

A National
Assessment
Collaboration

2018
NAC
Annual
Technical
Report



MEDICAL COUNCIL
OF CANADA

LE CONSEIL MÉDICAL
DU CANADA



TABLE OF CONTENTS

OVERVIEW	4
1. EXAM DEVELOPMENT	5
Blueprint and test specifications	5
Exam format	7
Exam content	7
Content validity.....	8
Competency descriptions and performance expectations	10
2. EXAM ADMINISTRATION	12
Candidate orientation.....	12
Exam sites and candidate numbers.....	12
Exam administration staff.....	13
Standardized Patient training.....	13
Chief Examiners.....	13
Common recruitment requirements for all MCC Exams.....	14
NAC PE recruitment requirements	14
3. EXAM SCORING	15
Standard quality assurance and quality control procedures	15
Exam result reporting.....	16
Scale scores	16
Pass/Fail status.....	20
Competency scores	21
4. PSYCHOMETRIC RESULTS	21
Scale scores	21
Estimates of score reliability and classification decisions	22
Historical comparisons.....	30
REFERENCES.....	31

TABLES AND FIGURES

Table 1: Test specifications for 2018 NAC exam.....	6
Table 2: Sampling of OSCE content by test specifications per test form for 2018.....	9
Table 2 (cont.): Sampling of OSCE content by test specifications per test form for 2018.....	10
Table 3: NAC exam rating scale and criteria	11
Table 4: NAC candidate numbers by test form 2018.....	12
Figure 1. Linking design for the 2018 NAC exam test forms	19
Table 5: Summary statistics of scale scores by form for each 2018 administration....	22
Figure 2. Score distribution by test form and for the whole year	22
Table 6: Reliability estimate, standard error of measurement, decision consistency and decision accuracy by form for each 2018 administration ...	24
Table 7: Summary statistics for OSCE stations by form for each 2018 administration.....	26
Figure 3. Competency score profile for test form 1 (all candidates), 2018	28
Figure 4. Competency score profile for test form 2 (all candidates), 2018	28
Figure 5. Competency score profile for test form 3 (all candidates), 2018	29
Figure 6. Competency score profile for test form 4 (all candidates), 2018	29
Table 8: Candidate performance data for 2013-2018.....	30

OVERVIEW

In response to the 2004 *Report from the Canadian Task Force on Licensure of International Medical Graduates* (IMGs; Federal/Provincial/Territorial Advisory Committee on Health Delivery and Human Resources, 2004), the Medical Council of Canada (MCC) began a series of related initiatives to support the assessment and training of IMGs in Canada. A steering committee was created and convened from 2005 to 2009 to develop a framework and governance structure for a National Assessment Collaboration (NAC).

The NAC is an alliance of Canadian organizations that are streamlining the evaluation process for international medical graduates (IMGs) seeking a license to practise medicine in Canada. A significant development of the NAC program is the pan-Canadian Objective Structured Clinical Examination (OSCE), known as the NAC Examination. The purpose of the NAC Examination (hereinafter referred to as “exam”) is to assist Canadian medical school clinical residency programs in selecting IMGs into the first year of postgraduate training. The intent of this national exam is to avoid duplication of assessments performed by provincial IMG assessment programs. Residency program directors are able to use candidate results to assist in making decisions about which IMG candidates are best qualified for entry into their programs. In 2018, the NAC exam was delivered in Alberta, British Columbia, Manitoba, Nova Scotia, Ontario, Saskatchewan, and Quebec, and was mandatory for application to The Canadian Resident Matching Service (CaRMS) in all provinces.

The National Assessment Central Coordinating Committee (NAC³) is the governing body responsible for the oversight and function of the NAC exam and reports to the MCC Executive Board.

The NAC Examination Committee (NEC), reporting to the NAC³, oversees the creation and maintenance of the NAC exam content. The NEC ensures that all content adheres to the NAC exam Blueprint, and that examiner rating

guidelines and the overall exam content and format meet NAC guidelines. In addition, the NEC approves the release of results, including adjudication on issues identified in scoring and quality assurance.

Policies and procedures have been established to ensure comparability of results from year to year, faster release of results over time and uniform quality control as well as quality assurance across exam dates and jurisdictions. To this extent, the MCC has developed and continues to update a library of structured procedures that help maintain uniformity in administration across regions and sites, as well as provide the basis for support material for Standardized Patients (SPs) and Physician Examiners (PEs). The standardization of procedures is necessary to support the validity argument that differences in test scores are due to differences in candidates' abilities as assessed by the NAC exam and not to site differences. Additionally, these policies and procedures are necessary for high-volume testing programs such as the NAC where the exam sessions may be geographically distributed and results must be comparable and uniform in quality.

This report summarizes exam administration aspects as well as key psychometric properties of all test forms for the NAC exam that took place in 2018.

1. EXAM DEVELOPMENT

This section describes the Blueprint and test specifications for the NAC exam, the format of the exam, how exam content is developed, and the scale and criteria used to rate competencies.

Blueprint and test specifications

The NAC blueprint was drafted over a series of meetings between 2009 and

2010 by a group of assessment experts, and ratified by the NAC Steering Committee in 2010. Since 2011, the steering committee’s successor group, the NEC, has maintained the original blueprint except for the testing of therapeutic knowledge. In 2013, the NAC Therapeutics Exam (a written exam) was removed from the Blueprint, and the testing of this knowledge was incorporated into the testing of clinical management skills to create a revised Management & Therapeutic competency.

Test specifications were developed for the NAC exam and approved by the NEC to meet the Blueprint and ensure that similar content is measured on each of the test forms. The process of creating a blueprint and test specifications ensures that candidates are measured on similar content across different test forms of the exam. All exam test forms are constructed by selecting OSCE cases/stations to best represent NAC test specifications.

Table 1 outlines the test specifications for the NAC exam and provides a summary of the required content and skills to be assessed in a test form, including the clinical competencies, systems, disciplines and patient age groups. An additional constraint of gender is also included to ensure the proportional distribution of patient gender across stations.

Table 1: Test specifications for 2018 NAC exam

CLINICAL COMPETENCY	Recommended # of stations	SYSTEM	Recommended # of stations
History Taking	7	Respiratory	min 1
Physical Examination	2-3	Cardiovascular	min 1
Combined Hx/Px	max 1	GI	min 1
Organization	10	MSK	} 2-3
Communication Skills	10	GU	
Language Fluency	10	Endocrine	
Diagnosis	min 3	Neuro	
Data Interpretation	min 3	Mental Health	} 2-3
Investigations	min 3	Reproductive Health	
Therapeutics and Management	Up to 20% must be TPx-specific	Multi-system	

DISCIPLINE	Recommended # of stations	AGE*	Recommended # of stations
Medicine	2-4	Newborn	1-2
Surgery	2-4	Infant 2-23 months	
Psychiatry	1-2	Child (preschool) 2-5 yrs	
OBGYN	1-2	Child (school age) 6-12 yrs	
Pediatrics	1-2	Adolescent 13-17 yrs	1-2
Geriatric Medicine	1-2	Young adult 18-44 yrs	4-5
		Adult 45-64 yrs	
		Older adult 65+ yrs	2-3
Urgent Care	1	*age of actual patient, not necessarily SP's age	

GENDER*

The ratio of M/F cases should be no greater than 6:4

* Gender of SP, not necessarily actual patient's gender.

