

STATISTICAL REPORT OF THE 2008 IBLCE EXAMINATION

Leon J. Gross, Ph.D.
Psychometric Consultant

In July 2008, the International Board of Lactation Consultant Examiners (IBLCE), administered its 24th annual credentialing examination in lactation consulting. The test was administered in 13 languages to 3,323 candidates in 152 locations across 37 countries and territories representing 5 continents.

This candidate population for the 2008 administration was the largest in the test's history, 4% higher than the prior record in 2006, and 13% higher than the prior year's population. The population was also larger than the first 8 test administrations combined, and the eighth consecutive administration with more than 2,000 candidates.

The 2008 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 seven of the eight most recent years, including 2008, 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 2008 administration is the ninth consecutive year in which candidates from these other countries have accounted for more than 30% of the candidates, and the second 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. This was the fourth administration in which candidates recertified by examination at the 20-year interval. Nearly 25% 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. In short, the

2008 marked a continuation and acceleration of important trends in the candidate population composition.

Table 1: Number of Test Centers and Candidates

<i>Region and Country</i>	<i>Number of Test Centers</i>	<i>Number of Candidates</i>
North America		
Canada	14	170
United States	59	1372
South America		
Brazil	5	28
Peru	1	15
Europe		
Austria	2	43
Belgium	1	38
Denmark	1	21
France	2	61
Germany	7	162
Greece	1	6
Hungary	1	12
Ireland	1	13
Italy	3	23
Netherlands	1	46
Norway	1	2
Poland	1	14
Portugal	1	7
Slovenia	1	12
Spain	1	13
Switzerland	1	65
United Kingdom	3	42
Africa		
Rwanda	1	15
South Africa	2	11
Middle East		
Egypt	1	85
Israel	1	34
Kuwait	1	7
Saudi Arabia	1	3
United Arab Emirates	3	113
Australasia		
Australia	20	300
China	1	18
Indonesia	1	42
Japan	2	147
Korea	4	328
Malaysia	1	2
New Zealand	3	28
Singapore	1	13
Taiwan	1	12

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

A total of 1075 candidates sat for one of the 12 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 for whom English is a second language where the test was not translated into their primary language. The IBLCE examination has now been administered in 17 languages (Arabic, Dutch, English, French, German, Hebrew, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Polish, Portuguese, Slovenian, Spanish, and Swedish) in more than 40 countries across all major continents, offering unparalleled geographical, cultural, and linguistic access.

The 2008 administration was the 18th 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. Of the 3323 candidates, 771 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 in the *Candidate Guide*.

The 2008 examination was developed according to IBLCE standard operating procedures. A 7-person examination committee met for four days to prepare, review, edit, and select test items. The Examination Committee includes representation from 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 12 languages. This marks the 22nd 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. These conditions existed for the following languages and test administrations.

- Arabic: 1994-95
- Dutch: 1993-08
- French: 1992-93, 1997-08
- German: 1987, 1989-91, 1993-08
- Hebrew: 2007
- Hungarian: 2006, 2008
- Icelandic: 2001-03
- Indonesian: 2008
- Italian: 1996-99, 2001-04, 2006-08
- Japanese: 2001-02; 2004, 2006-08
- Korean: 1999-08
- Polish: 2004-08
- Portuguese: 1998, 2000, 2002, 2004, 2006, 2008
- Slovenian: 2006-08
- Spanish: 1987, 1994-95, 1997, 1999-08
- Swedish: 2003

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.

Test Administration

The world-wide administration of the IBLCE examination is a complex process. Although culturally adapted versions of the test were deemed to be not necessary in 2008 as in some prior administrations, 12 translated versions of the test were offered, in addition to a linguistic adaptation (i.e., British English).

Examination Results

The results of the 24th administration of the IBLCE examination are displayed in Table 2. These statistics are based on the total population of 3323 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.

