



The logo for ExAC, featuring the letters 'ExAC' in a bold, sans-serif font. The 'x' is highlighted in a vibrant green color, while the other letters are black. A thin vertical green line is positioned to the left of the 'E'.

Examination for Architects in Canada  
Examen des architectes du Canada

A black and white photograph of a modern architectural interior, showing a staircase with glass railings and a large, open-plan space with concrete walls and ceiling.

TECHNICAL AND  
STATISTICAL  
REPORT FOR  
THE ExAC  
2023

A photograph of a modern office interior, featuring a desk, a chair, and large windows. The image is overlaid with a semi-transparent green filter.



- 01 Objective and Overview of this Report / page 3
- 02 Designing and Developing the ExAC / page 5
- 03 Administration of the Tests and Participation / page 9
- 04 Scoring / page 11
- 05 Equating, scaling and setting cut scores / page 21
- 06 Publication of the results / page 23
- 07 Validity Arguments / page 24
- References / page 32
- Glossary of Terms / page 33

# TABLE OF CONTENTS



















In contrast, “modifications”, which are not allowed, are changes made to the content of the tests and to the performance criteria. Modifications are not permitted because they would compromise the validity and reliability of the test scores.

It is permitted to provide the interns with clarifications concerning the instructions before a test has begun, but it is not permitted to provide clarifications concerning the items while a test is in progress (e.g. a reformulation or explanation).

4QFDJBMWFSTJPOTPGUIFFYBNJOBUIJPOTGPSJOUFSOTXIPBWFCCFOHSBOUFEBOBDDPNNPEBUJPO

The CExAC can provide the following accommodations to meet special needs:

- test papers with large print;
- additional time;
- unobstructed access to the testing site;
- access to snacks, for medical reasons, during the testing period;
- access to a private room;
- access to a transcriber.

2VBMJUPOUSPM

The CExAC has established quality-control procedures to ensure the uniformity and fairness of the tests administered across Canada as well as the validity and reliability of the data produced. CExAC has adopted a certain number of procedures so that the persons involved will have confidence in the validity and reliability of the scores that are awarded.

- Data analysis: The CExAC undertakes several types of statistical analysis of data taken from the answers given by the interns so that it can detect the presence of irregularities and verify the integrity of the data.
- Verification of the test material: Following each test, the CExAC attempts to determine if there were any possible irregularities during its administration at the testing sites or during the scoring phase. To achieve this, the consultants and the administrators visually inspect the test materials at different phases of the administration and scoring of the ExAC.

**Note** Since the ExAC serves a certifying role as an examination for admission to the profession, it is necessary to optimize the quality of the results obtained and to ensure that the examination conforms to the desired measurement characteristics. To this end, an analysis of the measurement properties of each of the sections of the ExAC and of each individual item is performed. Following this analysis, certain items that are not consistent with the others or that are found to be otherwise inadequate are excluded from the compilation of the overall scores.

The Committee for the ExAC restates its intention to offer an examination of the highest quality and, consequently, additional quality-control measures have been introduced in order to avoid errors in the test materials.

## SCORING

# 04

The CExAC items can be divided into three categories according to the type of answer that is required. There are multiple-choice items, short-answer items, and multiple-choice items which require a constructed response. Each category requires a different correcting and scoring method. Multiple choice answers are read and marked by an optical mark reader. The data thus collected is automatically recorded in a database. Insofar as the short-answer items are concerned, the CExAC follows rigorous scoring procedures to insure the validity and reliability of the test scores. Trained markers score all of the answers to the short-answer items.

The item-specific scoring rubrics and the exemplars are the principal tools which are used to score the short-answer items. Each answer provided by an intern to a short-answer item is scored in accordance with the best match to one of the score points outlined in the item-specific scoring rubric. The scoring rubrics provided by the CExAC present a working description of the different possible levels of performance by the interns. The exemplars illustrate the descriptors for each score point in a scoring rubric. The exemplars are selected and validated by architects from several licensing authorities during the selection process for the exemplars. The markers are trained to refer constantly to the exemplars in order to ensure uniformity in the scoring. The markers match the work of the interns to the descriptors for the score points shown in the scoring rubric and the exemplars which illustrate the work of the interns for the score points.

The principal steps in the scoring process are described below.

### THE SELECTION PROCESS FOR THE EXEMPLARS

The selection of the exemplars permits the range of acceptable performance levels for each score point in a scoring rubric to be defined so that a consensus can be reached on the scoring of the answers given by the interns. This range is used to train the markers. The process takes place in two phases: the pre-selection of the exemplars by two consultants and the selection of the exemplars during the training of the markers.

#### The selection phase for the exemplars

The selection of exemplars:

- involves the participation of consultants in the preparation of scoring tools of high quality and the training material for the markers, in order to ensure exactness, reliability and uniformity in the scoring of the short-answer items in the ExAC tests;
- requires representative examples of the answers from the interns which permit the definition and illustration of the range of performance levels of the interns corresponding to the scoring rubrics;
- provides a clear consensus for the coding of the answers provided by the interns which are used to train the markers of the short-answer items.

The members of the correction team, which consists of approximately twelve architects (the number of markers can vary from 12 to 16 from one year to the next, depending on the number of examinees) from several licensing authorities, meet prior to the scoring to prepare recommendations for the answers from the interns that will be used as exemplars during the scoring. The number of persons registering for the examination has an impact on the number of markers required for the session.

The qualifications required for markers are as follows:

- expertise and experience in architecture;
- the ability to explain clearly and concisely the reasons why an answer provided by an intern has been placed at a given performance level within a scoring rubric.

The markers:

- refer to the item-specific scoring rubrics and use their scoring expertise to determine the score to be awarded to an intern;
- collaborate with other markers, under the supervision of consultants, in order to arrive at a consensus on the appropriate scoring codes for each answer provided by an intern;
- prepare recommendations for refining the item-specific scoring rubrics;
- formulate the list of reasons which support the scoring codes that have been assigned.

### **Overview of the selection process for the exemplars**

1. Markers are recruited and selected.
2. The consultants to the CExAC conduct meetings of the markers.
  - In order to ensure productivity and efficiency during the meetings, the markers are divided into two groups, anglophone and francophone. Before the scoring process begins, the two groups meet together to share information.
  - Each group discusses a set of items and tasks as well as the corresponding item-specific scoring rubric and determines, using scores, the appropriate answers to be used.
  - The discussions cover the following points:
    - the content and requirements for each item or task;
    - the group consensus on the scores assigned to the answers provided by the interns;
    - the scoring rules, where needed, to ensure uniform scoring for each item or task.