Exam format

In 2018, the NAC exam test forms were composed of 10 operational 11-minute OSCE stations, designed to assess nine clinical competencies: communication skills, data interpretation, diagnosis, history taking, investigations, language fluency, organization, physical examination, and therapeutics and management.

In each station, an SP portrayed the clinical scenario, and each candidate's performance was evaluated by a PE. Each station measured between five and nine clinical competencies.

Standardized procedures, including training for PEs and SPs as well as data analyses, were followed to ensure that the NAC exam results were comparable across test forms for all candidates.

Exam content

NAC exam content is developed by a panel of clinical subject matter experts along with experts in medical education and assessment. In 2018, there were four NAC content development workshops where OSCE cases/stations were written, peer-reviewed and approved for piloting. These workshops focus on

developing content required to fulfill the NAC exam Blueprint and test specifications.

A NAC exam contains both operational and pilot stations. The operational stations, each including multiple items, are the basis for the candidates' reported score. Each exam also contains "blind" pilot stations that are completed under normal exam conditions but for which no score is reported to the candidates. Data obtained from these pilot stations are reviewed and analyzed after the exam. Feedback from PEs, SPs and SP trainers regarding pilot stations is also reviewed, and, if necessary, adjustments are made to the stations as a result. If significant adjustments are made to a station, the station is piloted a second time before it is used operationally.

To ensure that all NAC exams are comparable, each test form or iteration of the exam must meet specific testing criteria (see **Table 1** for test specifications).

Content validity

Measuring how well a test form matches the test specifications is one piece of evidence supporting the validity of score interpretations for the intended purpose of the examination (Kane, 2006; 2013). This section highlights the test specifications and how well each test form measures the required content and skills.

The NEC works with the MCC staff to select and approve the stations for a given test form. The test forms are drafted by the NAC TDO in accordance with the test specifications. The NEC then reviews the test forms, including individual stations, to ensure that test specifications are met and that content is at the appropriate assessment level (that of a recent graduate from a Canadian medical school). The NEC approves the final version of the content for each test form. For security reasons, multiple test forms are used each year.

Table 2 shows the sampling of test specification characteristics, clinical competencies, and number of stations for each of the five forms administered

in 2018. The *Recommended* column specifies the desired number of stations for each test form for each clinical competency, discipline, gender, system and age groups. There was one test form administered in March (Form 1) and three test forms administered in September (Forms 2, 3, and 4).

**Table 2: Sampling of OSCE content
by test specifications per test form for 2018**

		MARCH	SEPTEMBER		
CLINICAL COMPETENCIES	Recommended # of stations	Form 1	Form 2	Form 3	Form 4
History Taking	7	7	7	7	7
Physical Examination	2-3	2	2	2	2
Combined Hx/Px	max 1	1	1	1	1
Organization	10	10	10	10	10
Communication Skills	10	10	10	10	10
Language Fluency	10	10	10	10	10
Diagnosis	min 3	9	9	7	10
Data Interpretation	min 3	3	4	3	4
Investigations	min 3	6	6	6	7
Therapeutics and Management	min 3	7	8	9	8
DISCIPLINE	Recommended # of stations	Form 1	Form 2	Form 3	Form 4
Medicine	2-4	5	6	5	5
Surgery	2-4	3	2	2	3
Psychiatry	1-2	2	2	2	1
OBGYN	1-2	2	1	2	1
Pediatrics	1-2	1	2	2	2
Geriatric Medicine	1-2	2	1	1	1
Urgent Care	1	2	3	2	3
GENDER*		Form 1	Form 2	Form 3	Form 4
The ratio of M/F cases should be no greater than 6:4		M = 6	M = 4	M = 6	M = 4
		F = 4	F = 6	F = 4	F = 6

* Gender of SP, not necessarily actual patient's gender.

Table 2 (cont.): Sampling of OSCE content by test specifications per test form for 2018

SYSTEM	Recommended # of stations	Form 1	Form 2	Form 3	Form 4
Respiratory	min 1	1	1	1	1
Cardiovascular	min 1	1	1	1	2
GI	min 1	1	2	1	2
MSK	2-3	4	3	4	4
GU					
Endocrine					
Neuro					
Mental Health	2-3	4	5	4	3
Reproductive Health					
Multi-system					
AGE*	Recommended # of stations	Form 1	Form 2	Form 3	Form 4
Newborn	1-2	1	1	2	1
Infant 2-23 months					
Child (preschool) 2-5 years					
Child (school age) 6-12 years					
Adolescent 13-17 years	1-2	1	2	2	1
Young adult 18-44 years	4-5	6	5	4	6
Adult 45-64 years					
Older adult 65+ years	2-3	2	2	2	2

* Age of actual patient, not necessary SP's age.

Competency descriptions and performance expectations

In 2018, PEs used rating scales (described below) to score the candidates' proficiency on each of the competencies included in each station. Each station had one PE and, by the conclusion of the exam, each candidate had been evaluated by twelve PEs. The ratings provided by each PE were used to calculate all scores as described in the *Exam Scoring* section.

In 2018, PEs rated candidate performance for each competency relative to the standard of *a recent graduate from a Canadian medical school* using a

five-point Likert-type scale. The five rating points, along with the Acceptable performance level for each competency, are described in **Table 3**.

Orientation and training materials were given to PEs to provide more specific context for these ratings as well as for each station.

Table 3: NAC exam rating scale and criteria

UNACCEPTABLE as compared to a recent graduate from a Canadian medical school	BORDERLINE UNACCEPTABLE as compared to a recent graduate from a Canadian medical school	BORDERLINE ACCEPTABLE as compared to a recent graduate from a Canadian medical school	ACCEPTABLE as compared to a recent graduate from a Canadian medical school	ABOVE the level expected of a recent graduate from a Canadian medical school
○	○	○	●	○

The following are descriptors of ACCEPTABLE performance levels per competency.

HISTORY TAKING

Expectations: Acquires from the patient, family or other source a chronologic, medically logical description of pertinent events. Gathers information in sufficient breadth and depth to permit a clear definition of the patient's problem(s).

PHYSICAL EXAMINATION

Expectations: Elicits physical findings in an efficient logical sequence that documents the presence or absence of abnormalities, and supports a definition of the patient's problem(s). Sensitive to the patient's comfort and modesty; explains actions to the patient.

ORGANIZATION

Expectations: Approaches interaction in a coherent, succinct and systematic manner. Manages time effectively (including oral questions).

COMMUNICATION SKILLS

Expectations: Uses a patient-centered approach: establishes trust and respect, and shows sensitivity. Provides clear information and confirms patient's understanding throughout clinical encounter: encourages questions, and uses repetition and summarizing to confirm and/or reinforce understanding. Respects confidentiality. Avoids use of jargon/slang. If applicable, negotiates a mutually acceptable plan of management and treatment. Demonstrates appropriate non-verbal communication (e.g., eye contact, gesture, posture and use of silence).

