ExAC Examination for Architects in Canada
Examen des architectes du Canada

Statistical and technical report

Examination for Architects in Canada (ExAC 2024)
Winter 2025

Table of content

SUMMARY

This report contains validity and reliability evidence related to the Examination for Architects in Canada (ExAC). Validity and reliability are important considerations for all high-stakes exams. The Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) state that reliability refers to the consistency of scores across replications. Validity refers to the interpretations and decisions made from the test scores, and supporting evidence of several contributing factors such as fairness, security, feasibility, content, relevancy, and others.

The objective of this report is to comply with the Standards by providing “test users with the information needed to help them assess the nature and quality of the test, the resulting scores, and the interpretations based on the scores.”

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: Author.

INTRODUCTION

This report contains validity and reliability evidence related to the Examination for Architects in Canada (ExAC). Validity and reliability are important considerations for all high-stakes exams. The Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) state that reliability refers to the consistency of scores across replications. Validity refers to the interpretations and decisions made from the test scores, and supporting evidence of several contributing factors such as fairness, security, feasibility, content, relevancy, and others.

The term “validation” can be viewed as developing a scientifically sound validity argument to support the intended interpretation of test scores and their relevance. Validation is the joint responsibility of the test developer and the test user. Kane (2006) adds that the test developer must demonstrate the validity of the interpretations and the uses which result from the test.

The 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 consider the type of test and the available data which is pertinent to the technical and practical aspects of the test.

The objective of this report is to comply with Standard 7 (AERA, APA, NCME, 2014) by providing “test users with the information needed to help them assess the nature and quality of the test, the resulting scores, and the interpretations based on the scores.”

PURPOSE OF THE ExAC

Standard 1.1 prescribes that the intentions of an examination be clearly stated. To fulfil its mandate, the ExAC Committee (CExAC) offers the Examination for Architects in Canada (ExAC). The ExAC consists of four (4) tests (sections) which cover thirteen (13) themes, thirty-eight (38) general objectives, and one hundred and fifty (150) specific objectives. The four tests, each of which lasts a maximum of three hours, are offered at the same time throughout Canada over a period of two days. The ExAC is offered once a year in both of Canada's official languages.

The ExAC is intended for candidates who wish to gain access to the architectural profession. It is an examination for admission which is recognized by the licensing authorities which oversee the architectural profession in Canadian provinces and territories. Successful completion of the ExAC is one of the requirements intended to evaluate the competence of architects, and the ExAC is one of the final steps to be taken by candidates who wish to gain access to the profession.

The licensing authorities which oversee the architectural profession in the Canadian provinces and territories have a mandate to protect the public by controlling access to and the practice of the profession within their respective jurisdictions.

HISTORY

In order to provide architectural candidates with an examination that is best suited to the context of the practice of architecture in Canada, the Alberta Association of Architects (AAA), the Ontario Association of Architects (OAA), and the Ordre des architectes du Québec (OAQ) decided to develop a new examination for admission over which they would exercise complete control, in particular with respect to its preparation, updating, and administration. To date, in addition to the AAA, the OAA, and the OAQ, the following licensing authorities recognize and subscribe to the ExAC:

- Architects' Association of Prince Edward Island (AAPEI);
- Architects' Association of New Brunswick (AANB);
- Manitoba Association of Architects (MAA);
- Northwest Territories Association of Architects (NWTAA);
- Architects Licensing Board of Newfoundland and Labrador (ALBNL);
- Nova Scotia Association of Architects (NSAA);
- Saskatchewan Association of Architects (SAA);
- Architectural Institute of British Columbia (AIBC).

Here are links to separate related documents and resources available in both official languages in Canada:

- **ExAC Framework Document - ExAC**
- **ExAC website**
- **Canadian Architectural Certification Board (CACB)**
- **Committee for the Examination for Architects in Canada (CExAC)**
- **ExAC Accommodations**
- **Examination for Architects in Canada (ExAC)**
- **Internship in Architecture Program (IAP)**
- **Preparation Guide**
- **Regulatory Organizations of Architecture in Canada (ROAC)**

TEST STANDARDS FOR HIGH-STAKES EXAMINATION

Validity

The Standards for Educational and Psychological Testing defines validity as follows:

> *Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.*

The validation processes involve arguments and evidence obtained through metrological studies of processes, content, and interpretation of the results. Here are some common concepts:

- Test security (cheating, leaking, confidentiality, destruction of exam material);
- Quality control (test material, processes, data integrity);
- Fairness - Differential functioning between groups of individuals that create an advantage to a group due to the nature of the task, gender, origin, medical issues, language;
- Accessibility and adaptation;
- Test content – Assuring that the content is relevant and measurable (refer to items that are obsolete or irrelevant to the intent behind the test);
- Evaluator training and efficacy of the scoring algorithms and the rubrics;
- Access to proper tools, environment, facilities;
- Efficacy of directives and instructions;
- Transparency;
- Equity.

Many of the concepts are addressed by committees. Some are measured and monitored empirically. The rigorous application of protocols and processes are the key to assuring test validity. Just as for reliability, the continuous improvement cycle strategy (PDCA/PDSA: Shewhart; 1939 & Deming; 1950) is rigorously applied by the CExAC.

This report presents reliability and validity evidence as well as general information.