### **PREPARATION OF THE TRAINING MATERIAL FOR THE SCORING: EXEMPLARS, TRAINING COPIES, QUALIFYING TEST, VALIDATION COPIES**

The ExAC consultants prepare the material to be used to train the markers for the field-test of the short-answer items. They take into account the recommendations and decisions concerning the scoring that were made during the selection process for the exemplars and come to an irrevocable decision concerning the answers from the interns which will be used for the exemplars, the training of the markers, the qualifying tests and the assessment of the validity and reliability of the scoring .

The training material includes the following elements:

- exemplars which represent rigorously the score points in the scoring rubrics;
- training copies illustrating answers that correspond clearly to a score point as well as answers that correspond less clearly (e.g. answers that are shorter than normal, unusual approaches, a mix of good and poor answers, and writing that is difficult to read);
- justifications for each exemplar and each training copy used;
- answers which correspond clearly to a score point on at least one of the qualifying tests;
- documents used for continuing education during the daily activities related to standard setting;
- documents used to assess validity.

The consultants will take the necessary steps to ensure that there are sufficient copies of the material to permit the training of the markers.

### **SCORING OF THE SHORT-ANSWER ITEMS**

The CExAC has adopted stringent policies and procedures for the scoring of the items and tasks of the test in order to ensure the reliability of the test scores.

Architects from several licensing authorities score the tests. The CExAC recruits as many markers as possible (who are members in good standing of their respective licensing authority) and supplements this team with consultants who hold pertinent degrees and have substantial experience in large-scale testing.

### **Scoring rooms for the short-answer items**

All of the items are scored in a large room which is usually equipped with two groups of large tables. One group is used by the French-language team and the other is used by the English-language team. In 2012, an experiment was performed in which one large group of tables was used. In 2013 the marking session returned to the model of using separate table groups. Each scoring table group is under the supervision of a consultant. The consultants supervise the training, the scoring and the additional training of the markers. All of the markers are trained to use the CExAC scoring guide (scoring rubrics and exemplars) for each item that they have to score. Following the training, the markers must pass a qualifying test. Real-time monitoring of the validity and reliability of the scoring is performed at the scoring centre and additional training is provided if needed.

The markers perform the scoring individually. Each test paper is scored at least twice by two different markers; if the scores which have been awarded do not agree exactly, the paper is scored a third time by a third marker. The markers may discuss unusual answers with their designated consultant.

Each item is scored separately by at least two trained markers who use the same scoring rubric. An anonymous scoring procedure is used (i.e. a marker does not know what score was awarded by the other marker). The routing process ensures that the answers are read by two different markers answers. If the two markers award the same score (i.e. if there is perfect agreement between them), this score is awarded to the intern. If the two scores differ, whether closely or not, the copy is scored by a third marker, and the average of the scores is awarded to the intern. Some copies are corrected more than 3 times when the disparity between the scores of the first two markers is significant. This stringency enables the interns and the licensing authorities to be confident in the validity of the scores awarded to the interns.

### **Training for the scoring of the short-answer items**

The training is intended to develop a clear and shared understanding of the scoring material so that all of the consultants and markers interpret and apply the scoring material in the same manner; the goal is to obtain a level of scoring of the answers provided by the interns which is reliable, uniform, valid and precise, no matter which marker participated in the scoring.

### **TRAINING FOR THE SCORING OF THE SHORT-ANSWER ITEMS**

The consultants hold diplomas in measurement and evaluation and have proven experience in large-scale testing. They must encourage the markers to reject pre-conceived ideas concerning scoring procedures and to harmonize their thinking and judgment with the scoring procedures and material for the items to be scored. The consultants have the following responsibilities:

- training the markers;
- supervising the scoring of the items;
- ensuring that the scoring material is applied in a uniform manner;
- resolving problems arising during the scoring;
- answering questions from the markers.

In addition, the consultants are responsible for reviewing and analyzing the daily data reports to ensure that the scoring in their scoring rooms is performed to a high standard of quality. The consultants are responsible for the scoring material, the management of the room and any problems which arise during the scoring.

### **Training the markers for the scoring of the short-answer items**

The training of the markers for the scoring of the short-answer items is intended to develop specialists in the scoring of specific items or of subsets of items. Each item of the test requires a complete set of scoring materials, namely: item-specific scoring rubrics; exemplars (answers from interns illustrating each point on the scale) and the justifications for them; training copies; qualifying test and several validation copies.

In order to obtain high levels of validity and reliability, the CExAC has established stringent criteria for all of its selection, training and qualifying procedures. Several other quality-control procedures (described below) are used during the scoring process in order to identify markers who require additional training or who may eventually need to be removed from the scoring process.

The CExAC has devised scoring procedures in order to ensure uniformity. The training is intended to develop a clear and shared understanding of the scoring material so that every consultant and marker interprets and applies the scoring material in the same manner; the goal is to obtain a level of scoring for the answers provided by the interns which is reliable, uniform, valid and precise, no matter which marker scored the answers. The markers are all trained to perform scoring with the same scoring material. During the training, the markers are told that they may need to make adjustments with respect to the scoring of the performance of the interns so that they can accept the standards and practices of the CExAC for the scoring of the tests.

The training of the markers on a limited number of short-answer items takes about half a day and covers the following points:

- an overview of the tests to be scored (in particular, the purpose of the tests and the specific characteristics of the tests and of the population to be evaluated);
- general instructions concerning the security, the confidentiality and the pertinence of the scoring material;
- instructions on how to enter the scores onto the forms used to collect the scoring data

**Note** It should be noted that, since the 2012 ExAC, an electronic system has been used to record the scores. The identification numbers of the interns are coded in accordance with Code 39 barcode symbology. An electronic barcode reader is used to read and enter the identification number of the intern and the marker on a computerized form. The scores are then entered on this form by the marker. When the marker has finished scoring a test paper, the data is sent automatically to a central data base.