LANGUAGE FLUENCY

Expectations: Speaks clearly (volume and rate) with clear pronunciation; accent does not hinder interaction. Uses coherent flow of words, phrases, sentences and verb tenses to convey intended meaning. Uses appropriate choice of words and expressions for the context (e.g., vocabulary).

DIAGNOSIS

Expectations: Discriminates important from unimportant information and reaches a reasonable differential diagnosis and/or diagnosis.

DATA INTERPRETATION

Expectations: Interprets investigative data appropriately in the context of the patient's problem(s).

INVESTIGATIONS

Expectations: Selects suitable laboratory or diagnostic studies to elucidate or confirm the diagnosis; takes into consideration associated risks and benefits.

THERAPEUTICS AND MANAGEMENT

Expectations: Discusses therapeutic management, including but not limited to pharmacotherapy*, adverse effects and patient safety, disease prevention and health promotion, when appropriate. Selects appropriate treatments (including monitoring, counseling, follow-up); considers risks and benefits of therapy and instructs the patient accordingly.

2. EXAM ADMINISTRATION

This section describes procedures to standardize exam administration, including candidate orientation, responsibilities of exam administration staff, standardized patient training, role of Chief Examiners (CEs), and PEs recruitment and training.

Candidate orientation

The MCC provides detailed information about the NAC exam for candidates on the MCC website. Topics include what to expect on exam day, scoring and results, as well as application information.

Exam sites and candidate numbers

The exam sites, as well as, the number of candidates for each test form are depicted in **Table 4**.

Table 4: NAC candidate numbers by test form 2018

		SITES	Total # of candidates	# of first-time test takers	# of repeat test takers
MARCH	Form 1	BC, AB-Calgary, MB, ON-Ottawa, ON-Toronto, QC	615	434	181
	Form 2	BC, AB-Edmonton, NS, ON-Toronto	351	297	54
SEPT.	Form 3	AB-Calgary, AB-Edmonton, BC, MB, ON-Ottawa, ON-London, ON-Toronto, QC, SK	676	563	113
	Form 4	AB-Calgary, AB-Edmonton, MB, ON-Ottawa, ON-London, ON-Toronto, QC, SK	728	577	151
TOTAL			2370	1871	499

Exam administration staff

Each exam site is responsible for hiring and supervising exam administration staff. These individuals in turn work with the MCC to ensure the security of exam materials and the quality of performance of all people involved in the exam (SP Trainers, SPs, CEs, PEs, exam day staff, caterers). NAC policies and procedures provided by the MCC ensure the standardization of the exam administration. MCC staff oversees site staff, either in person or via electronic communication, on exam days across the country in addition to offering an assistance line.

Standardized Patient training

Each site is responsible for hiring and supervising the SP Trainers who, in turn, oversee the SPs and assure the quality of their standardized performance on exam day(s). SPs are trained at each site using standardized NAC training material provided by the MCC. Training support is provided centrally to SP Trainers by MCC staff, primarily by the NAC Training Coordinator.

Chief Examiners

All NAC exam sites employ physicians as CEs. The role of the CE depends on exam site size and on how the Site Administrator chooses to delegate tasks.

Each CE is responsible for:

1. Assisting with PE recruitment and training, if needed.
2. Assisting with the dry runs of SPs prior to exam day, including a final assessment of SPs' readiness to perform in a standardized manner according to their patient scripts on exam day.
3. Overseeing examiners and candidates on exam day.
4. Addressing, where appropriate, candidates' questions, concerns and complaints on exam day.

5. Reviewing and signing all incident reports recorded on exam day.

Note: One exam site (Nova Scotia) also hires a deputy registrar to share responsibilities with the CE.

Common recruitment requirements for all MCC Exams

- PEs or markers must be registered and in good standing with a Medical Regulatory Authority (MRA) in Canada.
- PEs or markers may be retired, but they must have an active licence with an MRA in Canada.
- PEs or markers must be practising in Canada, or they must have practised in Canada within the last 5 years.
- All PEs and markers must adhere to the MCC Code of Business Conduct.
- PEs or markers must have the ability and stamina to complete the task (e.g., uncorrected hearing loss is a serious handicap).

All exceptions must be approved by the examination manager.

NAC PE recruitment requirements

Physicians must meet ALL of the Common Recruitment Requirements for all MCC Exams.

1. Physicians must have the Licentiate of the Medical Council of Canada (LMCC) and must provide their LMCC registration number. Physicians who DO NOT have their LMCC will be accepted as examiners under the following conditions:

- Non-licentiate examiners must be faculty members (e.g., faculty lecturer, assistant professor, associate professor or professor).

AND

- Non-licentiate examiners must be certified by one of the following organizations and must provide their certification number:

- Royal College of Physicians and Surgeons of Canada (RCPSC)
- Collège des Médecins du Québec (CMQ)
- College of Family Physicians of Canada (CFPC)

AND

- Non-licentiate examiners must sign a waiver indicating that they have no intention of taking the NAC examination.
2. Physicians must have recent experience supervising clerks and/or PGY1s, and/or they must have experience as an examiner at this level of training.
 3. Physicians may be community physicians (i.e., they do not need to be faculty members if all other criteria are met).
 4. Physicians must be currently practising medicine in Canada; if they are a resident physician, they must be PGY4 or higher OR have CCFP certification at the time of the examination.
 5. If retired, physicians must be within three years of practising in Canada.

The MCC provides training to standardize PE scoring to the exam standard using a scoring exercise with guided discussions. It provides a pre-exam, online training for all new and returning PEs, and an exam-day PE training delivered by the exam sites' CEs and senior site staff.

3. EXAM SCORING

In this section, we describe the scoring quality assurance (QA) and quality control procedures as well as which scores are reported and how they are calculated.

Standard quality assurance and quality control procedures

To ensure the accuracy and integrity of the candidates' exam day electronic records, a number of data QA steps are performed as outlined below.

PEs complete a Rating Scale Form for every candidate seen in their OSCE station. These forms are scanned at the exam sites and transmitted securely to the MCC. MCC staff import the forms into Teleform (a Cardiff Software Inc. program) where they are reviewed. Scanning anomalies are identified (for example, a non-readable candidate barcode, PE's pencil marks too faint) and corrections are made. The data is then exported electronically into a scoring application to generate a list of all candidates who fall within +/- five per cent of the pass score (approximately one standard error of measurement above or below the pass score). Once the paper copies of the Rating Scale Forms arrive at MCC, all the sheets for this candidate group are reviewed by staff for discrepancies against the electronic data reports. Although rare, any differences are corrected in the electronic data files to reflect the paper Rating Scale Forms. Any updated electronic files are then re-imported into the scoring application.

Exam result reporting

For each candidate, several elements are analyzed and reported: (1) scale score, (2) pass/fail status and (3) competency scores.

The following sections outline the standard steps in creating the scale scores, establishing the cut score and calculating competency scores reported to candidates, IMG programs and the Canadian Resident Matching Service (CaRMS).