Content of the ExAC

Standard 11 indicates that licensure examination programs should provide a description of the content with evidence linking it to the profession's requirements. The content in the General and Specific Objectives of the ExAC is founded on the *Internship in Architecture Program* (IAP). The IAP was developed by the Regulatory Organizations of Architecture in Canada (ROAC) with the mandate of providing a structured and effective pathway through internship. The ExAC tests the experience of candidates gained during the IAP to determine if they meet the professional standards required to be an Architect in Canada. The candidates practical experience is supplemented and supported by several key documents including:

- Canadian Handbook of Practice for Architects;
- National Building Code of Canada;
- National Energy Code of Canada for Buildings.

These sources are generally accepted sources of critical knowledge and practical application related to architecture in Canada. The documents are recognized by the Regulatory Organizations of Architecture in Canada (ROAC).

Characteristics of the ExAC Tests

The four sections of the ExAC measure the degree of success with which the candidates attain or exceed the minimal level of competence for certain expectations and practical experience 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, and business/practice management and sustainable design literacy. For more information, please refer to the **Preparation Guide**.

Format of the ExAC

The tests include performance-based activities which require written answers to short-answer items as well as multiple-choice items, all of which allow the candidates to demonstrate their knowledge and aptitudes with respect to the expectations and practical experience content mentioned above. A version of each test is prepared for French-language candidates, and another version is prepared for English-language

candidates. The two versions contain the same number of items and involve the same types of tasks. Part of the item bank is initially developed in French, and the other part is first produced in English. Each group of items is then adapted to the other language. The measurement properties are calculated independently for the French-language and English-language versions of the items. It is not simply a matter of translation but rather of adaptation, since the translated tests may contain biases which favour members of a particular linguistic group. The adaptation allows the test items to be calibrated independently with respect to languages and by sections.

Examination results are provided for each candidate, for each province and territory, and for all of Canada. On behalf of the CExAC, the co-administrators posts aggregate national results on the website and in this report. Detailed reports are provided to each licensing authority. Each candidate receives an individual report of their results from the registrar of their licensing authority. In the event of a failure in at least one of the four tests, the individual report indicates the learning objective(s) requiring remediation.

The members of CExAC rely on the contribution of architects in good standing from all participating jurisdictions for most aspects of the ExAC, especially the development of items and item-specific scoring rubrics, the review of the content of the items, the weighting of the items, and the scoring of the answers provided by the candidates for the short-answer items. The CExAC and the participating architects are always guided by external consultants who are specialized in the preparation and administration of large-scale tests. Thus, the candidates who aspire to join the architectural profession are evaluated by tests designed by the consultants and vetted and marked by architects. Architects who wish to be involved in the CExAC testing process can submit a request to their respective licensing authority. The selection of architects to participate in the activities of the CExAC is made in accordance with the following criteria:

- Geographic location (to maintain a balance among jurisdictions and between languages);
- Private and public practice;
- Number of years of experience (some variety is desirable); and
- Previous work experience with the CExAC.

TEST DESIGN AND DEVELOPMENT

Specifications for the tests

The ExAC examinations are based upon a set of general and specific objectives developed, reviewed, and approved by a selected group of experienced architects from Canada. These objectives define the areas and levels of knowledge used to establish competency in the architectural profession. The general and specific objectives are posted on the ExAC website as part of the Preparation Guide. These objectives are subsequently reviewed periodically by groups of architects from various regions of Canada to check relevance to current practice in Canada.

The table of specifications is largely based on the document Canadian Standard Competency for Architects and developed with the support of consultants. It is used in the preparation of the items to maintain a uniform number and type of items from year to year. The table of specifications defines the concepts to be measured by each test, determines the type of response mechanism for each one, and presents 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.

It should be noted that not all the expectations and learning content areas can be measured during large-scale testing. The measurable expectations and learning content are grouped together by subject and the items are then linked to them. All the measurable expectations and learning content within a group cannot be measured during an individual examination but, over time, they will eventually all be measured. Consequently, it is important that the candidates include all these expectations and learning content in their preparation.

Table of specifications (blueprint)

<table>
<tr><th>Section 1</th><th>Section 3</th></tr>
<tr><td>Approximately 96 multiple-choice and approximately 8 short-answer questions</td><td>Approximately 96 multiple-choice and approximately 6 short-answer questions</td></tr>
<tr><td>Programming (18 items out of 95)</td><td rowspan="3">Final Project (92 items out of 107)</td></tr>
<tr><td>Site and Environmental Analysis (7 items out of 95)</td></tr>
<tr><td>Coordinating Engineering Systems (37 items out of 95)</td></tr>
</table>

Cost Management (19 items out of 95)	Sustainable Design Literacy (15 items out of 107)
Schematic Design (6 items out of 95)	
Design Development (7 items out of 95)	
Section 2	**Section 4**
Approximately 75 multiple-choice questions	Approximately 130 multiple-choice questions
National Building Code of Canada (NBC) National Energy Code of Canada for Buildings (NECB) (75 items out of 75)	Bidding and Contract Negotiations (31 items out of 130
	Construction Phase – Office Functions (41 items out of 130)
	Construction Phase – Field Functions (27 items out of 130)
	Management of the project and business/practice management (31 items out of 130)

Item development

Standard 4

> *Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.*

The following section presents such evidence.

