



A National
Assessment
Collaboration

NAC
EXAMINATION
ANNUAL
TECHNICAL
REPORT

2020

MARCH



MEDICAL COUNCIL
OF CANADA

LE CONSEIL MÉDICAL
DU CANADA

TABLE OF CONTENTS

OVERVIEW	4
1. EXAM DEVELOPMENT	6
Blueprint and test specifications	6
Exam content	7
Content validity.....	8
Exam format.....	10
Scoring candidate performance	10
2. EXAM ADMINISTRATION	12
Candidate orientation.....	12
Exam sites and candidate numbers	12
Exam administration staff.....	12
Standardized participant training	13
Chief examiners	13
Common PE recruitment requirements for all MCC exams.....	14
NAC examiner recruitment requirements.....	14
3. EXAM SCORING	16
Standard quality assurance and quality control procedures	16
Exam result approval	16
Exam result reporting	17
Scale scores.....	18
Step 1: Calculate total raw scores	18
Step 2: Linking	19
Step 3: Scale score transformation.....	20
Pass/Fail status.....	21
Domain subscores	21
4. PSYCHOMETRIC RESULTS	23
Scale scores.....	23
Estimates of score reliability and classification decisions.....	24
Cronbach alpha.....	24
Standard error of measurement.....	24
Decision accuracy and decision consistency.....	25
OSCE station statistics	26
Examiner analyses.....	27
Step One	27
Step Two	28
Step Three	28
Domain subscore profiles	29
Historical comparisons	30

REFERENCES	31
APPENDIX A:	
NAC Examination competency ratings	32
APPENDIX B:	
NAC Examination competency descriptors	33
APPENDIX C:	
NAC Examination Statement of Results (SOR) sample	34
APPENDIX D:	
NAC Examination Supplemental Information Report (SIR) sample	35

TABLES AND FIGURES

Table 1: Test specifications for the 2020 NAC Examination	7
Table 2: Sampling of OSCE content by test specifications for the March 2020 NAC Examination	9
Table 2 (cont.): Sampling of OSCE content by test specifications for the March 2020 NAC Examination	10
Table 3: NAC Examination candidate numbers for the March 2020 NAC Examination.....	12
Table 4: Summary statistics of scale scores for the March 2020 NAC Examination.....	23
Figure 1. Score distribution for the March 2020 NAC Examination	23
Table 5: Decision Consistency, Decision Accuracy, Reliability Estimate, and SEM for the March 2020 NAC Examination	25
Table 6: Summary Statistics for OSCE Stations for the March 2020 NAC Examination.....	27
Figure 2. Domain subscore profile for the March 2020 NAC Examination.....	29
Table 7: NAC Examination candidate performance data for March 2019 to March 2020.....	30

OVERVIEW

In response to the 2004 Report of the Canadian Task Force on Licensure of International Medical Graduates (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 international medical graduates (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 seeking a licence 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 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 March 2020, the exam was delivered in Alberta, British Columbia, Manitoba, Ontario, and Quebec, and was mandatory for application to the Canadian Resident Matching Service (CaRMS) in all provinces.

The National Assessment Central Coordinating Committee (“NAC³”)¹ is the governing body responsible for the oversight and function of the exam and reports to the MCC Executive Board.

The NAC Examination Committee (NEC), reporting to the NAC³, oversees

¹ The oversight and mandate for the NAC³ was completed July 2020.

the creation and maintenance of the exam content. The NEC ensures that all content adheres to the exam Blueprint, and that the overall exam content and format meet NAC guidelines. In addition, the NEC approves the release of results and adjudicates 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 end, the MCC has developed and continues to update a library of structured procedures that help maintain uniformity in administration across regions and sites. The MCC also provides the basis for support materials for standardized participants (SPs) and examiners. 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 exam and not to extraneous differences. Additionally, these policies and procedures are necessary for high-volume testing programs such as the NAC exam 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 the test form for the exam that took place in March 2020.

