

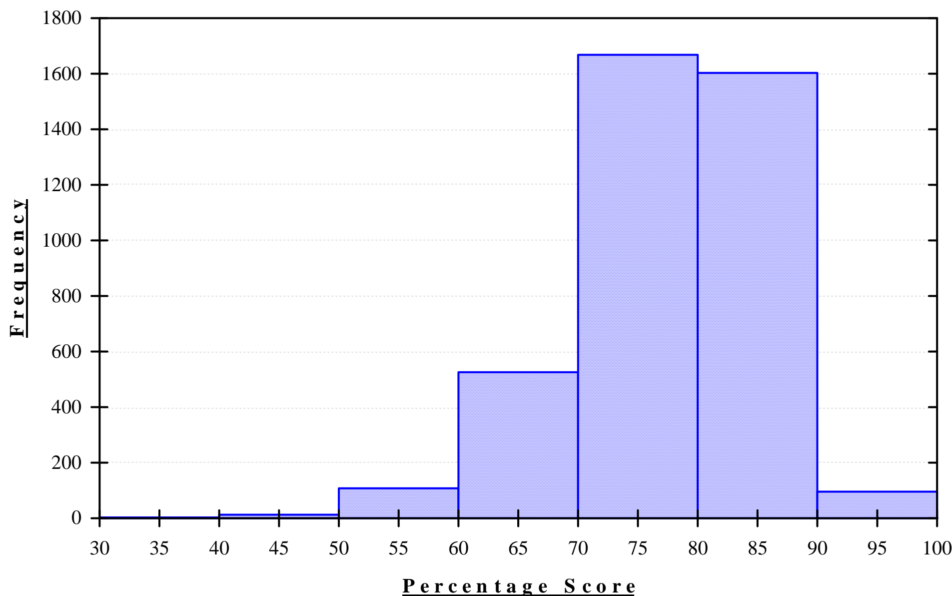
The final statistics are based on the combined written and visual portions. A graph of the total test scores, in percentages, is displayed in Figure 1. The statistics are presented for the overall test rather than for the two subtests because the examination was designed and intended to be *one* comprehensive test.

The statistics are not based on the 200 items that were administered. Some items were deleted from scoring on the basis of a flaw that is typically revealed by a computerized procedure known as an item analysis.

One of the functions of the Board's psychometrician is to "flag" or identify all test items with questionable performance for review by the Examination Review Committee of the Board. These individuals receive a copy of the statistical analysis and hold a meeting with one of the psychometricians.

The Examination Review Committee reviews all flagged items, and items that candidates have critiqued during the test administration. Items determined to have been defective are deleted, and items identified as having been initially miskeyed are rekeyed.

Figure 1: Distribution of Overall Test Scores



Subsequent to this meeting, the examinations are rescored. This review of test and item statistics and candidate critiques, and subsequent rescoring, are quality control procedures routinely performed by most credentialing boards to promote reliability, validity, and fairness.

Table 2: Summary Statistics of Overall Test Scores*

Descriptive Data	
Score Range	34-97
Mean Score**	77.91
Median	79.00
Standard Deviation	7.98
Reliability Data***	
KR-20 Reliability	0.87
Kappa Reliability	0.55
Standard Error of Measurement	
of all scores	2.61
at the pass-fail cutoff score	4.77
Pass-Fail Data	
Pass-Fail Score	68
Candidates Passing	89.25
Candidates Failing	10.75

** Statistics are based on percentage scores for all 4,017 candidates.*

*** Subtest means are 76.39 for the text-only portion, and 80.45 for the image-based portion.*

**** Reliability data are based on the 186-scored-item version of the test.*

Additional quality control procedures are applied to assess the adequacy of the translated versions of the test. These quantitative and qualitative procedures are designed to identify any items for which the translation was inadequate, and resulted in a significant performance decrement. When this occurrence is

confirmed, the item is deleted selectively. This procedure is applied *only* to the affected linguistic version. For the 2009 examination, 14 items were deleted from the examination, overall, for psychometric reasons. Due to linguistic flaws, additional items were deleted from several of the translated versions.