Scale scores

In general, there are two steps to deriving scale scores. The first step involves computing a total score. The second step can be divided further in two to show how candidates' total scores are converted to scale scores to report to candidates and IMG programs.

Step 1: Total scores

The first step in deriving a total score is to calculate the station score for each OSCE station with the following formula:

$$\text{station score} = \frac{((\text{mean of rating scales})-1)}{((\text{max score on rating scale})-1)} * 100,$$

where the *mean of the rating scales* is the average of the rating scales for the competencies of that station for a candidate and the *max score on rating scale* (maximum possible score on each rating scale) is 5. The subtractions in the numerator and the denominator are for converting each station's scores to a per cent correct range of 0 to 100. For example, a station containing five competencies that are scored with ratings of 4, 4, 4, 5 and 5 would result in the following score:

$$\text{station score} = \frac{((4 + 4 + 4 + 5 + 5)/5) - 1}{(5) - 1} * 100 = \frac{((4.4) - 1)}{(5) - 1} * 100 = 85.0$$

The OSCE station scores are used first and foremost to calculate the total score for each candidate. It is calculated using the mean of the 10 OSCE station scores. This total score is then used as part of the second set of steps explained in the next section.

Before getting to those next steps, it is important to mention that calculations take into account missing data in the above-mentioned instance. Since station scores are based on the competency mean for that station, missing data needs to be taken into account so as not to negatively impact a candidate's score. Missing data occurs when the PE does not provide a competency rating for a given candidate on the scannable Rating Scale Form. When rating scales do not have a score provided by the PE, the station score is based only on the mean of the ratings provided by the PE. For example, a station containing five competencies of which only four were scored with ratings of 4, 4, 4 and 5 would result in the following score:

$$\text{station score} = \frac{((4 + 4 + 4 + 5)/4) - 1}{(5) - 1} * 100 = \frac{((4.25) - 1)}{(5) - 1} * 100 = 81.25.$$

The station score would have been 60 per cent if the missing rating scale was treated as zero and the adjustment not applied.

Step 2: Linking and scale scores

The second set of steps to obtain a scale score can be divided further in two, in order to show how candidates' total scores are converted to scale scores to report to candidates and IMG programs. The first step (2a) is to link through common stations the scores from current 2018 test forms to scores from previous test forms through a chain of linking steps dating back to test form 1 in 2013 that was used for setting the cut score and establishing the scale. The second step (2b) is to convert the linked total scores for the current test forms to scale scores that are reported to candidates and IMG programs.

2a: Linking

As described in the *Exam Development* section, for security purposes, multiple test forms are used each year. Each form contains a different sample of stations to meet the test specifications. However, test forms containing different stations may result in one test form being slightly more difficult or easier than another.

The process of linking total scores statistically takes into account the small differences in test form difficulty and adjusts the total scores for the test form being linked so that all scores are on the same metric. Linking also provides a way to apply the same performance standard to candidates who take different test forms. A detailed description of linking designs, issues and methods can be found in *Test equating, scaling, and linking: Methods and practice* (2nd Edition) authored by Kolen and Brennan (2004). Linking scores allows for a comparison of scale scores reported to candidates and IMG programs over time.

One method to link test forms is to have a subset of content appear identically across test forms. This is a common-item non-equivalent group design. The subset of content that is presented identically is called an anchor set and it is used to statistically estimate the overall ability of candidates that took each test form and the test form difficulty. For the NAC 2018 test forms, anchor sets were included across the four test forms and used to link

candidates' total scores. Test form 1 in 2018 was selected as the base form to which all subsequent test forms in 2018 were linked. A graphical representation of the linking design is shown in **Figure 1**, demonstrating that test form 1 was used as the base for the sets of anchor stations. Approximately 30 per cent of the content was selected as anchors to statistically link the test forms. The rule of thumb is to have at least 20 items or 20 per cent of the test as the anchor.

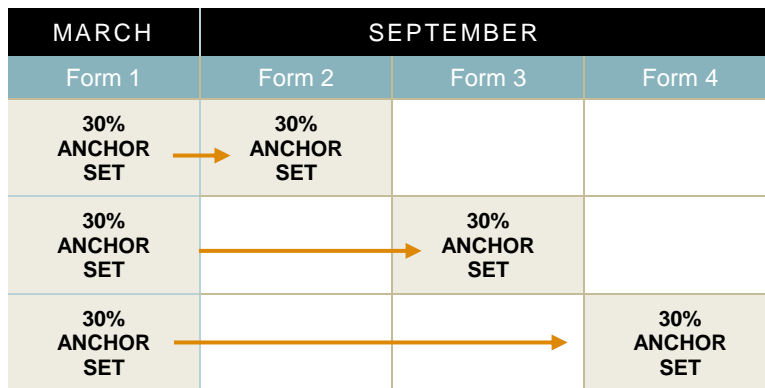


Figure 1. Linking design for the 2018 NAC exam test forms

The Levine observed score method is used to link scores between NAC test forms (Kolen & Brennan, 2004).

2b: Scale score transformation

Once total scores are established and linked to the base test form, they are transformed to scale scores for reporting purposes. A slope and intercept are used to transform all linked scores. This final transformation ensures that any differences in scale score means and standard deviations on the current test forms can be directly compared to test form 1 of 2013. For example, a mean increase from 70 to 75.0 or decrease from 70 to 65.0 would indicate that the general performance of the candidates who took the subsequent test forms is of higher or lower ability, respectively. The final transformation formula for forms 1 through 5 scores is as follows:

$$ScaleScore_x = (slope)(LinkedScore_x) + (intercept),$$

where

$ScaleScore_x$ is defined as the linear function to calculate the scale score for candidate X,

slope is equal to 1.06 applied to all test forms (based on the initial transformation of form 1, 2013),

intercept is equal to 3.34 applied to all test forms (based on the initial transformation of form 1, 2013), and

$LinkedScore_x$ is the linked score for candidate X.

After these transformations for test forms 1 through 5, all scale scores are rounded to a whole number between 0 and 100 (note that this is not a percentage scale). The reported scale scores as seen by candidates and IMG programs are this rounded value. For example, a linked score of 81.25 would result in the following scale score:

$$ScaleScore_x = (1.06) * (81.25) + (3.34) = 89.47 \text{ rounded to } 89.$$

Pass/Fail status

The cut score for this exam was set by a panel of 18 physicians from across the country, representing faculties of medicine and different specialties, and with years of experience supervising students and residents. The panel members reviewed stations, content and score information to provide judgments for establishing the recommended cut score. Test form 1 from March 2013 was used to establish the cut score, and a borderline group method was used for standard setting. Full details of the standard setting exercise can be found in “Recommending a Passing Score on the National Assessment Collaboration (NAC) Examination: Procedures and Results” (Hambleton & De Champlain, 2013). The reported scale score and cut score of 65, which was recommended by the panel of physicians in 2013 and approved by the NEC in April 2013, was used to assign each candidate from the 2018 administration either a pass or fail status. A score of 65 or greater was deemed a pass.