- a detailed review and discussion of the scoring material for each item to be scored ( the item, the item-specific scoring rubric, the exemplars and the justifications for them); the scoring material is approved by the CExAC and cannot be altered by the markers once the scoring process has begun:
  - an emphasis is placed on providing the marker with an understanding of the way in which answers differ incrementally in quality and the manner in which each answer reflects the description of the score point in the scoring rubric;
  - the exemplars are answers which are typical (as opposed to unusual or uncommon), clear (as opposed to controversial or borderline) and correct (i.e. they cannot be changed by anyone);
- the scoring of a set of training copies, consisting of a selection of answers from interns which have been competently scored:
  - sets of training copies that generally contain clear answers as well as uncommon and less- clear ones (e.g. answers which are shorter than normal, use an atypical approach, contain a mix of good and poor answers, or are presented in writing which is difficult to read);
  - the markers score the answers first, and then they compare their scores with good scores with justifications;
  - Next, the markers analyze the justifications and the characteristics of each correct answer with their consultant;
  - during the process, the markers must internalize each item-specific scoring rubric and adjust their individual scoring to conform to the scoring rubric which has been provided to them.

The markers also receive training which encourages them:

- to read the entire answer before coming to any decisions with respect to the scoring;
- to view each answer as a whole, without paying special attention to details (such as spelling);
- to remain objective and fair and to view the answer through the filter of the scoring rubric;
- to avoid allowing themselves to be influenced by other corrected copies (Halo effect);
- to score all of the answers in the same way and not to adjust the scoring to take into account a particular characteristic affecting one intern;
- to always give the benefit of the doubt to the intern in the event of a doubt or a borderline situation.



## EXCEPTIONAL PROCEDURES DURING THE SCORING

### Indications of inappropriate content or of interference by staff persons during in the administration of the ExAC

Occasionally the answers of the interns to short-answer items contain content which is inappropriate or offensive or has different handwriting or indicates possible interference by staff persons involved in the administration of the ExAC or some other problem. The copies which indicate such problems are set aside once they have been scored. They are evaluated by members of the CExAC, who examine the incident reports and consult with the appropriate jurisdiction as needed.

### Damaged or misprinted test papers

When the test papers are distributed to the interns, it is possible that these may be torn, poorly stapled or missing pages. In such cases, the interns will not be penalized. The material is replaced at the time of administration. If these damaged papers are discovered during the scoring session, members of the CExAC will examine the damaged papers and determine the procedure for scoring these papers. Severe penalties will be imposed in the event that an intern attempts to remove test papers from the room.

In regards to printing errors in the papers, see the explanatory note in Chapter 3 concerning the management of printing errors.

## DAILY ADDITIONAL TRAINING

The consultants frequently provide explanations concerning the scoring of specific items and of key elements of the item-specific scoring rubrics in order to ensure that everyone applies the material exactly and uniformly, from one day to the next, as well as before and after the breaks.

### Daily morning review of the exemplars

The consultants begin each day with a partial or complete review of the scoring rubrics and the exemplars. The purpose of this review is to reorient the markers and to highlight each section of the scoring rubrics which requires special attention.

### Daily activities to establish a scoring standard

Without prior notice, the markers will be asked to score copies that they themselves have already scored or copies that have received perfect agreement from other markers. This procedure serves to ensure uniformity in the scoring among the markers over time and among themselves.

## REPORTS TO MONITOR THE QUALITY OF THE SCORING OF THE SHORT-ANSWER ITEMS

On a daily basis, the consultants review all of the data concerning the productivity of the members of their team and the degree of agreement among them (the rate of perfect agreement between the scores obtained by an intern from two different markers).

During the scoring, the markers must maintain a certain level of precision in the validation copies. The minimum required level of perfect agreement to be maintained is as follows: 75% perfect agreement between at least two markers. The third correction, which is performed when there is not perfect agreement between the first two markers, contributes significantly to the reliability of the scoring process.

Markers who do not meet the required standards for the precise degree of agreement are given additional training, which includes an attentive review of the exemplars. If this additional training does not correct the situation, it is possible that the scores awarded by these markers will not be used and that a fourth scoring of the copies in question will be required.

### Cumulative reports on the mean score and on the distribution of the scores

The cumulative data for the mean score and the distribution of the scores are used to monitor for deviation in the scoring of a marker. These data reports permit the identification of potential problems that need to be taken into consideration during subsequent analyses. The data are used to confirm the validity of the scores. The report indicates and summarizes (by item or group of items, scoring rubrics and marker) the cumulative mean score and the distribution of the scores awarded by the markers.

### Cumulative reports on reliability

During the scoring, the consultants monitor reliability and uniformity by using an anonymous reinsertion (unidentified) of test papers. All of the short-answer items are sent for a second scoring, which permits a rating of the reliability among markers. The reports indicate and summarize the daily and cumulative levels of agreement among markers for two independent scorings which include perfect or close agreement. The reports provide summaries by item or by group of items, group, scoring rubrics and marker. The markers are ranked according to their level of reliability.

### Cumulative reports on productivity

These reports rank the markers according to their productivity level (from low to high) and name the markers who have not attained the minimum level of productivity. The consultants review the data which is highlighted in this report in order to determine whether additional training is required for a specific marker.

### Reports of aggregate cumulative data for each marker

These reports combine validity data with secondary data on each marker. The reports include not only daily and cumulative data on validity but also data on daily and cumulative reliability, the mean score and productivity data. This report ranks markers according to their level of validity (from low to high). Markers who do not meet the standard of 75% perfect agreement for the three-point scoring rubrics, 70% for the four-point scoring rubrics, and 60% for the six-point scoring rubrics are highlighted in these reports.

### RELIABILITY AND VALIDITY OF THE MARKERS

The procedures which are used to assess the reliability and validity of the ExAC tests are summarized below.

#### Reliability of the markers

The reliability of the tests is compromised by different sources of measurement error. In the case of the short-answer items, the absence of consistency in the scoring is the main source of errors. The percentage of agreement between the scores awarded by a pair of markers is identified by the term interrater reliability. The percentage of perfect agreement is used as an index of interrater reliability.

The process used to determine interrater reliability for the ExAC does not require the recirculation of selected test papers because all of the answers provided by the interns are scored by at least two markers. The interrater reliability for the ExAC is reflected in the scores awarded by two markers (or three, where applicable) and is measured by the rate of perfect agreement between at least two markers.

#### Reliability estimates for the short-answer items on the ExAC 2023

The percentages of perfect agreement between the first two markers for Section 1 vary between 70.9% and 85.8% for the English-language and French-language versions. Overall, the percentages of perfect agreement for the items in Section 1 of the English-language and French-language versions exceed the CExAC objective.

The percentages of perfect agreement between the first two markers for Section 3 vary between 70.7% and 90.5% for the English-language and French-language versions except for one item. The percentages of perfect agreement for the items in Section 3 of the English-language and French-language versions exceed the CExAC objective.