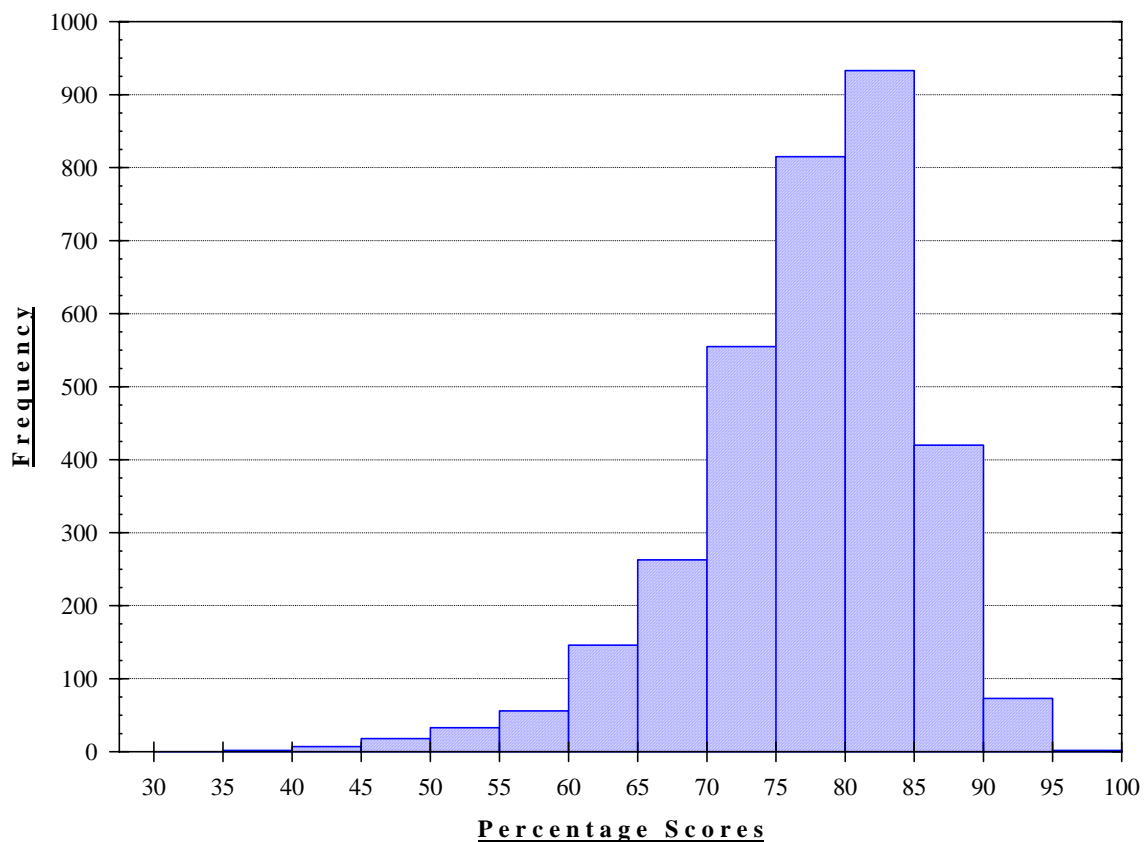
The Examination Review Committee reviews all flagged items, and items that candidates have critiqued during the test administration. Items determined to have been defective are deleted, and items identified as having been initially miskeyed are rekeyed.

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.

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 2008 examination, nine items were deleted from the examination, overall, for psychometric reasons. Due to linguistic flaws, additional items were deleted from several of the translated versions.

Figure 1: Distribution of Overall Test Scores



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.

Table 2: Summary Statistics of Overall Test Scores*

Descriptive Data

Score Range	39-96
Mean Score**	77.87
Median Score	79.00
Standard Deviation	8.21

Reliability Data***

KR-20 Reliability	0.86
Kappa Reliability	0.52
Standard Error of Measurement	
Of all scores	2.53
At the pass-fail cutoff score	5.55

Pass-Fail Data

Pass-Fail Score	65
Candidates Passing	93.56
Candidates Failing	6.44

* Statistics are based on percentage scores for all 3,323 candidates.

** Subtest means are 76.21 for the didactic portion, and 80.89 for the visual portion.

*** Reliability data are based on the 191-scored-item version of the test.

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 ongoing analysis continues to replicate the findings of the cutoff score validation study.

For this 24th administration of the IBLCE examination, the overall rounded pass-fail cutoff score was 65%. 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 2008 test was the same (i.e., 65%), despite variance in the number of items.

The pass-fail cutoff score of 65% was in the middle of its historic range, and nearly identical to the cutoff score of the 2005 and 2006 examinations. The rounded mean score of 78% was several percentage points below the mean of the 2007 examination, but was equal to the third-highest mean in the test's history. The rounded pass rate of 94% was comparably in the upper end of its historic range. This is attributable primarily to the relatively high mean score.

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 mean score and pass rate were near record high levels. The results indicated that the test was approximately average in difficulty in comparison with prior tests; however, candidate preparedness was at a higher level than that of prior populations. The high pass rate was primarily the result of the superior level of candidate ability, which is attributable primarily to more rigorous eligibility requirements than existed during the early years of the program, and a relatively high percentage of recertification candidates. These candidates typically perform better than candidates sitting for initial certification.

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.