Competency scores

The calculation of a competency score is similar to that of a station score. It is calculated by converting the rating scale average (across stations) for that competency to a score of 0 to 100. Here is an example of the general formula to calculate the competency score for “Organization”:

$$\begin{aligned} & \textit{Competency Score}_{\textit{Organization}} \\ &= \frac{((\textit{mean of Organization rating scales}) - 1)}{((\textit{max score on rating scale}) - 1)} * 100 \end{aligned}$$

Competency scores are used to create the graphical representation of each competency as reported in the Supplemental Feedback Report (SFR). Competency scores are not used to calculate the total score or scale scores (as outlined above); therefore, competency scores cannot be directly compared to the candidates’ scale scores. Competency scores only provide some general feedback to candidates on their performance on the NAC exam.

4. PSYCHOMETRIC RESULTS

In this section, summary statistics for scale scores and pass rates are provided, as well as estimates of reliability and classification decisions, and a summary of item quality and competency score profiles. Results reviewed and approved by the NEC following the March and September administrations are used in this section, excluding candidates whose status is “No Standing” or “Denied Standing” or who missed more than one station.

Scale scores

Summary statistics of scale scores and pass rates from each test form are presented in **Table 5**. The score distributions by test form and for the whole year are displayed in Figure 2. These statistics are based on the scale scores reported to candidates. The minimum, maximum, and standard deviation are indicators of the variation in scale scores.

Table 5: Summary statistics of scale scores by form for each 2018 administration

Administration		N	Min.	Max.	Mean	Median	SD	Pass Rate
MARCH	Form 1	615	35	91	71.9	73	7.9	86.2%
	Form 2	351	38	91	70.3	71	8.1	78.6%
SEPT.	Form 3	675	32	93	73.5	74	8.4	86.8%
	Form 4	726	38	91	71.0	72	8.9	80.3%
TOTAL		2367	32	93	71.8	73	8.5	83.5%

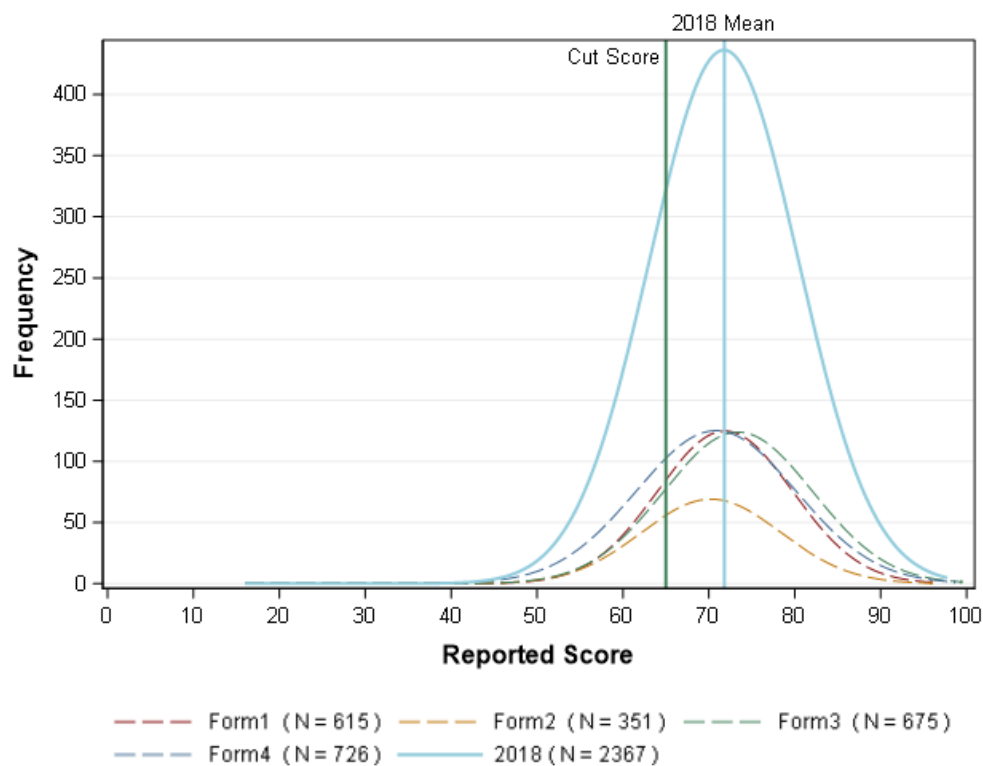


Figure 2. Score distribution by test form and for the whole year

Estimates of score reliability and classification decisions

Table 6 shows the reliability estimates, the Standard Error of Measurement (SEM), the decision consistency and decision accuracy estimates along with the associated false positives and false negatives by test form. The

estimated false positives indicate the expected proportion of candidates who pass based on their observed score, but who should fail based on their true ability. The estimated false negative rate indicates the expected proportion of candidates who fail based on their observed score, but who should pass based on their true ability.

Cronbach's alpha

Cronbach's alpha was used to estimate score reliability for the NAC exam, as it is appropriate for evaluating the effects of stations and candidate performance. This reliability estimate is described in *Educational Measurement* by Haertel in section 2.4.4 (Brennan, 2006). The formula for Cronbach's alpha is:

$$\alpha\rho_{XX'} = \frac{n}{n-1} \left(1 - \frac{\sum \sigma_{X_i}^2}{\sigma_X^2} \right)$$

where n is the number of stations, $\sigma_{X_i}^2$ is the variance for each station, σ_X^2 is the variance of the total score (Haertel, 2006, p. 74). A score reliability estimate indicates the desired consistency (or reproducibility) of exam scores across replications of measurement (Crocker & Algina, 1986; Haertel, 2006). Cronbach's alpha for each form is included in **Table 6**. The reliability estimate in conjunction with the total exam SEM can provide further evidence of the reliability of the candidate's scale score.

Standard error of measurement

The SEM provides a value within a certain confidence range (for example 68 per cent or 95 per cent) that a candidate's true score is captured by the observed score. SEM values should be as small as possible so that the measurement of the candidate's ability contains as little error as possible. The SEM is calculated as follows:

$$SEM = \sigma_X \sqrt{1 - \alpha\rho_{XX'}}$$

where σ_X is defined as the standard deviation for the total score (square root of the variance), and $\alpha\rho_{XX'}$ is defined as the reliability estimate as shown

above. The SEM on the reported score scale for each form is listed in **Table 6**.

Decision accuracy and decision consistency

A critical concern for a high-stakes exam such as the NAC exam is the pass/fail decision. Reliability of the NAC exam can also be assessed by examining the consistency and accuracy of pass/fail decisions based on exam scores. Decision consistency and decision accuracy can be estimated using the Livingston and Lewis (1995) procedure. Decision consistency is an estimate of the agreement between the pass/fail classifications on potential parallel forms of the exam. Decision accuracy is an estimate of the agreement between the pass/fail classifications based on observed exam scores and those that would be based on their true scores (that is, the expected score [i.e., average score] if the candidates could be tested an infinite number of times).