When the scores from two markers for an intern's answer do not agree 100%, the answer is scored once again by a third marker. The markers do not know the scores attributed by the other markers, nor do they know how many times a particular copy has been scored.

## EQUATING, SCALING AND SETTING CUT SCORES

# 05

### EQUATING

For reasons of security, the CExAC creates different tests each year while still ensuring that the content and the statistical characteristics are comparable to those of the tests used in previous years. Despite significant efforts undertaken to ensure similarity, the tests may differ to some extent in difficulty from one year to another. To determine if this has occurred, the CExAC makes use of a technique which is known as equating; this permits an adjustment to be made as a function of the differing degree of difficulty in the tests from one year to another. Equating ensures that interns in one year are not unjustly advantaged or disadvantaged relative to those of another year, and that any variations measured in performance levels are the result of differences in the performance of the interns rather than of differences in the difficulty of the tests.

We can make the assumption that there is a linear relationship between the annual editions of the ExAC since the table of specifications, the types of items and the patterns of answers remain relatively constant from one year to another. That being the case, methods of linear test equating are indicated. In order to determine the multiplication factor used to perform the linear transformation, a sample of items is repeated and the mean scores of these items are tracked from one year to another. This method makes it possible to verify whether the fluctuations of the mean scores of an intern are due to the average degree of difficulty of the tests or whether they are due to changes in the average performance of a cohort of interns from one year to another.

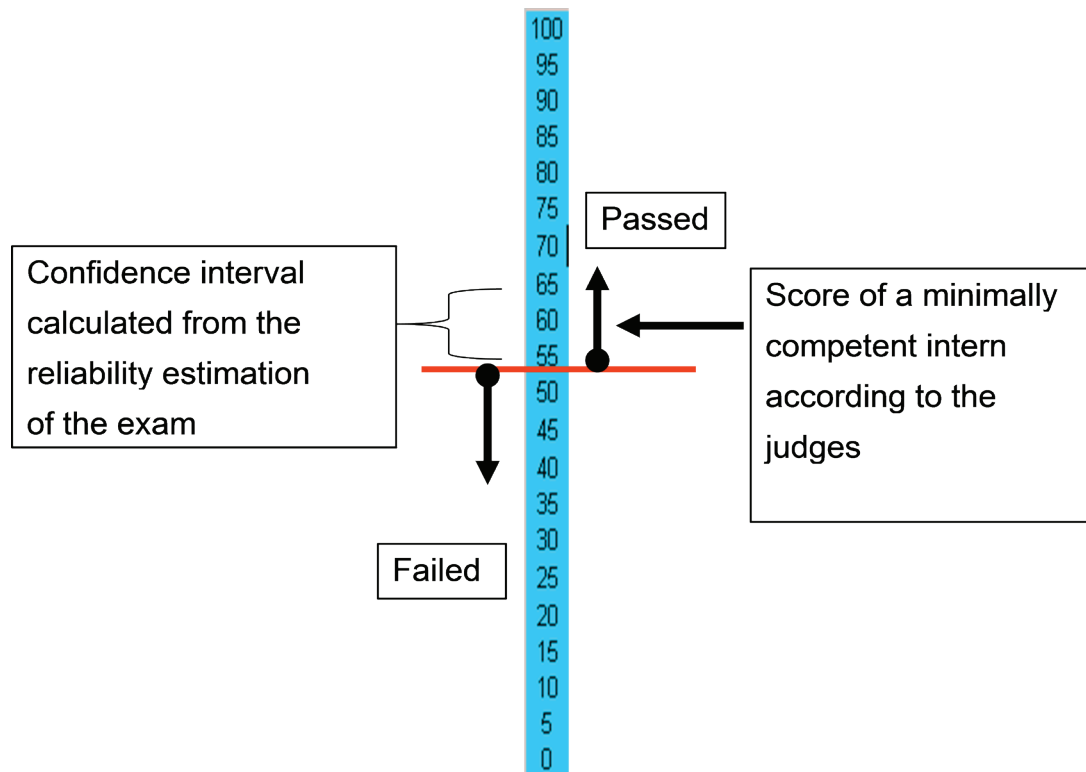
### SCALING

Since the architectural profession and the criteria for training may evolve over time, it is possible that the CExAC may eventually make modifications to the objectives measured by the various ExAC tests. Therefore, the content and duration of the CExAC tests may also be changed to reflect modifications made to the objectives. Should this occur, the modified tests will be different, both in terms of their content and in terms of their statistical characteristics, from the tests given in previous years. In such cases, the CExAC will use a process called scaling, in order to establish a relationship between the tests of previous years and the modified tests of the subsequent year. The processes used for equating and scaling are similar, but the intended effects are different. Equating permits an adjustment to be made according to the differences in difficulty between tests that are similar in content and statistical characteristics. When necessary, scaling is used to establish a relationship between two tests that have different content and statistical characteristics.

### SETTING CUT SCORES

The process for setting cut scores consists of a series of steps which permit the determination of the passing score for each test. The process is begun by asking several expert markers (architects who have been trained as markers during a scoring session) to examine each item on the tests and to estimate the likelihood of success of an intern with a minimum level of competence. The average of the estimates of the markers for all of the tests constitutes the minimum score that must be obtained on each test by an intern with a minimum level of competence. In addition, items which are repeated from one year to the next are tracked to ensure that the cut scores remain equivalent from one year to the next. Since the tests are measuring instruments with an imperfect level of reliability, an interval of confidence which is situated around the limiting score estimated by the judges is calculated. This gives the benefit of the doubt to interns who have an obvious minimum level of competence. An intern who fails a test will have received a score which was lower than the limiting score determined by the judges, and this same result is also lower than the lowest interval of confidence that could be considered as a margin of error. Figure 5.1 illustrates this concept.

Figure 5.1  
Cut scores



## SUCCESS RATES FOR THE 2023 ExAC

Sixteen markers participated in the scoring session for the 2023 ExAC. Several of the markers are bilingual. The table below shows the success rate for each section. In all, of the 851 interns who participated in the 2023 ExAC, 627 of them passed all of the sections for which they were eligible; thus, 73.6% of the interns were successful in all sections. The overall success rate for all of the 2023 ExAC tests was 85.2%.