Initial field-testing of the items

An initial field-testing session for the four tests was held in the spring of 2008 in three provinces (Quebec, Ontario, and Alberta). The testing process permitted a rapid determination as to whether the items functioned as intended. Information derived from the answers was used to revise and refine the stems of the multiple-choice items, the short-answer items, the selection of the multiple-choice items, the item-specific scoring rubrics for the short-answer items, and to assess whether the length of the tests were adequate. The results of the field-testing of the items were provided to those responsible for content to assist them in reviewing, revising, and correcting the items.

Each year, new items are developed and vetted so that they can become items on future tests. Architects from various jurisdictions assist the CExAC in all aspects of the preparation of the items, and these include:

- Finding contextual situations and relevant examples that comply with the requirements of the table of specifications which includes the general and specific objectives;
- Drawing the plans and sketches, preparing the multiple-choice and short-answer items, defining tasks, and preparing item-specific scoring rubrics for the short-answer items;
- Evaluating the value and interest of the items by submitting them to other contributors; and
- Reviewing the items and the item-specific scoring rubrics, paying particular attention to content and possible prejudices towards subgroups of individuals.

Item developers

Standard 4.8 indicates that qualifications, characteristics, and training received by the development team are provided. Standard 4.7 indicates the process that is used to develop, review, and try out items.

On behalf of the CExAC, the co-administrators recruit architects, and the consultant trains them to participate in committees of five to fourteen architects. These committees meet several times a year to develop and review the multiple-choice and short-answer items used in the tests.

All items must conform to the expectations and learning content described in the table of specifications. The architects are provided with a copy of the *Guide for the Development of Items for the ExAC* to assist them in developing the multiple-choice and short-answer items and the item-specific scoring rubrics.

The developers of the items for the ExAC tests are selected in accordance with the following criteria:

- Extensive knowledge of and experience in the practice of architecture in Canada at a variety of scales, project types, and locations;
- Good knowledge of the principal sources of content for the examination, including the Intern Architect Program, Canadian Handbook of Practice for Architects, National Building Code of Canada, and the National Energy Code of Canada for Buildings;

- Proven experience in collaborating with others and in accepting instructions and feedback.

Training of item developers

Group training, lasting about 3 hours and dealing with the importance of developing valid and reliable items, congruent with the specific objectives, is provided to item developers before item development sessions begin. Particular attention is paid to the choice of verbs to correspond to taxonomy levels described in Bloom's Taxonomy (Bloom & al., 1956). A copy of the Guide for the Development of Items for the ExAC is provided to those who are responsible for developing. This table defines the objective and the taxonomy level of the cognitive demands which are to be targeted during the item-development process.

The ExAC content evaluation committee

The content evaluation committee consists of about seven architects from several licensing Authorities including members of CExAC. The members meet up two times a year to check that newly developed items correctly evaluate the objectives and the learning content described in the table of specifications under the supervision of consultants. The members also provide knowledgeable advice with respect to the specialized content and the assessment of the quality of the material which is proposed for the tests. In addition, the members make suggestions concerning the inclusion, revision, or exclusion of items. For this purpose, the members of the content evaluation committee must possess the above-mentioned skills and have considerable expertise in the practice of architecture. The committee also examines the items for inappropriate content (sexism, agism, etc.) and regionalism (expression or reference peculiar to a particular region of Canada).

Revisions made by the content evaluation committee

At the end of the item-development session, the consultant and the item developers examine the draft versions of the items and item-specific scoring rubrics to make adjustments before they are presented to the content evaluation committee, composed of members of CExAC. Further revisions are made after the evaluation committee has completed its review and the questions are then subjected to a selection process for field-testing.

For the multiple-choice items, they consider the clarity and the completeness of the stem of an item, the preciseness of the correct answer, the plausibility of the distractors, and the rules which apply to the development of items. For the short-answer items, they consider the correspondence between the items and the item-specific scoring rubrics and determine if the items solicit the expected range of answers The revision checks that items are appropriate for use in all the provinces and territories and that the correct answer remains the same irrespective of the origin of the candidate.

Field testing process for new items

CExAC subjects new items to a field-testing process. Items being field-tested are indistinguishable from the actual test items, so candidates do not know which category of item they are answering, which prevents demotivation that can result when new items are field-tested outside of the normal testing context. Scores from items being subjected to field-testing are not necessarily used to determine the results for the candidates. Subjecting test material to field-testing is part of the process for selecting good quality items from a psychometric and metrological point of view.

Tests assembly: Selecting the items for the tests

The test items consist of approximately 30 % to 40 % items that were not used in the previous ExAC. It is important to note that the items correspond to content groups, but that not all the content in a group is measured in a given test; nonetheless, all the measurable content is eventually included in a test. These are important factors to be considered when selecting items for an examination which is national in scope:

- The measurement parameters for each item demonstrate that the test item is fair and of comparable difficulty to those used in previous tests;
- The items conform to the test specifications, their content is balanced, and there is no differential functioning in the items or in the test; and
- The preparation of the items is guided by a table of specifications which is similar from one year to the next.

A committee examines the proposed assembled tests and verifies if instructions are sufficient and that items do not give the answers to other items. The final assembly of the ExAC is then sent for publishing.

Sample items can be found in the **Preparation Guide**.