Table 3: Aggregate Performance on Content Disciplines

<i>Discipline</i>	<i>Number of Items Scored*</i>	<i>Mean % of Items Correct</i>
A. Maternal and Infant Anatomy	27	77.7
B. Maternal and Infant Normal Physiology and Endocrinology	29	75.4
C. Maternal and Infant Normal Nutrition and Biochemistry	11	73.5
D. Maternal and Infant Immunology and Infectious Disease	12	71.6
E. Maternal and Infant Pathology	20	77.9
F. Maternal and Infant Pharmacology and Toxicology	13	71.5
G. Psychology, Sociology, and Anthropology	11	81.8
H. Growth Parameters and Developmental Milestones	16	81.8
I. Interpretation of Research	8	63.5
J. Ethical and Legal Issues	6	68.0
K. Breastfeeding Equipment and Technology	12	79.2
L. Techniques	20	83.2
M. Public Health	6	75.8

* 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.

For the content disciplines, the highest performance level was in discipline L (Techniques), with a mean score of 83.2%. The lowest performance level was in disciplines I (Interpretation of Research), with a mean score of 63.5%. For the chronological periods, the highest performance level was in period 4 (Prematurity), with a mean score of 87.0%. Period 1 (Preconception) had the lowest performance level, with a mean score of 65.0%.

Table 4: Aggregate Performance on Chronological Periods

<i>Chronological Period</i>	<i>Number of Items Scored*</i>	<i>Mean % of Items Correct</i>
1. Preconception	3	65.0
2. Prenatal	16	82.6
3. Labor/Birth (Perinatal)	12	84.0
4. Prematurity	13	87.0
5. 0-2 Days	25	82.6
6. 3-14 Days	21	80.9
7. 15-28 Days	23	74.9
8. 1-3 Months	14	70.2
9. 4-6 Months	15	78.4
10. 7-12 Months	5	77.6
11. Beyond 12 Months	2	76.4
12. General Principles	42	72.2

* 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.

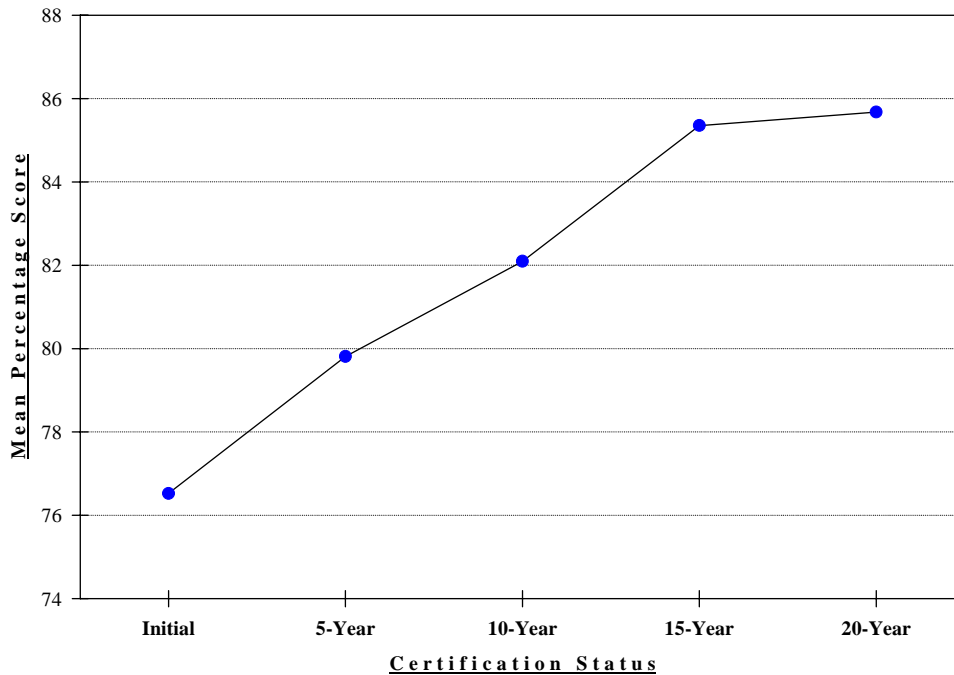
The performance of the candidates sitting for recertification was excellent. Each of the four recertification groups exhibited a higher mean and pass rate than the candidates sitting for initial certification. In addition, there was a linear progression in the mean performance among the three of the four recertification groups. There was little distinction between the 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 771 recertification candidates, all but four passed (99.5%); 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

<i>Type (Number) of Candidates</i>	<i>Mean % Correct</i>	<i>% Pass Rate</i>
Initial Certification (2,552)	76.5	91.8
Recertification by Exam at 5 Years (96)	79.8	99.0
Recertification by Exam at 10 Years (564)	82.1	99.5
Recertification by Exam at 15 Years (48)	85.4	100.0
Recertification by Exam at 20 Years (63)	85.7	100.0
Recertification Total (771)	82.3	99.5

Figure 2: Distribution of Scores by Certification Status



The 24th 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.

The next administration of the IBLCE examination will be in July 2009.