Table 5.1

Combined success rates for the French-language and English-language versions

Section 1	Section 2	Section 3	Section 4
643/755	659/772	651/764	657/772
85.2%	85.4%	85.2%	85.1%

## PUBLICATION OF THE RESULTS

06

The results of the CExAC tests are published by the language of the intern, by province or territory, and for all of Canada. The results describe the performance of the interns at this crucial step in their training. The stakes are very high for the interns since passing the ExAC (except in the event that an equivalence has been granted) is an essential condition for gaining access to the profession. The results obtained by the interns are also crucial to the licensing authorities from an economic point of view, and they will serve as a guide in their decision-making process. For all of these reasons, a high degree of confidentiality is maintained in order to protect the integrity of the ExAC and the privacy of the personal information of the interns. The CExAC publishes annually, in both official languages, both this general report and a summary version of it. These reports are public and are available on the ExAC website: [www.exac.ca](http://www.exac.ca) within a few months following the administration of the ExAC. Tables also provide data concerning the overall success rates.

The licensing authorities receive data files which provide detailed results of the performance of each intern who is registered in their province or territory. The registrar of each licensing authority transmits the results to the interns in a performance report. This document indicates the status of the intern, the sections of the ExAC that were passed, the equivalences that were awarded (where applicable) and the sections that need to be taken again. The report does not include raw scores. The results are indicated in pass or fail form.

Where a section must be taken again, the report provides constructive information on the portions of the test that were less-successfully completed. The intern can then take the test again the next time that the ExAC is offered. The tests that were passed do not have to be rewritten. It is important, however, to communicate with the pertinent licensing authority in order to obtain information about exceptions which are linked to time limitations.

## REQUESTS FOR REVIEW

An intern may request a review. The precise rules which apply to this process are contained in the guide. Once a request has been made, two consultants perform a multi-point review of the bubble answer sheets. The short-answer items are scored once again by two experienced markers. The data bases containing the intern data are checked in order to verify the integrity of the data. Every year, about 20 to 40 interns request a review after receiving their results.

## VALIDITY ARGUMENTS

# 07

### INTRODUCTION

The previous chapters of this report contain information concerning the validity of the Examination for Architects in Canada (ExAC), in that they deal with several topics related to the construction and delivery of the examination: the preparation of the examination and its components, the administration of the ExAC sessions, scoring, equating and scaling, item analysis, reliability, results, publication and delivery of individual scores, and requests for review. The purpose of this chapter is the document the validity of the ExAC by setting out the arguments contained in the previous chapters and by providing additional pertinent information.

### THE MANDATE OF THE CExAC

In order to fulfill its mandate, the Committee for the ExAC (CExAC) offers the ExAC once a year. The ExAC consists of four tests (sections) which cover 13 themes and 160 specific objectives. The four tests, each of which lasts a maximum of three hours, are offered at the same time throughout Canada, during the fall, over a period of two days. Interns may take the ExAC in either of the two official languages of Canada.

The ExAC is intended for interns who have met all of the necessary requirements and who wish to gain access to the architectural profession. The ExAC is an examination for admission which is recognized by the licensing authorities which oversee the architectural profession in most of the Canadian provinces and territories. Successful completion of the ExAC is one of the criteria for ensuring the competence of architects. It is the last step which needs to be taken by interns before they can gain access to the profession and obtain the right to practice as an architect.

The licensing authorities which oversee the architectural profession in Canada have a legal mandate to ensure the protection of the public by controlling access to and the practice of the profession within their respective province or territory. In order to provide architectural interns with an examination which is best suited to the context of the practice of architecture in Canada, the licensing authorities in most of the provinces and territories agreed to develop and administer an examination for admission which is referred to as the ExAC.

Each year, detailed results are provided to each intern who has taken the ExAC. The results are also compiled in aggregate form for each province and territory and for each of the official languages. The reports which are made public present only the overall results for all of Canada and for each official language. In several of the provinces and territories, fewer than 16 interns take the ExAC each year. It is therefore necessary to ensure identity protection in the results in order to prevent the identification of the interns and to avoid the temptation to make inappropriate comparisons between the licensing authorities.

#### THE CONCEPT OF VALIDITY

In 1999 (revised in 2014), the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME) established principles and standards for testing (Standards for Educational and Psychological Testing). A French-language adaptation of this document, entitled *Normes de pratique du testing en psychologie et en éducation*, was published in 2003 by the Institut de recherches psychologiques (IRP). In these publications, it is stated that the term “validity” refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests (page 9). The term “validation” can be viewed as developing a scientifically sound validity argument to support the intended interpretation of test scores and their relevance to the proposed one (page 9). Validation is the joint responsibility of the test developer and the test user (page 11). Kane (2006) adds that the test developer must demonstrate the validity of the interpretations and the uses which result from the test (page 17).

The previous references (AERA, APA and NCME, 1999; IRP, 2003; Kane, 2006) provide information concerning the sources of data which should be considered during the preparation of validity arguments. These data sources cover the content of tests and their consequences, as well as methods of answering and interactions with other tests. Even though the sources of the data which constitute an ensemble are distinct, they are not considered as different types of validity. The validation process must take into account the type of test and the available data which is pertinent to the technical and practical aspects of the test.



## VALIDITY ARGUMENTS BASED ON THE CONTENT OF THE ExAC

### Characteristics of the ExAC Tests

The four ExAC tests measure the degree of success with which the interns attain or exceed the minimal level of competence for certain expectations and certain learning content areas (13 themes). Globally, these content areas reflect the spheres of the architectural profession which can be measured by the selected evaluation method: programming, site and environmental analysis, coordination of engineering systems, cost management, National Building Code, schematic design, design development, final project, bidding and contract negotiations, construction phase (office), construction phase (site), management of the project & business/practice management and sustainable design literacy. The principal sources of content for the ExAC are: the Internship in Architecture Programme, the Canadian Handbook of Practice for Architects and the National Building Code of Canada.

Tables of specification are used in the preparation of the items in order to ensure that the number and type of items remain uniform from year to year. These tables of specification define the concept to be measured by each test, determine the type of response mechanism for each one and present the distribution of the items in the targeted content areas. The content areas covered by the ExAC tests are limited to those which can be measured by written examinations.

### Preparation of the test items

The CExAC ensures that the test items are pertinent to the content areas and to the practice of the architectural profession by using the following methods in their preparation:

- obtaining the participation of architects in the preparation and vetting of the items;
- ensuring that these architects come from different regions of Canada;
- subjecting all of the items to testing and retaining those which offer the desired metric and metrological characteristics.

The CExAC recruits and trains experienced architects from various regions of Canada to develop and vet the items. Each year, development and vetting sessions are held in various cities in either French or English. Each development session may take place over a six-day period. Each vetting session may last two days. From four to eight experienced architects and at least one consultant participate in a typical development or vetting session.