EXAM DEVELOPMENT

ExAC calendar cycle

January	Presentation and acceptance of exam results to CExAC.
February	Exam scores are released to the jurisdictions and candidates are advised of their results and the Request for Review period opens.
March-April	The current year's exam is assembled using existing questions from the item bank and approved new items from the previous spring vetting by the CExAC. Request for Review period closes; exam reviews are completed; and results delivered to candidates. The upcoming exam is vetted by the CExAC (late April to mid-May) .
May	Based on the results of the vetting session, the final version of the exam is established and is internally field-tested by independent architects for quality control. New questions for consideration in future exams are written in separate French and English item development sessions (May-June).
June	Registration for the fall writing of the ExAC opens. Final proofreading of the exam, including linguistic review (editorial) and confirmation of print-ready proofs.
July	Registration for the exam closes; a tentative "master list" of registrants is prepared and updated as necessary. Regional jurisdictions forward accommodation and exemption requests for their registrants to the national co-administrators.
August	CExAC reviews accommodation and exemption requests (by virtual meeting) and results are communicated to jurisdictions Recruitment of Test Supervisors and Invigilators begins
September	Jurisdictions provide final confirmation of Test Supervisors and Invigilators. Review of items written in the spring; items are either approved for inclusion in a future exam or are rejected or potentially sent back to be re-worked.
October	Exam booklets and supplementary test materials are shipped to jurisdictions.

November	ExAC is written (normally the first Monday and Tuesday of November). Multiple choice bubble sheets are machine marked (optical mark reader) and short-answer booklets are shipped to one of the co-administrators for human scoring.
December	Short-answer booklets are marked (normally over four days on and/or around the first weekend of December).

The development cycle for continuous improvement

CExAC uses the plan-do-check-act (PDCA /PDSA) approach (Shewhart, 1939; Deming, 1950), a four-step framework for implementing incremental change for the purpose of improvement based on the scientific method of “hypothesis-experiment-evaluation-adjustment”. The ExAC’s development cycle embeds several opportunities for review (checks) to identify redundancies, vulnerabilities, and flaws, and to improve processes moving forward.

ADMINISTRATION OF THE TESTS

ExAC Administration manual

The *ExAC Administration Manual* is developed in French and English and provided to each licensing authority and for each ExAC administration site to describe in detail the procedures for administering the exam.

The procedures described in the manual must be followed by ExAC administrators and supervisors to provide consistency, test security, and fairness in test administration for candidates in Canada. The manual includes:

- The procedures to be followed (e.g., instructions to be followed during the testing phase);
- The specific accommodations and provisions that are permitted for candidates who have submitted a request for an accommodation;
- The instructions to the candidates (e.g. instructions for presenting the four tests);
- The responsibilities of those involved in administering the exams.

Candidates are provided with a quiet space in which to work and must be always supervised. Severe penalties will be imposed if a candidate attempts to remove test papers from the room. It is permitted to provide the candidates 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).

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, as well as provide precision in the scoring.

A system of instantaneous communication among administration sites has been established to permit unforeseen circumstances, such as printing errors, to be managed. The Administration Manual 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 candidates, directives to be followed during the testing phase);

- The specific accommodations and provisions which are permitted for candidates who have submitted a request for an accommodation;
- Directives to be given to the candidates (e.g. providing initial instructions for the four tests);
- The responsibilities of the personnel involved in the testing phase.

The administration manual specifies the physical arrangement of the rooms and tables. The supervisors note the position of each candidate within the room and provide CExAC with an incident report if warranted. CExAC takes investigative steps when determined to be appropriate.

Support provided to candidates who require accommodation

The administration manual provides information and instructions to assist CExAC in accepting accommodations for candidates as recommended by appropriate professionals and indicates the documentation expected to be provided.

Modifications, or changes made to the content of the tests and to the performance criteria, are not permitted because they would compromise the validity and reliability of the test scores.

Definition of an “accommodation”

> *An accommodation is both a support mechanism and a service which permit candidates requiring accommodation to demonstrate their competence with respect to the skills that are evaluated by the tests. The accommodation only modifies the way in which a test is administered or the manner in which the candidate provides answers to parts of the test. It does not modify the content of the tests or compromise their validity or reliability.*

Examples of available accommodations

CExAC can provide the following accommodations to meet an individual’s needs:

- Test material with large print;
- Additional time;
- Unobstructed access to the testing site;
- Access to snacks, for medical reasons, during the testing period;
- Access to a smaller room or separate exam space; and/or

- Access to a transcriber.

Quality control

The consultants hold degrees in measurement and evaluation and have proven experience in high stakes and in large-scale testing. The Committee for the ExAC strives to offer a high-quality examination so has introduced quality-control measures to avoid errors in the test materials, provide uniformity and fairness of the tests administered across Canada, and support the validity and reliability of the data produced. CExAC has adopted procedures so persons involved will have confidence in the validity and reliability of the scores awarded.

- Data analysis: The consultants undertake several types of statistical analysis of data taken from the answers given by the candidates so that it can detect the presence of irregularities and verify the integrity of the data.
- Verification of the test material: Following each test, CExAC attempts to determine if there were any 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. Architects, not involved in the development process, review the assembled tests to detect errors in material, directives, or references to charts, plans, and tables.

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 verify 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, 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.

Professional linguistic reviewers and translators are used.

SCORING

The ExAC items can be divided into two categories according to the type of answer that is required: multiple-choice and short-answer, each requiring a different correcting and scoring method. Multiple choice answers are read and marked by an optical mark reader, automatically recording data in a database. On behalf of the CExAC, the consultant follows rigorous scoring procedures for short answer items to support the validity and reliability of test scores, including scoring by specially trained markers.