The examination score data display the score range, the mean and median scores, and the standard deviation, in percentage scores. The score range indicates the lowest and highest scores on each test, the mean is the arithmetic average, and the median is the middle score in the distribution, or 50th percentile, determined by placing all test scores in numeric order and selecting the middle score. The standard deviation describes the variability of the test score distribution. Approximately 68% of the test scores lie between plus-and-minus one standard deviation from the mean, and approximately 95% lie between plus-and-minus two standard deviations.

The reliability data contain four statistics. The KR-20 reliability statistic quantifies the internal consistency or replicability of the test results. In short, if the same test could be administered to the same candidates under the same conditions, a high level of agreement would be expected for the results to be considered replicated (i.e., reliable). The KR-20 statistic provides an estimate of this hypothetical situation.

The kappa index is similar to the KR-20 statistic. Both statistics quantify the degree of agreement that results from the hypothetical readministration of the same test to the same candidates. However, while KR-20 reliability is concerned with agreement among scores, kappa focuses on the agreement between the resulting pass-fail decisions. Kappa is interpreted in a manner similar to KR-20, although kappa is typically lower than KR-20 for the same test results.

The standard error of measurement is used to identify the score range within which a candidate's "true score" would be if the test had a perfect level of reliability. Two types of standard errors are reported: one for all scores and one at the pass-fail cutoff score.

The last portion of the statistical table displays the pass-fail data for the examination, which includes the raw and percentage pass-fail score, and the number (raw) and percentage of candidates passing and failing the examination. The pass-fail cutoff score was determined using a form of the Nedelsky procedure, a criterion-referenced technique. Applying this procedure, a performance standard or index is determined for each test item (known as a minimum pass index, or MPI)

on the basis of its perceived level of difficulty. The examination pass-fail standard is then computed as the average of each of the item performance indices.

The unique consideration of the Nedelsky technique is that it treats item difficulty as a function of the degree to which the wrong answers (i.e., distractors), approximate the correct response, and are therefore too difficult for the minimally competent candidate to eliminate. Items with a greater number of these difficult or "sophisticated" distractors are considered to be more difficult to answer correctly and thus, have a lower standard or performance expectation than items determined to be easier on the basis of having fewer, if any, "sophisticated" distractors. This standard setting methodology is applied by the Board *before* the test is administered, thus eliminating "grading on the curve" or inappropriate competition among candidates for acceptable scores.

A study of this process was conducted for the initial administration of the IBLCE examination and indicated that the standard setting technique yields a valid pass-fail score. With the exception of the few items with a relatively low standard for which candidate performance was relatively high, the anticipated relationship between item standard and item performance was attained. Specifically, as the number of sophisticated distractors in an item increased, the percentage of correct responses decreased. Thus, the Board's *a priori* assessment of item difficulty was deemed to be a valid basis for setting performance standards. This analysis is replicated for each examination and continues to support the findings of the cutoff score validation study.

For this 25th administration of the IBLCE examination, the overall rounded pass-fail cutoff score was 68%. As the pass-fail scores are determined on an item-by-item basis, the deletion of additional items in the translated versions of the test does not necessarily affect the percentage pass-fail score. Certainly, the additional deleted items do not make the pass-fail cutoff score more difficult to attain. This is because when items are deleted from computing the candidate scores, the deleted item MPIs are deleted from the passing-score computation also. The rounded pass-fail cutoff score for all translated versions of the 2009 test was the same (i.e., 68%), despite variance in the number of items.

The pass-fail cutoff score of 68% was in the high end of its historic range. The rounded mean score of 78% was equal to the mean of two of the prior three examinations. However, as the pass-fail cutoff score was higher than the cutoff of recent examinations, the rounded pass rate of 89% was lower than the pass rate of

recent examinations. This is the first examination in the past eight, inclusive, in which the pass rate was below 90%, although it was only nominally lower.