The architects who participate in these sessions are selected because of their expert knowledge and their experience in the field, their knowledge of laws and regulations, their skill and experience in the use of scoring rubrics, their skill in written communications and their experience in the preparation of contracts, plans and specifications. Training is provided at the beginning of each session for new item developers. An item development guide has been prepared by one of the consultants for this purpose. At the end of each development session, the items are assembled and the architects provide comments on the items prepared by their colleagues. This procedure enables the detection of errors in the items and permits an initial selection to be made. Later on, all of the items are vetted by the members of the CExAC.

The members of the CExAC and two or three consultants review all of the new items which are to be incorporated into one or another of the ExAC tests during the year. The members of the CExAC examine the items one at a time in order to ensure that they conform to the development guidelines, that they do not contain errors in the stems or in the distractors, that they do not demonstrate any regional distinctiveness or include any inappropriate references to a subgroup and that there is no unessential information in them. The role of the CExAC is to ensure that the items in the ExAC properly measure the expectations outlined in the general and specific objectives.

The new items which are incorporated into the ExAC for the current year can be considered to be in a field-testing phase. Those items which demonstrate sound and fair metric properties will be retained and will be included in the overall score of the interns. Where the opposite occurs, those items will not be counted in the overall score of the interns, and they will be re-evaluated and revised by the various development and vetting committees. The revised items will be retained in the database and may be used in a future version of the ExAC. Each year, about 10% of the new items are not included in the overall score because they lack the required metric properties.

#### Fairness and uniformity in the examination

The CExAC has adopted quality-control procedures for fairness and uniformity in the organization of the ExAC examination sessions. These procedures are also intended to ensure precision in the scoring. Supervisors are required to be present at each site where the examination is being administered to the interns. A system of instantaneous communication between administration sites has been established to permit unforeseen circumstances such as printing errors to be managed. Verification procedures have been put in place to ensure that the examination papers are of the highest quality. A guide describes in detail what the supervisors and administrators of the tests must do, including:

- the procedures to be followed (e.g. the preparation of the materials which are to be distributed to the interns, directives to be followed during the testing phase);
- the specific accommodations and provisions which are permitted for interns who have submitted a request for an accommodation;
- directives to be given to the interns (e.g. providing initial instructions for the four tests);
- the professional responsibilities of the personnel involved in the testing phase.

The guide specifies the physical arrangement of the rooms and tables. The supervisors note the position of each intern within the room. They must provide the CExAC with an incident report for every incident which occurs. The CExAC deals with each of these incidents individually and takes the necessary steps when an investigation is required. Thus far, the CExAC has been required to undertake two or three more-or-less intensive verifications each year.

### Scoring the items

The ExAC items are divided into three categories, according to the type of answer which is required. The first category is multiple-choice items. The answers for the multiple-choice items are read and scored by an optical mark reader. Several analyses are performed in order to ensure the exactitude of the optical reading.

The second category consists of short-answer items, and the third category consists of multiple-choice questions which also require a short answer. For the short-answer questions, the CExAC follows rigorous scoring procedures in order to ensure the validity and the reliability of the scores. All of the answers to the short-answer items are scored by markers (architects) who have been trained for this purpose.

Item-specific scoring rubrics and exemplars are the main tools which are used to score the short-answer items. The scoring rubrics for the ExAC provide a working description of the different possible levels of performance by the intern. The exemplars are samples of intern responses which illustrate the descriptors for each score point in the scoring rubric. The exemplars are selected and validated by architects at the beginning of the scoring process. The markers are trained to refer constantly to the exemplars in order to ensure uniformity in the scoring. The markers practice on several copies and contend with different scenarios before beginning the actual scoring.

In order to ensure uniform scoring, the test belonging to each intern is first scored by two different markers with neither marker knowing the score awarded by the other. Then the consultants verify the scores which were awarded by the two markers for each item on the intern's copy. If the scores awarded by the two markers do not match perfectly, the items are scored a third time. The copies may be scored by more than three markers when the disparity between scores is pronounced. The verification process for uniform scoring also enables the detection of any possible deviation by markers who tend to score significantly less or more severely than their colleagues. In such cases, the consultants prescribe additional training. The items which are being subjected to field testing are scored using the same scoring requirements as those for the regular test items in order to ensure accurate and uniform marking for all of the items.

The percentages of perfect agreement for most of the short-answer items in the English-language and French-language versions of the tests exceeded the objectives set by the CExAC. As previously mentioned, when the scores awarded by two markers are not in 100% agreement, the answer is scored by a third marker.

## Equating

Equating is a procedure which is used to ensure that the results of the ExAC are comparable from one year to another. It could be hypothesized that a linear relationship exists between the annual versions of the ExAC to the extent that the table of specifications, the types of item and the answering patterns remain relatively constant from one year to the next. Since this is the case, the use of linear equating methods is indicated. In order to determine the multiplying factor which needs to be applied in order to perform the transformation, a sample of common items is repeated and the average scores obtained for these items are tracked from one year to the next in order to establish a common scale. This method also makes it possible to verify whether fluctuations in the average scores of the interns from one year to another are the result of the average level of difficulty of the tests or of variations in the average level of performance of a cohort of interns.

### VALIDITY ARGUMENTS BASED ON THE METRIC PROPERTIES OF THE ExAC

#### Technical quality of the examination

The method used to analyze the technical quality of the ExAC conforms to classical test theory (CTT). This theory was first stated in 1904, notably in the work of Spearman. CTT evolved considerably thereafter and was formalized during the 1960s. More modern testing theories and measurement models were developed since then, including item response theory (IRT). Theories such as IRT offer considerable advantages with respect to the technical analysis of items and tests. However, these theories are demanding with respect to the assumptions and to certain conditions of use. For example, to calibrate items as a function of a few parameters, a considerable sample of cases (i.e. interns) is needed; unfortunately, the annual number of participants in the ExAC is insufficient to establish these parameters. We have therefore opted for CTT, since it is less demanding in terms of the size of the sample. It should be noted that CTT is a valid theory which is well-known and still widely used today.

Several parameters are considered when choosing the items to be included in the overall scores of the interns. The first parameter to be considered is the difficulty ( $p$ ) of the item. The degree of the difficulty ( $p$ ) ranges from 0 to 1; the items which are closest to 0 are the most difficult and those which are closest to 1 are the easiest. An attempt is made to retain items which are between 0.20 and 0.95. For example, an item which has a degree of difficulty ( $p$ ) of 1 is an item which everyone has answered correctly, and this is not very useful in evaluating the performance of an intern. Such an item would therefore not be included in the total scores of the interns, and it would be revised before it could be included in a subsequent version of the ExAC.