The item-specific scoring rubrics and exemplars are the principal tools used to score short-answer items. Scoring rubrics present a working description of the different possible levels of performance by the candidates, and exemplars illustrate the descriptors for each score point. Markers are trained to refer to the exemplars to provide uniformity in scoring, matching the work of the candidates to the descriptors for the score points and the exemplars. Each answer provided by a candidate is scored in accordance with the best match to one of the score points outlined in the item-specific scoring rubric.

In the scoring process, the consultants have the following responsibilities:

- Training markers;
- Supervising the scoring of items;
- Overseeing the application of scoring material in a uniform manner;
- Resolving problems arising during scoring;
- Answering questions from the markers;
- Checking that scoring material cannot be altered by the markers once the scoring process has begun;
- Performing 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 consultants are also responsible for the scoring material, test security, and the management of the room and any problems which arise during the scoring. They review and analyze daily data reports to check that the scoring in their scoring rooms is performed to a high standard of quality.

Similarly, markers have the following responsibilities:

- To read an 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 by viewing each answer through the filter of the scoring rubric;
- To avoid allowing themselves to be influenced by other corrected copies (Halo effect);
- To not to adjust the scoring to consider a particular characteristic affecting one candidate.

Details of the principal steps in the scoring process follow.

Selection process for the exemplars

Exemplars are selected to illustrate the range of acceptable performance levels for each score point in a scoring rubric and support a consensus on the scoring of answers between markers. The process takes place in two phases: the preselection of the exemplars by a consultant 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, to provide exactness, reliability, and uniformity in the scoring of the short-answer items in the ExAC tests;
- Requires representative examples (exemplars) of the answers from the candidates to define and illustrate the range of performance levels of candidates, corresponding to the scoring rubrics; and
- Provides consensus for the coding of answers used to train the markers of the short-answer items.

The members of the scoring team (the number of which can vary from one year to the next, depending on the number of examinees) from several licensing authorities, meet at the beginning of the scoring to prepare and train to use exemplars during the scoring.

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 a candidate 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 a candidate;
- Collaborate with other markers, under the supervision of consultants, to arrive at a consensus on appropriate scores for each answer;
- Prepare recommendations for refining the item-specific scoring rubrics, if appropriate;
- Formulate the list of reasons which support the scoring codes that have been assigned.

Overview of the training process

The consultants conduct meetings with the markers:

- Markers are divided into two groups, anglophone and francophone, to support productivity and efficiency. Before the scoring process begins, the two groups meet to share information;
- Each group discusses a set of items and corresponding scoring rubrics and agreed-upon appropriate answers and related scores;
- The discussions cover include:
  - The content and requirements for each item;
  - The group consensus on the scores assigned to the answers;
  - The scoring rules, where needed, to provide uniform scoring.

PREPARATION OF THE TRAINING MATERIAL FOR THE SCORING

Exemplars, training copies, qualifying test, and validation copies

The ExAC consultant prepares material to be used to train the markers that includes:

- Exemplars which represent the score points in the scoring rubrics;
- Training copies illustrating answers that correspond clearly or less clearly (e.g. answers that are shorter than normal, unusual approaches, a mix of good and poor answers, and /or writing that is difficult to read) to a score point;
- Justifications for each exemplar and each training copy used; and
- Documents used to assess validity.

Scoring rooms for the short-answer items

Scoring takes place in a large room which is usually equipped with two groups of large tables: one for the French-language team and the other for the English-language team. The marking teams always include at least one architect who practices architecture in both English and French, whose role is to :

- Score copies in both languages
- Facilitate communication between both teams
- Clarify content, if needed
- Help with the alignment and performance between both teams

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 the markers are trained to use the scoring guide (scoring rubrics and exemplars). Real-time monitoring of the validity and reliability of the scoring is performed, and additional training is provided if consultants determine it is needed.

The markers perform scoring individually using an anonymous scoring procedure (i.e. a marker does not know what score was awarded by the other marker). Each test paper is scored at least twice by two different markers; if the awarded scores do not agree exactly, the paper is scored a third time by a third marker, reinforcing and contributing to the reliability of the scoring process.

Some copies are corrected more than three times when the disparity between the scores of the first two markers is significant. In the event of disagreement between scores after these processes, the average of the scores is awarded to the candidate.

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 markers interpret and apply the scoring material in the same manner and the consultants are able to supervise effectively. The goal is to obtain reliable, uniform, valid, and precise scoring of answers , no matter which markers participated in the scoring.

The consultants work to encourage markers to reject pre-conceived ideas concerning scoring procedures and harmonize their thinking and judgment with the scoring procedures and materials.

Daily additional training

The consultants frequently provide explanations concerning the scoring of specific items and of key elements of the item-specific scoring rubrics to provide exact and uniform application of material across individual markers, from one day to the next and before and after breaks.

Daily review of the exemplars

The consultants begin each day with a partial or complete review of scoring rubrics and exemplars to reorient markers and highlight sections of the scoring rubrics which may require special attention.

Routine activities to monitor intrarater and interrater reliability

Without special notice, markers may be asked to score copies that they themselves have already scored or copies that have received perfect agreement from other markers to protect and develop uniformity in the scoring over time and with other markers.