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. From 2011 to 2018, the steering committee's successor group, the NEC, maintained the original Blueprint except for the testing of therapeutic knowledge. In 2014, the NAC Therapeutics Examination (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 and Therapeutics competency. In 2015, the NEC struck a subcommittee to consider and recommend updates to the NAC exam. In 2019, those changes took effect. Those changes included the removal of Language Fluency and Organization as measured competencies, the use of key featured checklist items and a more streamlined scoring process took effect. See **Table 1** for the updated Blueprint and test specifications for the March 2020 exam.

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. Adhering to 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 and 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 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 the 2020 NAC Examination

DISCIPLINE	Recommended stations, No.	SYSTEM	Recommended stations, No.
Medicine	2–4	Respiratory	≥ 1
Surgery	2–4	Cardiovascular	≥ 1
Psychiatry	1–2	Gastrointestinal	≥ 1
OB/GYN ^a	1–2	Musculoskeletal	2-3
Pediatrics	1–2	Genitourinary	
Geriatric medicine	1–2	Endocrine	
Urgent care	1	Neurologic	
^a OB/GYN: Obstetrics and Gynecology		Mental health	2-3
		Reproductive Health	
		Multisystem	
CLINICAL COMPETENCY	Recommended stations, No.	AGE ^b	Recommended stations, No.
History taking	6–7	0–2 mo (newborn)	1-2
Physical examination	1	2–23 mo (infant)	
Combined history and physical examination	2–3	2–5 yr (preschool child)	
Communication skills	≥ 6	6–12 yr (child)	
Diagnosis	≥ 3	13–17 yr (adolescent)	1-2
Data interpretation	≥ 3	18–44 yr (young adult)	4-5
Investigations	≥ 3	45–64 yr (adult)	
Management ^c	≥ 3	≥ 65 yr (older adult)	2-3
		GENDER ^d	
		Of 10 stations, no more than 60% should be male or female	

^b AGE of actual patient, not necessarily SP's age

^c Up to 20% must be therapeutics-specific

^d GENDER of actual patient. not necessarily the SP's gender

Exam content

NAC exam content is developed by a panel of clinical subject matter experts along with experts in medical education and assessment. In this reporting year, there were several content development workshops where OSCE cases and stations were written, peer reviewed, and approved for piloting.

The 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 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 examiners, SPs and SP trainers regarding pilot stations is also reviewed, and, if necessary, adjustments are made to the stations. 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 specification).

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 a test form measures the required content and skills. Only one test form (Form 1) was administered in March 2020.

The NEC worked with MCC staff to select and approve the stations for a given test form. The test forms were drafted by the NAC test development officer (TDO) in accordance with the test specifications. The NEC then reviewed the test forms, including individual stations, to ensure that test specifications were met and that content was at the appropriate assessment level—that of a recent graduate from a Canadian medical school. The NEC approved the final version of the content for each test form. For security reasons, each exam sitting uses a different test form.

Table 2 shows the sampling of test specification characteristics, clinical competencies, and number of stations for test form 1 administered in March 2020. The “Recommended Stations” column specifies the desired number of stations for each test form for each clinical competency, discipline, gender, system, and age group.

Table 2: Sampling of OSCE content by test specifications for the March 2020 NAC Examination

		Recommended stations, No.	Form 1
CLINICAL COMPETENCY	History taking	6–7	6
	Physical examination	1	1
	Combined history taking and physical examination	2–3	3
	Communication skills	≥ 6	9
	Diagnosis	≥ 3	8
	Data interpretation	≥ 3	4
	Investigations	≥ 3	5
	Management ^a	≥ 3	9
DISCIPLINE	Medicine	2–4	7
	Surgery	2–4	2
	Psychiatry	1–2	2
	OB/GYN ^b	1–2	2
	Pediatrics	1–2	1
	Geriatric medicine	1–2	1
	Urgent care	1	3
GENDER^c	Of 10 stations, no more than 60% should be male or female		F = 5 M = 5
SYSTEM	Respiratory	≥ 1	2
	Cardiovascular	≥ 1	1
	Gastrointestinal	≥ 1	1
	Musculoskeletal	2–3	3
	Genitourinary		
	Endocrine		
	Neurologic	2–3	6
	Mental health		
	Reproductive health		
	Multisystem		