The coefficient of discrimination ( $D$ ) must also be considered. The discrimination of an item refers to the property of this item to be able to distinguish between interns with a higher level of performance and those with a lower level of performance. For example, an intern who has a higher level of overall performance on a test should have a greater chance of performing well on a particular item than would an intern who has a lower level of overall performance. The parameter  $D$  ranges from -1 to 1. Items with a coefficient which is close to or lower than 0 are not good discriminators. Items with a coefficient higher than 0.20 are generally retained because these are considered to be good discriminators.

In order to ensure that a test is metrically consistent, the correlation between the results for an item and the total score on the test is examined. This involves determining whether all of the items measure the same thing or the same field. An item which correlates poorly with the others may measure the knowledge of a fact which has nothing to do with the architectural profession. Generally, the items which have an item/total correlation above 0.20 are retained.

The parameters of difficulty (p), discrimination (D) and item/total correlation each contribute, in their own way, to the accuracy and internal consistency of a test. Optimizing techniques are used to identify items which show weaknesses with respect to one or another of these parameters. Such items are withdrawn and subjected to revision in order to eliminate these weaknesses, and they are not included in the overall scores of the interns.

Since the outcome of the examination has important consequences, precise measurements are performed to monitor its progress and to provide a confidence interval which ensures that all of the interns who should have been successful were, in fact, successful. Cronbach's alpha, a conservative indicator of reliability, is calculated for each test and for each language. For the 2023 ExAC, the reliability coefficients noted below were obtained. While the ExAC has not been in existence for very long, its reliability coefficients have increased steadily since its initial version.

Cronbach's alpha ( $\alpha$ )	Section 1	Section 2	Section 3	Section 4
English-language version	0.880	0.895	0.900	0.915
French-language version	0.854	0.879	0.885	0.885

Other analyses, such as standard error measurement and differential test functioning, are also performed. A follow-up of interns who have obtained very low scores is undertaken in order to determine if some subgroup of interns is at a disadvantage for linguistic or other reasons.

As mentioned above, several factors contribute to the preciseness of the ExAC: the quality of the items, the accuracy and uniformity of the scoring, and the correlation among the items. All of the items used in the ExAC are directly linked to the general and specific objectives. The CExAC selects items of suitable difficulty which permit interns with a higher level of performance to be distinguished from those with a lower level. Several procedures have been put in place to contribute to the accuracy and uniformity of the scoring.

## VALIDITY ARGUMENTS BASED ON THE INTERPRETATION OF THE RESULTS

### Setting cut scores

The process for setting cut scores consists of a series of steps that enable the passing score for each test to be determined. It is begun by asking several markers who have received training during the scoring session to examine each test item and to estimate the likelihood of success of an intern with a minimum level of competence. The average of the estimates of the markers for all of the tests constitutes the minimum score that must be attained on each test by an intern with a minimum level of competence. This is a procedure which conforms to that devised by Angoff. Another procedure is also used to track a group of repeated items to ensure that the cut scores remain constant from one year to another. Next, since tests are measuring instruments with an imperfect level of reliability, an interval of confidence is calculated around the cut score determined by the markers. By proceeding in this manner, the benefit of the doubt is given to interns with an apparent minimum level of competence. Thus, an intern who fails a test will have received a score which was lower than the limiting score determined by the markers and which was lower than the lowest confidence interval (cut-off point) that could be considered as a margin of error. For the past few years, we have observed that the cut-off point tends to be situated around 50 % (i.e. a score of 50% of the points obtained from among the items which were retained).

### Publication

The results of the ExAC are published by the language of the interns (French and English), for both the provinces and the territories, and for all of Canada. The results describe the performance of the interns.

A high level of confidentiality is maintained in order to protect the integrity of the ExAC and the privacy of the personal information of the interns. The CExAC publishes annually, in the two official languages, both this general report and an executive summary of it. The results are public and are published on the ExAC website ([www.exac.ca](http://www.exac.ca)) a few months following the administration of the examination. Tables are also provided to show the aggregate success rates.

The licensing authorities receive detailed results of the performance of each intern registered in their province or territory. The registrar of each licensing authority transmits the results to the interns in a performance report. This report indicates the status of the intern, the ExAC tests which were passed, the equivalences which were awarded (where applicable), and the tests which need to be taken again. The report does not include raw scores. The results are indicated in pass or fail form.

Where a test must be taken again, the report provides constructive information concerning the portions of this test which were less-successfully completed. The intern can take the test again during the next presentation of the ExAC. The tests which were passed do not have to be rewritten. However, the intern must communicate with the licensing authority in order to obtain information concerning any applicable time limits and the number of times that it is permitted to rewrite a test. The intern has up to three chances to pass the four tests; thereafter, all four tests, including the ones which have already been passed, must be rewritten.

### Requests for a review

An intern may request a review. Once a request has been made, the consultant performs a multi-point review of the bubble answer sheets. The short-answer items are scored again by two experienced markers. The data concerning the interns are tracked in order to verify their integrity. Every year, about 20 to 40 interns take advantage of this option.

### REFERENCES

The American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education (1999). Standards for Educational and Psychological Testing. Washington, DC, American Educational Research Association.

Institute of Psychological Research (2003). Normes de pratique du testing en psychologie et en éducation. French adaptation of Standards for Educational and Psychological Testing by the AERA, the APA and the NCME. Montreal.

Kane, M.T. (2006). "Validation" in R. L. Brennan (Ed.), Educational Measurement, Fourth Edition. Westport, CT. American Council on Education and Praeger Publishers, pp. 17-64.