Table 6 includes the decision consistency and decision accuracy estimates for each form. Ideally, both of these values should be high, such as 0.80 and above, suggesting reliable and valid pass/fail classifications. A value of 0.80 indicates that either the accuracy or the consistency of the decision is being met for 80 percent of the candidates.

Table 6: Reliability estimate, standard error of measurement, decision consistency and decision accuracy by form for each 2018 administration

	MARCH	SEPTEMBER		
	Form 1	Form 2	Form 3	Form 4
Reliability estimate	0.73	0.74	0.77	0.67
SEM (reported score scale)	4.09	4.14	4.02	5.11
Decision consistency	0.89	0.85	0.90	0.84
False positive	0.07	0.08	0.05	0.09
False negative	0.05	0.07	0.04	0.07
Decision accuracy	0.92	0.90	0.93	0.89
False positive	0.03	0.04	0.03	0.04
False negative	0.04	0.06	0.04	0.07

OSCE station statistics

Summary statistics for each of the OSCE stations by administration and form are provided in **Table 7**. The percentages of missing data, the proportion correct (p-values), standard deviation and station total correlations (STCs) are presented.

P-values are the average proportion correct scores that candidates achieved on each of the stations. In general, p-values indicate station difficulty. P-values range between 0 and 1. Station p-values that are low (< 0.20) indicate a difficult station and those that are high (> 0.90) indicate an easy station. P-values are sample dependent, where comparisons across different samples of candidates do not take into account potential differences in overall ability. Therefore, p-values should not be overinterpreted or used as the only indicator of difficulty. Rather, p-values provide a general sense of the range of difficulty of stations on a particular test form.

Standard deviations indicate the general variability of station scores. The STCs are indicators of discrimination between low- and high-ability candidates for a given station. A low or negative STC (< 0.15) indicates that on a given OSCE station, high-ability candidates are not scoring as well as expected, and that low-ability candidates are scoring better than expected. This information, along with p-values, may indicate stations that should be removed from scoring, as these stations are not adequately discriminating low- and high-ability candidates. A moderate to high STC (> 0.30) indicates that high-ability candidates are performing well on a given OSCE station. Stations with STCs that are below 0.15, as well as with negative values are flagged for review. Stations are retained if content is deemed acceptable.

Table 7: Summary statistics for OSCE stations by form for each 2018 administration

Station	MARCH				SEPTEMBER			
	FORM 1				FORM 2			
	% Missing	p-value	SD	STC	% Missing	p-value	SD	STC
1	0.19	0.69	0.11	0.40		0.66	0.12	0.45
2	0.03	0.59	0.14	0.38	0.24	0.66	0.14	0.41
4	0.08	0.63	0.15	0.29	0.09	0.64	0.14	0.37
5	0.11	0.66	0.12	0.47	0.04	0.56	0.13	0.30
7	0.06	0.65	0.12	0.40	0.14	0.62	0.12	0.43
8		0.62	0.14	0.44	0.08	0.58	0.12	0.47
9	0.03	0.57	0.18	0.32	0.08	0.58	0.15	0.36
10		0.63	0.14	0.43	0.06	0.60	0.17	0.46
11	0.05	0.57	0.18	0.38		0.57	0.16	0.36
12	0.12	0.60	0.15	0.44	0.04	0.61	0.14	0.41
Mean	0.08	0.62	0.14	0.40	0.10	0.61	0.14	0.40

Station	SEPTEMBER							
	FORM 3				FORM 4			
	% Missing	p-value	SD	STC	% Missing	p-value	SD	STC
1	0.21	0.54	0.14	0.41	0.08	0.66	0.13	0.39
2		0.63	0.16	0.49	0.12	0.55	0.17	0.39
4	0.12	0.64	0.15	0.52	0.42	0.62	0.14	0.43
5	0.15	0.59	0.15	0.49	0.05	0.61	0.16	0.45
7	0.11	0.65	0.13	0.38	0.16	0.62	0.14	0.34
8	0.05	0.63	0.16	0.44	0.28	0.62	0.13	0.45
9	0.05	0.62	0.16	0.53	0.02	0.58	0.17	0.40
10	0.15	0.61	0.14	0.38	0.20	0.61	0.14	0.43
11	0.15	0.65	0.14	0.40	0.07	0.59	0.17	0.48
12	0.34	0.62	0.14	0.42	0.10	0.60	0.16	0.47
Mean	0.15	0.62	0.15	0.45	0.15	0.61	0.15	0.42

An STC value less than 0.30 in test form 1 was brought forward for review. Following a review by exam content experts, the station was deemed acceptable and included in the score calculation.

Competency score profiles

A competency score profile is presented in the form of a graph to each candidate in the SFR. A sample of the Statement of Results and the Supplemental Feedback Report is provided in Appendix A and B respectively. The graph shows the competency score for each of the nine competencies and the standard error of measurement around the competency score (communication skills, data interpretation, diagnosis, history taking, investigations, language fluency, organization, physical examination, and therapeutics and management). The calculation of the competency scores for each candidate is outlined in the Competency Score section within the Exam Scores section in this report.

This section provides competency score profiles for each of the four test forms in 2018. The range of competency scores for each test form is shown graphically in **Figures 3 through 6**. The boxes for each competency indicate the range for 50 per cent of candidates' competency scores. The vertical line represents the median or 50th percentile competency score. The remaining 25 per cent of competency scores are shown to the right or left of the box as a line. The mean competency score is shown by the diamond. The note under each figure indicates the naming convention for each competency.

To summarize, **Figures 3 to 6** show the competency scores for each test form for all candidates. These graphs indicate that some competency scores are more variable (for instance, the range of competency scores is quite wide with very high and low scores). In addition, the graphs show the relative strengths and weaknesses of the competencies being measured on each test form. Note that competency profiles cannot be compared across test forms as they are created for individual test forms and they are not linked statistically across forms like the total exam scores.

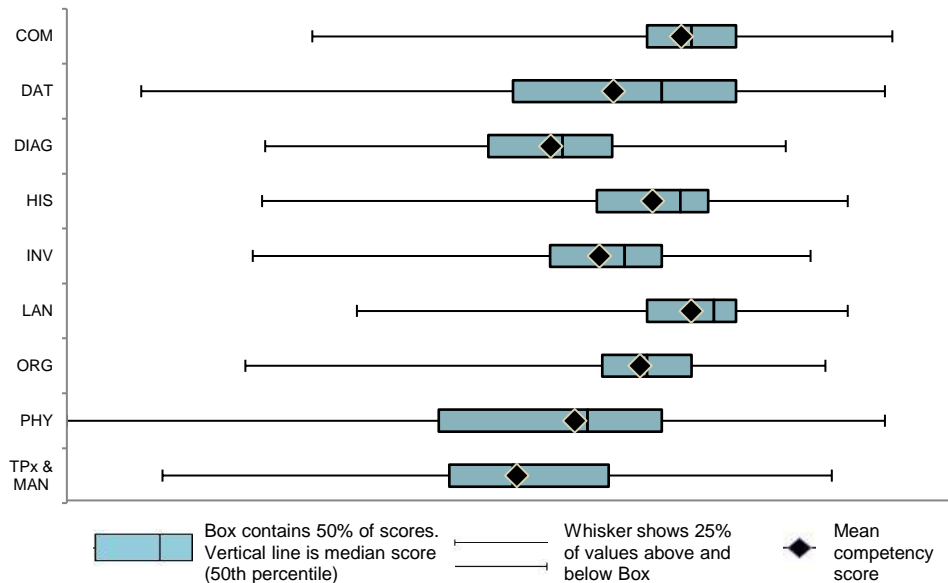


Figure 3. Competency score profile for test form 1 (all candidates), 2018

Note: COM = communication skills, DAT = data interpretation, DIAG = diagnosis, HIS = history taking, INV = investigations, LAN = language fluency, ORG = organization, PHY = physical examination, TPx & MAN = therapeutics and management