^a OB/GYN: Obstetrics and Gynecology

^b Up to 20% must be therapeutics-specific

^c GENDER of actual patient, not necessarily the SP's gender

Table 2 (cont.): Sampling of OSCE content by test specifications for the March 2020 NAC Examination

		Recommended stations, No.	Form 1
AGE ^d	0–2 mo (newborn)	1–2	1
	2–23 mo (infant)		
	2–5 yr (preschool child)		
	6–12 yr (child)		
	13–17 yr (adolescent)	1–2	1
	18–44 yr (young adult)	4–5	6
	45–64 yr (adult)		
	≥ 65 yr (older adult)	2–3	2

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

Exam format

For each administration, the NAC exam test form is composed of 10 operational 11-minute OSCE stations. The overall exam is designed to assess seven clinical competencies: communication skills, data interpretation, diagnosis, history taking, investigations, physical examination, and management.

In addition to completing 10 operational stations, candidates completed two pilot stations that did not count towards the final score.

In each station, a standardized participant (SP) portrayed the clinical scenario, and each candidate's performance was evaluated by an examiner. Each station measured up to seven clinical competencies.

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

Scoring candidate performance

Examiners rated candidate performance relative to the standard of a recent graduate from a Canadian medical school. The scoring tools use a

combination of short key-features checklists and rating scales.

The key features methodology gives score points to only the key steps a physician must take in order to manage the patient's issue effectively. Both the patient interaction component and the oral question component (if applicable by station) are scored in this key-features format.

Examiners also scored the candidates' proficiency on a number of competencies on a five-point Likert-type scale. The five rating points, along with a description of the acceptable performance level for each competency, are described in **Appendix A** and **Appendix B**.

Orientation and training materials were given to examiners to provide more specific context for these scoring tools.

Each station had one examiner, and, by the conclusion of the exam, each candidate had been evaluated by 12 examiners in 10 operational and two pilot stations. The scores from the 10 operational stations provided by each examiner were used to calculate all scores as described in Section 3, Exam Scoring.

2. EXAM ADMINISTRATION

This section describes procedures to standardize exam administration, including candidate orientation, responsibilities of exam administration staff, SP training, role of chief examiners (CEs), and examiner 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 registration information.

Exam sites and candidate numbers

The exam sites and the number of candidates for the March 2020 session are shown in **Table 3**.

Table 3: NAC Examination candidate numbers for the March 2020 NAC Examination

SITES	Total candidates, No.	First-time test takers, No.	Repeat test takers, No.
AB – Edmonton	435	316	119
BC – Vancouver			
MB – Winnipeg			
ON – Ottawa			
ON – Toronto			
QC – Montreal			

Abbreviations:

AB, Alberta; BC, British Columbia; MB, Manitoba; ON, Ontario; QC, Quebec.

Exam administration staff

Each exam site is responsible for recruiting 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, examiners, exam day staff, caterers). NAC policies and procedures provided by the MCC ensure the standardization of the exam administration. On exam days across the country, MCC staff oversee exam site staff either in person or via electronic communication; MCC also provides a telephone line for assistance.

Standardized participant training

Each site is responsible for hiring and supervising the SP trainers who, in turn, oversee the SPs to ensure 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 officer.

Chief examiners

All NAC exam sites employ physicians as chief examiners (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 PE recruitment requirements for all MCC exams

- Examiners or markers must be registered and in good standing with a medical regulatory authority in Canada
- Examiners or markers may be retired, but they must have an active licence with a medical regulatory authority in Canada
- Examiners or markers must be practising in Canada or they must have practised in Canada within the last five years
- All examiners and markers must adhere to the MCC Code of Business Conduct
- Examiners and markers must have the ability and stamina to complete their tasks

Any exceptions must be approved by the examination manager.