Routine equating analyses were conducted to evaluate the consistency of the test's difficulty and cutoff score relative to prior administrations. Equating analyses were particularly important, since the cutoff score was relatively high and the pass rate was lower than that of recent examinations. The results indicated that the test was slightly easier in comparison with prior tests, which led to the cutoff score being relatively high; however, candidate preparedness was at a lower level than that of prior candidate populations. The pass rate decrement was primarily the result of the lower level of candidate preparedness.

Candidates whose overall score was at or above the pass-fail cutoff score received the IBCLC credential as an International Board Certified Lactation Consultant if they were taking the test for initial certification. If they were vying for recertification, a passing score allowed them to retain their IBCLC status. Any candidates whose overall score was below the cutoff score are eligible for re-examination; however, if they were recertification candidates, their certification status was terminated.

All candidates, regardless of whether they passed or failed the examination, received a supplementary diagnostic performance report that indicated their number of correct responses for each discipline and chronological period. For failing candidates, this report is useful in identifying subject matter strengths and weaknesses, which may be particularly valuable in preparing for a subsequent examination. For passing candidates, this report may identify subject matter areas where continuing education is likely to be most useful.

The aggregate performance for each content discipline and chronological period is shown in Tables 3 and 4, respectively. These tables indicate the number of items scored for each of the disciplines and periods, and the average percentage of correct responses.

For the content disciplines, the highest performance level was in discipline E (Maternal and Infant Pathology), with a mean score of 82.4%. The lowest performance level was in discipline I (Interpretation of Research), with a mean score of 67.0%. For the chronological periods, the highest performance level was in period 2 (Prenatal), with a mean score of 87.4%. Period 1 (Preconception) had the lowest performance level, with a mean score of 69.4%.

Table 3: Aggregate Performance on Content Disciplines

Discipline	*Number of Items Scored**	*Mean % of Items Correct*
A. Maternal and Infant Anatomy	21	80.3
B. Maternal and Infant Normal Physiology and Endocrinology	29	74.4
C. Maternal and Infant Normal Nutrition and Biochemistry	13	79.4
D. Maternal and Infant Immunology and Infectious Disease	12	70.2
E. Maternal and Infant Pathology	24	82.4
F. Maternal and Infant Pharmacology and Toxicology	14	75.4
G. Psychology, Sociology, and Anthropology	11	80.7
H. Growth Parameters and Developmental Milestones	12	81.3
I. Interpretation of Research	6	67.0
J. Ethical and Legal Issues	6	69.0
K. Breastfeeding Equipment and Technology	10	82.4
L. Techniques	24	81.0
M. Public Health	4	71.9

* *Based on items deleted selectively from scoring on translated versions because of linguistic flaws, the number of items within some disciplines is lower for some translated versions of the test.*

Table 4: Aggregate Performance on Chronological Periods

Chronological Period	*Number of Items Scored**	*Mean % of Items Correct*
1. Preconception	4	69.4
2. Prenatal	14	87.4
3. Labor/Birth (Perinatal)	14	81.2
4. Prematurity	11	83.9
5. 0-2 Days	24	83.3
6. 3-14 Days	20	79.2
7. 15-28 Days	19	72.7
8. 1-3 Months	14	80.8
9. 4-6 Months	14	77.4
10. 7-12 Months	6	76.1
11. Beyond 12 Months	6	72.0
12. General Principles	40	71.5

* *Based on items deleted selectively from scoring on translated versions because of linguistic flaws, the number of items within some chronological periods is lower for some translated versions of the test.*

Table 5 displays the candidate means and pass rates based on candidate certification status. In this table, performance is compared for candidates taking the test for initial certification, and for recertification at 5-, 10-, 15-, and 20-year periods.