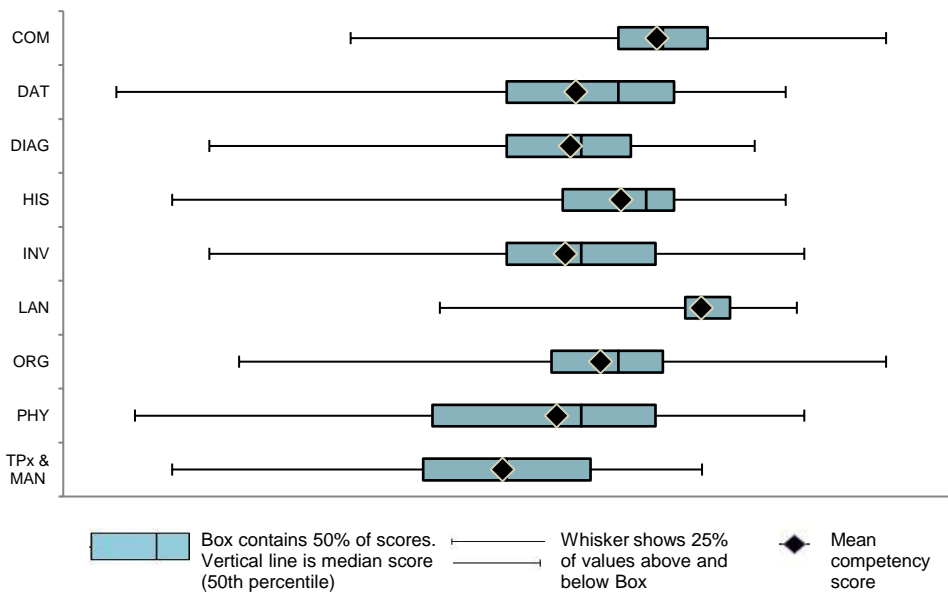


Figure 4. Competency score profile for test form 2 (all candidates), 2018

Note: COM = communication skills, DAT = data interpretation, DIAG = diagnosis, HIS = history taking, INV = investigations, LAN = language fluency, ORG = organization, PHY = physical examination, TPx & MAN = therapeutics and management

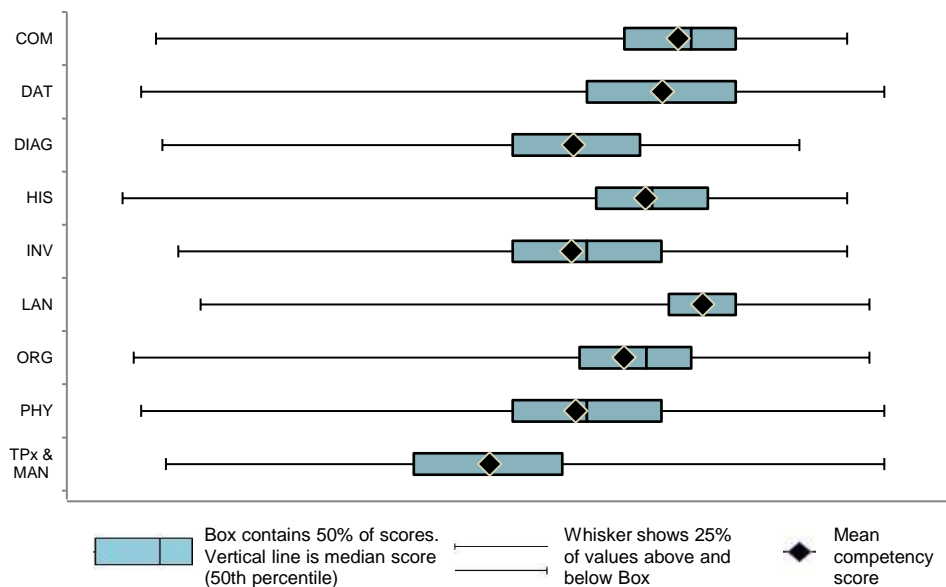


Figure 5. Competency score profile for test form 3 (all candidates), 2018

Note: COM = communication skills, DAT = data interpretation, DIAG = diagnosis, HIS = history taking, INV = investigations, LAN = language fluency, ORG = organization, PHY = physical examination, TPx & MAN = therapeutics and management

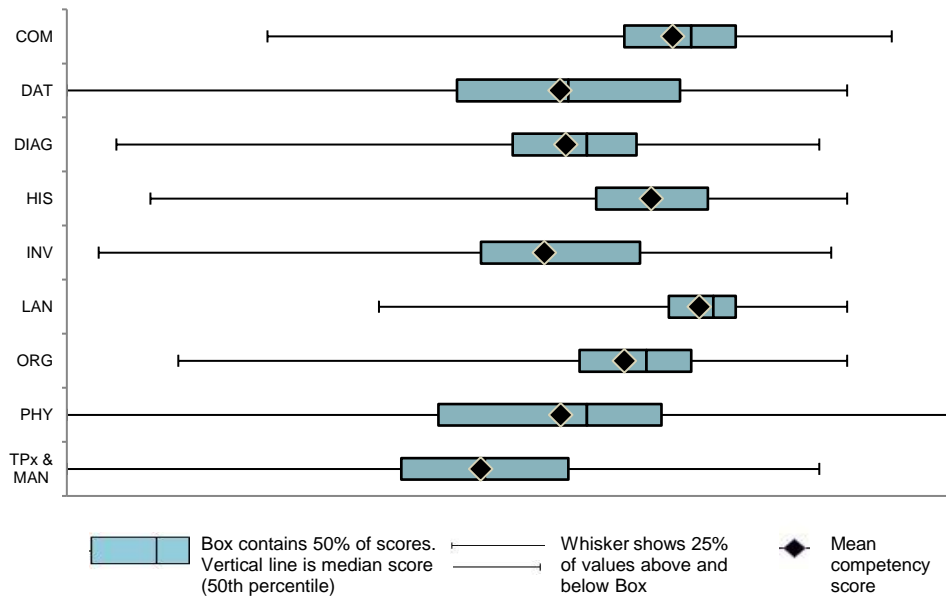


Figure 6. Competency score profile for test form 4 (all candidates), 2018

Note: COM = communication skills, DAT = data interpretation, DIAG = diagnosis, HIS = history taking, INV = investigations, LAN = language fluency, ORG = organization, PHY = physical examination, TPx & MAN = therapeutics and management

Historical comparisons

Table 8 presents candidate performance data for 2013-2018 comparison.