Reports to monitor the reliability of the scoring of the short-answer items

The consultants review 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 a candidate from two different markers) daily.

During scoring, each marker must maintain a minimum level of agreement of less than 1.5 overall score point average for a given section by the marker in comparison to the combined overall score point average by all other markers who scored the same copies. Markers who do not meet this requirement are given additional training, including re-review of exemplars. If this additional training does not correct the situation, scores awarded by these markers may not be used and a fourth scoring of the copies in question may be required.

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

Data reports that summarize the cumulative mean score and distribution of scores are used to monitor for deviation in the scoring of each marker and permit the identification of potential problems to be taken into consideration during subsequent analyses.

Cumulative reports on reliability

The consultants monitor reliability and uniformity during scoring by using an anonymous (unidentified) reinsertion of test papers.

All 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 markers according to their productivity level (by numbers of scored copies) and identify the markers who have a low level of productivity. The consultants review the data which is highlighted in this report to determine whether additional training is required for a specific marker.

Security

Standard 6 stipulates the importance of maintaining appropriate personal data and exam content security. Verification procedures have been put in place to check the identities of candidates. Identification numbers of the candidates are coded in accordance with Code 39 barcode symbology.

A complete closed-circuit wired network with data-entry forms is used to record the scores and calculate statistics in real-time while mitigating security risks. The use of a closed-circuit wired network prevents hacking and disturbance from external entities. During the scoring sessions, no computer, network, or database system contains personal information nor exam material. All personal identification information is coded and only item numbers and scores obtained are internally transferred. Public access networks and wi-fi are never used. An electronic barcode reader is used by markers to read and enter the identification number of the candidate and the marker identification number on a computerized form (the use of barcode technology considerably reduces transcription errors). 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 local database wired to the closed-circuit network. Test material used for scoring (test copies and training manuals) are always supervised or locked in a room.

Sensitive transmissions between the CExAC team and Associations are done using encrypted protocols.

EXCEPTIONAL PROCEDURES DURING THE SCORING

Indications of inappropriate content or of interference during in the administration of the ExAC

Occasionally the answers of the candidates to short-answer items contain content which is inappropriate or offensive, appears to indicate a change in handwriting, indicates possible interference by staff persons involved in the administration of the ExAC, or some other problem. Markers are trained to watch for anomalies and set aside copies they suspect might indicate such problems once they have been scored. They are evaluated by members of CExAC, who examine the incident reports and review with consultants, the appropriate jurisdiction, and or Senior Administrators from ROAC as needed.

Damaged or misprinted test material

When the test papers are distributed to the candidates, it is possible that these may be torn, poorly stapled, or missing pages. In such cases, the candidates should not be penalized, and the material should be replaced at the time of administration. If such damage is discovered during the scoring session, members of CExAC and consultants will review and determine the procedure for scoring these papers.

See the explanatory note earlier in this document concerning the management of printing errors.

RELIABILITY AND VALIDITY ARGUMENTS BASED ON THE METRIC PROPERTIES

Technical quality of the examination

The method used to analyze the technical quality of the ExAC conforms to classical test theory (CTT). The basis for this theory was first stated in 1907, notably in the work of Spearman (1907after which it evolved and was formalized during the 1960s. Theories and measurement models developed since then, including item response theory (IRT) can offer considerable advantages with respect to the technical analysis of items and tests in some testing programs, but are demanding with respect to assumptions and/or certain conditions of use. For example, to calibrate items as a function of a few parameters, a considerable sample of cases (i.e. candidates) is needed and the annual number of participants in the ExAC is insufficient. CTT has been chosen for the ExAC testing program because it is a valid theory, well-known and widely used, that is less demanding in terms of sample size.

Several parameters are considered when choosing the items to be included in the overall scores of the candidates.

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, which is not particularly useful in evaluating the performance of a candidate.

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 candidates with a higher level of performance and those with a lower level of performance. For example, a candidate with a higher level of overall performance on a test should have a greater chance of performing well on a particular item than would a candidate 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 considered to be good discriminators and are generally retained as part of the ExAC testing program.

The correlation between the results for an item and the total score on the test is analysed to check if a test is internally consistent. This involves determining whether all

the items measure the same thing or the same field or domain. An item which correlates poorly with the others may measure the knowledge of a topic which has little or nothing to do with the practice of architecture in Canada.

The parameters of difficulty (p), discrimination (D), and item/total correlation each contribute, in their own way, to determine 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 they are not included in the overall scores of the candidates.

Since the outcome of the examination has important consequences, precision measurements are performed to monitor its progress and to provide a confidence interval when interpreting the data. Cronbach's alpha, a conservative indicator of reliability, is calculated for each section of the ExAC and for each language. While the ExAC has not been in existence for very long, its reliability coefficients have increased steadily since its initial version. Even if a Cronbach's alpha of 0,7 or higher is considered acceptable, the high-stakes nature of the ExAC requires a much higher standard. The following table shows the progression of this indicator including for 2024 ExAC.