The performance of the candidates sitting for recertification was excellent. Each of the four recertification groups exhibited a higher mean and pass rate than the candidates sitting for initial certification. In addition, there was a linear progression in the mean performance among three of the four recertification groups. There was little distinction between the mean performance of candidates sitting for recertification at 15- and 20-year intervals; candidates in both groups exhibited excellent performance.

Generally, progressively longer recertification periods represent candidates with a longer career as a lactation consultant, and a commitment to continued competence and credential maintenance. Of the total 829 recertification candidates, all but 12 passed (98.6%); this replicates a pattern of superior performance by recertification candidates on prior examinations. The comparative means scores are depicted graphically in Figure 2.

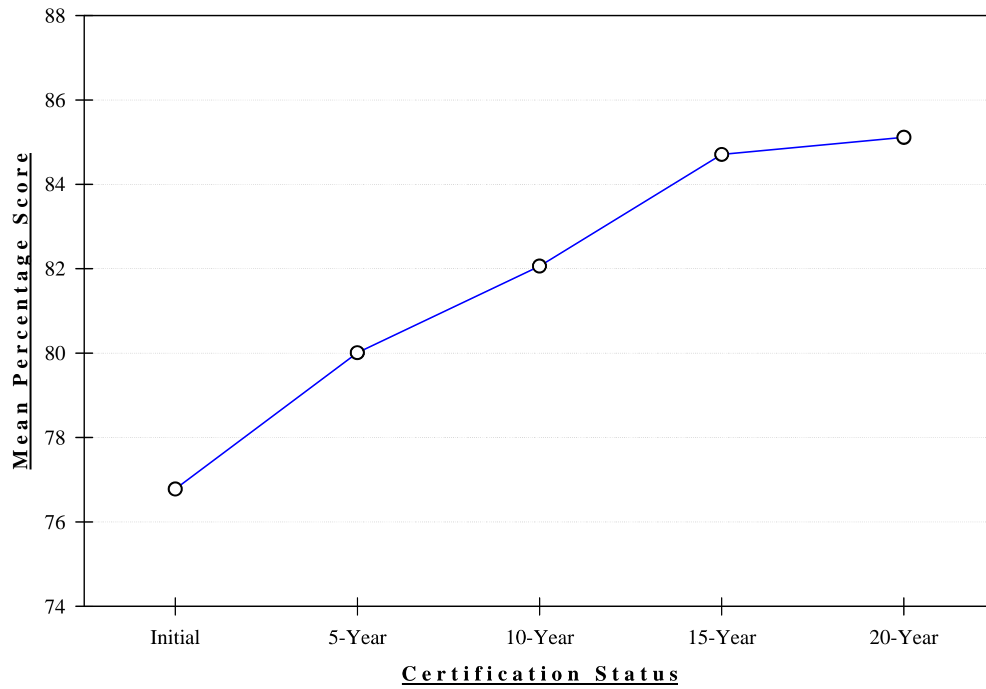
Table 5: Comparative Performance by Certification Status

Candidate Certification Status	*Number of Candidates*	*Mean % Correct*	*% Pass Rate*
Initial Certification	3188	76.8	86.8
Recertification by Exam at 5 Years	102	80.0	97.1
Recertification by Exam at 10 Years	604	82.1	98.5
Recertification by Exam at 15 Years	55	84.7	100.0
Recertification by Exam at 20 Years	68	85.1	100.0
Recertification Total	**829**	**82.2**	**98.6**

The 25th administration of the IBLCE examination for certifying lactation consultants was very successful, and the Board congratulates all candidates who sat for the examination, both for initial and continued certification. Regardless of whether they passed or failed, these candidates had the courage and fortitude to accept the challenge of the examination. The Board is also grateful to the members

of the Examination Committee and to the many professionals who contributed test items for the examination.

Figure 2: Distribution of Scores by Certification Status



The next administration of the IBLCE examination will be in July 2010.