Table 8: Candidate performance data for 2013-2018

Year	Session	TOTAL		First-time test takers		Repeat test takers	
		# Tested	Pass rate	# Tested	Pass rate	# Tested	Pass rate
2018	Total	2369	83.45%	1870	82.25%	499	87.98%
	March	615	86.18%	434	85.94%	181	86.74%
	Sept.	1754	82.50%	1436	81.13%	318	88.68%
2017	Total	1506	86.79%	1430	87.76%	76	68.42%
	March	330	87.88%	292	89.04%	38	78.95%
	Sept.	1176	86.48%	1138	87.43%	38	57.89%
2016	Total	1547	92.63%	1447	93.99%	100	73.00%
	March	380	91.84%	327	94.50%	53	75.47%
	Sept.	1167	92.89%	1120	93.84%	47	70.21%
2015	Total	1487	91.19%	1392	91.95%	95	80.00%
	March	335	84.48%	295	84.07%	40	87.50%
	Sept.	1152	93.14%	1097	94.07%	55	74.55%
2014	Total	2106	92.64%	1680	93.00%	426	92.00%
	March	420	83.81%	296	82.77%	124	86.29%
	Sept.	1686	94.84%	1384	95.09%	302	93.71%
2013	Total	1221	77.58%	986	76.00%	235	85.00%
	March	547	70.93%	439	68.79%	108	79.63%
	Sept.	674	83.09%	547	81.54%	127	89.76%

REFERENCES

- Bahry, L., Hambleton, R. K., Gotzmann, A., De Champlain, A. F. & Roy, M. (2012). National Assessment Collaboration Standard Setting Study Report. *Technical Report for the Technical Advisory Committee and National Assessment Collaboration Examination Committee*. MCC: Ottawa, Canada.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications Inc.: Thousand Oaks, CA.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich College.
- Federal/Provincial/Territorial Advisory Committee on Health Delivery and Human Resources. *Report from the Canadian Task Force on Licensure of International Medical Graduates* (February 2004).
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65-110). Westport, CT: Praeger.
- Hambleton, R. K. & De Champlain, A. F. (2013). Recommending a passing score on the National Assessment Collaboration (NAC) examination: Procedures and results. *Technical Report for the National Assessment Collaboration Examination Committee*. MCC: Ottawa, Canada.
- Kane, M.T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.17-64). Westport, CT: Praeger.
- Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kolen, M.J., & Brennan, R.L., (2004). *Test equating, scaling, and linking: Methods and practice* (2nd ed.). New York, NY: Springer Science Business Meeting.
- Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32(2), 179-197. doi: 10.1111/j.1745-3984.1995.tb00462.x

APPENDIX A: NAC Examination Statement of Results (SOR)

	National Assessment Collaboration Examination Statement of Results
Candidate name: Aaaaaaaaa, Aaaaaaaaaaaaa	
Candidate code: 0000000000	
Examination session: January 2018	Your final result: Pass
Pass score: 65	Your total score: 88

February 7, 2018

We are writing to inform you of your final result on the National Assessment Collaboration Examination.

Your total score, representing your overall performance, is reported as a scaled score ranging from 0 to 100. Your final result is based on your total score relative to the pass score.

For more information, please visit the exam's Scoring web page on our website, mcc.ca.

Supplemental information on your examination performance is reported to you in a separate document within your physiciansapply.ca account.

 mcc.ca
physiciansapply.ca
inscriptionmed.ca

APPENDIX B: NAC Examination Supplemental Feedback Report (SFR)

	
SUPPLEMENTAL FEEDBACK REPORT	
Candidate name: Aaaaaaaaa, Aaaaaaaaaaaaa	Examination: NAC
MCC candidate code: 0000000000	Examination date: 2018-01-19
Your total score: 75	

The purpose of this report is to provide you with supplemental information on your relative strengths and weaknesses, based on your performance across the different domains that were assessed by the test form of the National Assessment Collaboration (NAC) examination that was administered to you.

Figure 1 displays your performance as measured across stations for each of the following nine domains: Communication Skills, Data Interpretation, Diagnosis, History Taking, Investigations, Language Fluency, Organization, Physical Examination and Therapeutics (TPx) and Management. Each domain is sampled a number of times, with some being measured across all stations and others across a smaller number of stations.

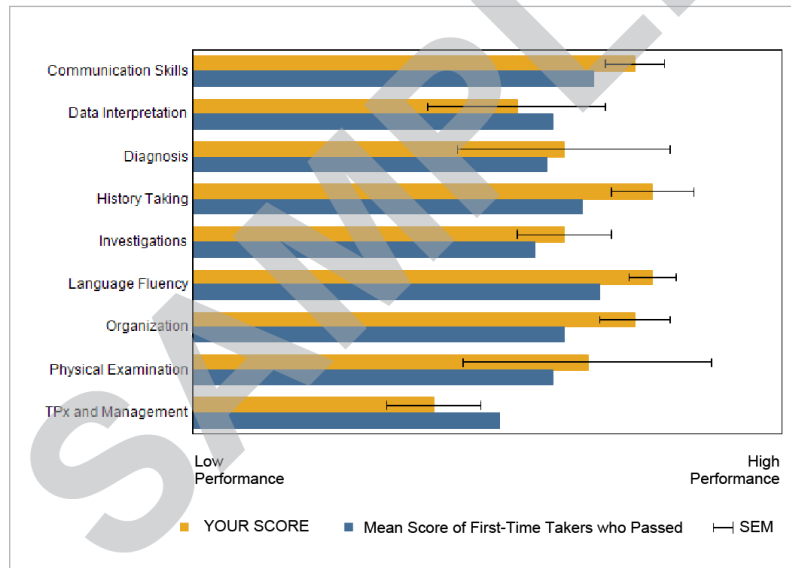
To help you better understand your performance, your subscore for each domain is shown along with the mean score of candidates from this examination session who were first-time takers of the NAC examination and who passed. The standard error of measurement (SEM) associated with each of your subscores represents the expected variation in your subscore if you were to take this examination again with a different set of stations covering the same or similar domains. Small differences in subscores or overlap between SEMs are indicative that performance in those domains was relatively similar. Likewise, overlap between the SEM for a domain subscore and the mean score of first-time takers who passed, within a given domain, signifies that performance is similar to the mean.

Report date: 2018-02-07 1/2

It is important to note that the subscores are based on significantly less data than the total score and that these do not have the same level of precision as the total score. If you have failed the examination and wish to retake it, preparation for all domains is important; otherwise you could improve some subscores and inadvertently lower others. Please also note that your subscores cannot be directly compared to your total score in that they are calculated differently and as such the pass score cannot be translated to Figure 1.

For more information, please visit the NAC examination web page.

* Figure 1. NAC Examination Score Profile



Report date: 2018-02-07

2/2

Aaaaaaaaa, Aaaaaaaaaaaaa / 0000000000