Historical Cronbach's coefficients

	English	French	Average alpha
Coefficients alpha (ExAC08)			0,807
Coefficients alpha (ExAC09)			0,825
Coefficients alpha (ExAC10)			0,822
Coefficients alpha (ExAC11)			0,813
Coefficients alpha (ExAC12)			0,835
Coefficients alpha (ExAC13)			0,851
Coefficients alpha (ExAC14)			0,846
Coefficients alpha (ExAC15)			0,846
Coefficients alpha (ExAC16)			0,851
Coefficients alpha (ExAC17)			0,877

Coefficients alpha (ExAC18)			0,876
Coefficients alpha (ExAC19)			0,883
Coefficients alpha (ExAC20)			0,879
Coefficients alpha (ExAC21)			0,868
Coefficients alpha (ExAC22)			0,891
Coefficients alpha (ExAC23)			
Section 1	0,880	0,854	
Section 2	0,895	0,879	
Section 3	0,900	0,885	
Section 4	0,915	0,885	0,887
Coefficients alpha (ExAC24)			
Section 1	**0,883**	**0,860**	
Section 2	**0,880**	**0,893**	
Section 3	**0,901**	**0,886**	
Section 4	**0,917**	**0,910**	**0,891**

Reliability estimates for the short-answer items on the ExAC 2024 specifically

We calculate the interrater reliability to measure the consistency among markers. Since all copies are scored at least twice, we verify if different markers provide similar scores.

The percentages of perfect agreement between the first two markers for Section 1 vary between 63,8 % and 95,0 % 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 generally accepted standards.

The percentages of perfect agreement between the first two markers for Section 3 vary between 68,2 % and 94,6 % 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 generally accepted standards.

When the scores from two markers for a candidate'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.

Other analyses, such as standard error measurement and differential test functioning, are also performed from time to time. A follow-up of candidates who have obtained exceptionally low scores is undertaken to determine if a subgroup of candidates is at a disadvantage.

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 the items used in the ExAC are directly linked to the general and specific objectives. The consultant selects items of suitable difficulty which permit candidates 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.

Equating

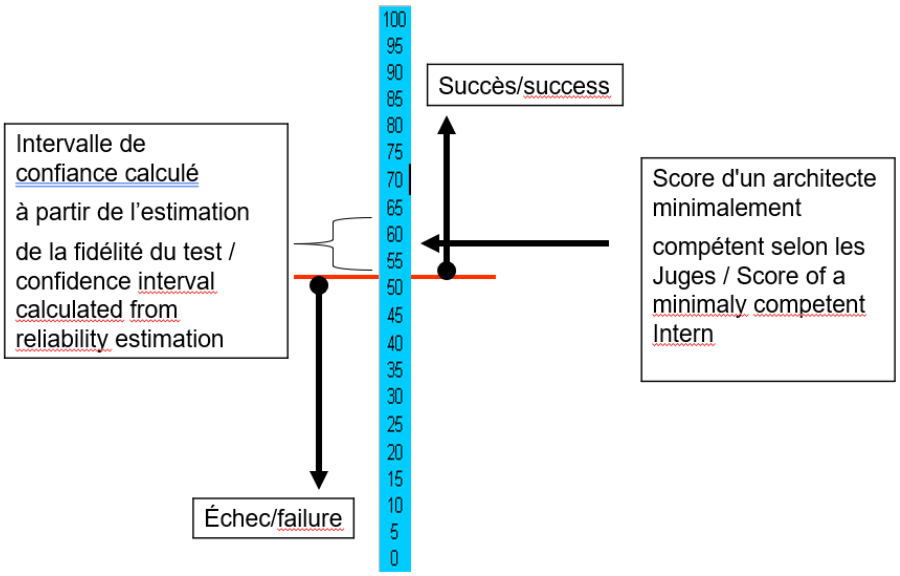
Equating is a procedure used to enable the results of the ExAC to be comparable from one year to another and across languages. It could be hypothesized that a linear relationship exists between the annual versions of the ExAC to the extent that the table of specifications, the types of items, and the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

Standard Setting

A series of steps which permit the determination of the passing score ("cut score") for each of the four tests is used for the ExAC program. The process requires taking into consideration contextual and historical references, the evolution of the average difficulty *(p)* of common items, and an estimate of the likelihood of success of a candidate with a

minimum level of competence. This process is performed in totality or partially every few years or when a major event occurs such as Covid-19 or a substantial modification in the content of the ExAC (for instance, updating the exam content to reflect new versions of the National Building Code of Canada).

The process was last performed in 2024. It began by asking several (architects who have been trained as markers during a previous scoring session to examine each item on the tests and to estimate the likelihood of success of a candidate with a minimum level of competence. The average of the estimates of the markers for the tests constitutes the minimum score that must be obtained on each test by a candidate with a minimum level of competence as proposed by Angoff (1971). 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 measurement instruments with an imperfect level of reliability, a confidence interval which is situated around the cut score estimated by the markers is calculated (the red line in the following figure shows the lower limit). This gives the benefit of the doubt to candidates who have an obvious minimum level of competence. A candidate who fails a test will have received a score which was lower than the limiting score determined by the markers, and this same result is also lower than the lowest interval of confidence that could be considered as a margin of error. The following figure illustrates the factors taken into consideration.



Summary of the 2024 Standard Settings exercise

Factors considered when revising the standards:

- Scores distribution;
- Precision of the instrument;
- Historical and contextual data;
- Evolution of relative difficulty levels of common items over time;
- Estimation of the performance of minimally competent architects.