## GLOSSARY OF TERMS FOR THE ExAC DOCUMENTS

Term	Explanation
Accommodation	A support mechanism and a service provided on request to interns who have special needs.
Aggregates	Depersonalized results which are classified by subgroup or by region.
Bias of an item or of a test	A distinctive feature of an item which favours or hinders individuals as a consequence of their belonging to a particular subgroup. For example, an item which makes reference to the construction of a skating rink might put Congolese at a disadvantage with respect to Canadians.
Bubble answer sheet	A standard answer sheet which allows an intern to record the answers to multiple-choice items. The sheets are rigid and can be read by an optical mark reader. The scores are transferred automatically to a data base management system.
Certification or admissions test	A test which is intended to certify an individual in accordance with well-defined criteria. Since the stakes are very high, the accuracy, the pertinence and the processes surrounding the design, the administration, the scoring and the interpretation of the scores must be given a high priority. The ExAC is a certification test.
CExAC	The Examination for Architects in Canada (ExAC) Committee.
Constructed-response item	A synonym for short-answer item.
Consultants	Specialists in testing. These specialists are not architects.
Correlation coefficient	Correlation refers to the strength of the relationship between items. The stronger the relationship between two items, the more the test forms a coherent whole. For example, if interns who are successful on item 22 are also successful on item 34, there is a positive correlation between the two items.
Diagnostic test	Bias which favours or hinders a group of individuals because of their membership in a subgroup.
Differential item functioning	Bias that favors or disadvantages a group of individuals based on their group of belonging.
Discrimination coefficient	An item provides good discrimination if it is answered correctly by interns who are successful on the ExAC as a whole and if it is answered incorrectly by interns who perform poorly on the ExAC as a whole.
Distractors	The false answers in a multiple-choice item.



Docimology	The science which is concerned with the preparation, administration and interpretation of tests.
Edometrics	The science which is concerned with the preparation, administration and interpretation of tests in education.
Equating	A method used to ensure that the results of tests are comparable over time.
ExAC	The Examination for Architects in Canada.
ExAC administrator	Individual who supervises the preparation, the administration, the resources and the activities associated with the various phases of the ExAC. There are two ExAC administrators.
Examination	In this report, the term “examination” typically refers to the entire ExAC, which is composed of four “sections” or “tests”.
Exemplar	An example of an answer to a question which is typical of a specific performance level.
Field-testing of items	Process whereby items are tested in the field in order to verify their measurement properties and their functioning.
High-stakes test	An expression which is used when the test results have serious consequences for those taking them. Certification tests such as the ExAC are examples of high-stakes tests.
Interrater concordance	This is a comparative index of the results of the correction of the same item by different markers. There is concordance when one marker awards the same score as another marker for the answer of a given intern to a specific question.
Intrater concordance	This is a comparative index of the results of the correction of several items by a single marker. There is concordance when the marker awards the same score at different times of the day for the answer of a given intern to a specific item.
Item	A question on a test. Tasks are associated with items.
Item bank	A collection of archived and recently-developed items which can be selected for use in the preparation of the four sections of the ExAC.
Item calibration	Assigning the appropriate statistical parameters to each item. Each item can be assigned its own degree of difficulty, its discrimination coefficient, and its correlation coefficient with the other items in a test. These are known as measurement characteristics. The decision to include an item on the ExAC is based largely on these statistics.
Item discrimination	The relationship between the number of interns who are successful in answering an item and the number of interns who are successful on the examination as a whole. This relationship is used to evaluate the difficulty of an item or the degree to which an objective was achieved.
Item weighting	The number of points assigned to an item. This indicates the importance of an item with respect to the total score.

Large-scale tests	These are distinguished from class-room tests by the larger numbers which are subjected to the testing. Certification tests such as the ExAC are examples of large-scale tests.
Licensing authority	A legal entity which is responsible for regulating the practice of a profession within a province or territory.
Marker	A synonym for judge and rater.
Measurement properties	Several statistical measurements related to a test and its items (e.g. reliability, validity).
Minimum level of competence	The total score on a test which represents the minimum level which an intern must attain in order to be considered competent.
Multiple-choice item	An item which consists of a question and a choice of answers.
NCARB	The National Council of Architectural Registration Boards produces an American certification examination. The ExAC has provided an alternative to this examination since 2008.
Optical mark reader	A specialized scanning device which is able to read bubble answer sheets.
Psychometrics	The science which is concerned with the preparation, administration and interpretation of tests in psychology and education. The word Edumetrics can also be used.
Rater	A synonym for marker and judge.
Reliability	Reliability is a concept which is associated with measurement accuracy and consistency among the items on a test.
Remediation	A compensatory or corrective strategy for a failure or a weakness in a particular area.
Qualifying test	Exemplars are used to test whether the markers are using the scoring rubrics properly.
Scaling	A method which ensures that test results are comparable in spite of the differences among them.
Score	Refers to the points obtained and to the number of correct answers. Classical Test Theory is used to interpret the answers on a test.
Score point(s)	Point(s) which are awarded to an intern for an answer which corresponds to a descriptor in a scoring rubric.
Scoring centre	During the scoring session, a separate room is provided with computers and other equipment to permit the performance of the calculations and follow-up measures which are needed for the proper functioning of the session.
Scoring rubric	An evaluation grid used during a scoring session for short-answer items. This grid uses descriptors to illustrate the different levels of performance for an item and the number of points (score points) to be awarded to each level.

Setting cut scores	The process of determining and confirming the minimum acceptable level of performance.
Short-answer item	An item which does not offer a selection of answers but which requires that an answer be formulated.
Special needs	A physical or psychological handicap or a temporary or permanent medical condition which could hinder the participation of an intern in the test sessions.
Specific objectives	The items in the ExAC are designed to measure the knowledge and skills of interns with respect to 130 criteria which are considered essential for architects to be judged competent to practice their profession. Example of a specific objective: Explain data from a legal land survey.
Standard setting	The process of determining and confirming the minimum acceptable level of performance.
Stem	The portion of an item which presents the question (excluding the choice of answers).
Subgroup	An identifiable group based on gender, ethnicity or culture.
Table of specifications	A table used during the preparation of items and the construction of a test. The table of specifications clearly indicates all of the content areas to be measured, their relationship to the specific objectives and the items which correspond to each content area.
Taxonomic levels	Refers to a hierarchy of cognitive objectives which characterize test items. Items are intended measure knowledge, comprehension and the capacities to apply concepts, to analyze, to synthesize and to evaluate.
Test supervisor	The individual who oversees the administration of the ExAC. Each licensing authority hires its own test supervisors.
Test version	There are French-language and English-language versions of the ExAC.
Training copies	Exemplars used for training purposes. These are taken from actual tests written by interns.
Validation copies	Exemplars used to verify the performance and the quality of the work of markers.
Validity	Validity is a broad and inclusive concept which refers to the idea of the pertinence of a test and its processes. For example, a test is valid when it measures only what it is intended to measure and nothing else. A test is valid when it is fair.
<a href="http://www.exac.ca">www.exac.ca</a>	The ExAC website





ExAC

Examination for Architects in Canada  
Examen des architectes du Canada

© Committee for the ExAC  
Comité de l'ExAC  
January 2024