Results

Item difficulty as monitored	Increased by 1,57 percentage point
Performance of minimally competent architects	Passing rate decrease by 2,1 percentage point
The result of these analyses and other contributing factors suggests that the performance of current candidates for the ExAC is lower. As a result, the overall passing rate has conservatively been decreased by approximately 1,5 percentage point.	

Publication of the results

The results of the ExAC tests are published in the language of the candidate, by province or territory, and for all of Canada. The results describe the performance of the candidates at this crucial step in their training. The stakes are very high for the candidates because passing the ExAC is an essential condition for gaining access to the profession. The results obtained by the candidates are crucial to the licensing authorities from an economic point of view and the protection of the public, and they will serve as a guide in their decision-making process. For all these reasons, a high degree of confidentiality is maintained to protect the integrity of the ExAC and the privacy of the personal information of the candidates. On behalf of he CExAC, the co-administrators 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** within a few months following the administration of the ExAC. Tables are also prepared and published to provide data concerning overall success rates.

The licensing authorities receive data files which provide detailed results of the performance of each candidate who is registered in their province or territory, and each licensing authority transmits results to the candidates in a performance report. This document indicates the exam status of the candidate, the sections of the ExAC that were passed, and the sections that need to be taken again. Results are indicated in pass or fail format, and raw scores are not provided.

Where a section must be taken again, the report provides constructive information on the portions of the test that were less-successfully completed. See the ExAC website for rules related to **retaking the ExAC**.

Success rate for the ExAC 2024

Seventeen markers, including several who were bilingual, participated in the scoring session for the ExAC 2024. The table below shows the success rate for each section. Of the 931 candidates who participated in the ExAC 2024, 672 passed all sections they were eligible to write. The overall success rate was **72,2 %**. The average pass rate per section was **83,5 %.**

Pass rate for the English and French form of the ExAC 2024

Section 1	Section 2	Section 3	Section 4
705 / 839	704 / 847	699 / 835	704 / 845
84,0 %	83,1 %	83,7 %	83,3 %

Requests for review

In compliance with Standard 8, a candidate may request a review. The precise rules which apply to this process are contained ion the **website**. Once a request has been made, two consultants perform a multi-point review of the bubble answer sheets. The short-answer items are scored twice again by two experienced markers. The databases containing the candidate data are checked to verify the integrity of the data.

REFERENCES

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC, American Educational Research Association.

Angoff, W.H. (1971). *Scales, norms and equivalent scores*. In R.L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals.* Handbook I: Cognitive domain. New York: David McKay Company.

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.

Kolesar, Peter J. (1994). *What Deming told the Japanese in 1950*. Quality Management Journal 2(1) (1994): 9–24.

Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality control.* Washington, DC: Graduate School of the Department of Agriculture.

Spearman, C. (1907). *Demonstration of formulae for true measurement of correlation*. American Journal of Psychology, 18, 161-169.

Swamidass, P. M. (2000). *Deming Cycle (PDCA).* In: Swamidass, P. M. (eds) Encyclopedia of Production and Manufacturing Management. Springer, New York.

GLOSSARY

GLOSSARY

Term	Explanation
Accommodation	A support mechanism and a service provided on request to candidates who have a condition which could hinder the participation of a candidate in the test sessions.
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 refers to a regional cultural concept might cause a disadvantage to individuals from other regions.
Bubble answer sheet	A standard answer sheet which allows a candidate to record the answers to multiple-choice items. The sheets are semi-rigid and can be read by an optical mark reader. The scores are transferred automatically to a database management system.
Certification, licensure or admissions test	A test which is intended to certify an individual in accordance with well-defined criteria. Since the stakes are extremely 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 licensure test.
CExAC	The Examination for Architects in Canada (ExAC) Committee.
Constructed response item	A synonym for short-answer item in this document.
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 candidates who are successful on item 22 are also successful

	on item 34, there is a positive correlation between the two items.
Differential item or test functioning	Presence of a 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 candidates who are successful on the ExAC as a whole and if it is answered incorrectly by candidates 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.
Edumetrics	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 National 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 National 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 to verify their measurement properties and their functioning.
High-stakes test	An expression which is used when the test results have profound consequences for those taking them. Certification tests such as the ExAC are examples of high-stakes tests.
Interrater reliability	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 candidate to a specific question.
Intrarater reliability	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 candidate 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 candidates who are successful in answering an item and the number of candidates 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 classroom tests by the larger numbers which are subjected to the testing. The ExAC is a large-scale test.
Licensing authority	A legal entity which is responsible for regulating the practice of a profession within a province or territory.
Marker	In this document, a synonym for marker 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 a candidate must attain 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 mechanical scanning device which can read bubble answer sheets and record the scores electronically.
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 marker in this document.
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.
Score	Refers to the points obtained and to the number of correct answers. Classical Test Theory is used for the ExAC to interpret the answers on a test.
Score point(s)	Point(s) which are awarded to a candidate 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 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 diverse levels of performance for an item and the number of points (score points) to be awarded to each level.
Setting cut scores / Standard setting	The processes of determining and confirming the minimum acceptable level of performance.

Short-answer item	An item which does not offer a selection of choices, but which requires that an answer be formulated.
Specific objectives	The items in the ExAC are designed to measure the knowledge and skills of candidates with respect to multiple criteria which are considered essential for architects to be marked competent to practice their profession. Example of a specific objective: Explain data from a legal land survey.
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 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 to 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 for each testing site.
Test version / Test form	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 candidates. 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, among other things.