NAC examiner recruitment requirements

Examiners must meet all of the common examiner recruitment requirements for all MCC exams. Additionally, examiners for the NAC exam must meet the following requirements.

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

and

 - b. Non-licentiate examiners must be certified by one of the following organizations and must provide their certification number:
 - i. Royal College of Physicians and Surgeons of Canada (RCPSC).

- ii. Le Collège des Médecins du Québec (CMQ).
- iii. The College of Family Physicians of Canada (CFPC).

AND

- iv. 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 postgraduate year 1 residents and/or they must have experience as an examiner at this level of training.
 3. Physicians may be community physicians (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 postgraduate year 4 residents or higher **or** have a Certificate of the College of Family Physicians at the time of the examination.
 5. If retired, physicians must be within three years of having practised in Canada.

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

3. EXAM SCORING

In this section, we describe the quality assurance and quality control procedures related to the scoring of the NAC exam as well as what 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 quality assurance steps are performed as outlined below.

Examiners complete a score sheet for every candidate seen in their OSCE station. These score sheets are scanned at the exam sites and transmitted securely to the MCC. The MCC staff import the score sheets into TeleForm (OpenText), a form-processing program, where they are reviewed. Scanning anomalies are identified (for example, an unreadable candidate barcode, examiners' pencil marks that are too faint) and corrections are made. The data are then exported electronically into a scoring application for preliminary scoring and the results are used to generate a list of candidates who fall within 10 points above and below the pass score. Once the paper copies of the score sheets 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 score sheets. The updated electronic files are then reimported into the scoring application for final scoring and scale score transformation for all candidates. All scores are also calculated independently in parallel using the Statistical Analysis System (SAS) and compared with the results from the scoring application. All values must match before results are released to candidates.

Exam result approval

The results for each administration of the NAC exam are reviewed by the NEC. The NEC approves the release of results after each administration,

including special cases. Once the results have been approved by the NEC, they are imported to physiciansapply.ca and released to candidates.

When an incident occurs during the exam that may impact a candidate's performance, it is presented to the NEC as a special case. Depending on the nature of the incident (e.g., illness, fire alarm, SP misportrayal, a candidate's inappropriate behaviour), the NEC may decide to remove a station from a candidate's exam or award a candidate a No Standing result or a Denied result.

A No Standing result indicates that procedural irregularities in the examination process may have seriously affected the performance of the candidate and/or may have prevented a reliable assessment of the candidate's knowledge and abilities. A No Standing result does not count towards a candidate's number of attempts.

A Denied result indicates that a candidate has been found to have committed an infraction related to the MCC's examination process and/or breached confidentiality of the examination. A Denied result counts as an attempt towards a candidate's total number of attempts. Additionally, candidates that are given a Denied result may be denied eligibility to one or more future examinations of the MCC for a specified period.

Exam result reporting

Approximately one week after results are released to candidates, the MCC issues a Statement of Results (SOR) and a Supplemental Information Report (SIR) to each candidate through their physiciansapply.ca account (see **Appendix C** for an SOR example and **Appendix D** for an SIR example). The SOR includes the candidate's final result and total score, as well as the pass score. The SIR includes the candidate's final result, total score and additional information and graphics to display the candidate's domain subscores and comparative information.

The total score is reported on a standard-score scale ranging from 300 to 500. In contrast, the score profile in Figure 1 of the sample SIR in Appendix B

displays a candidate's domain subscores in terms of a percentage. As a result, total scores cannot be compared with domain subscores in the SIR as they are reported on different scales. Additionally, it is important to note that because subscores have fewer items than total scores, subscores have less measurement precision. Subscores are provided to individual candidates for feedback only and are not meant to be used by organizations for selection.

The following sections outline the steps in creating the results reported to candidates, IMG programs and the CaRMS.

Scale scores

The scale score is a candidate's total score reported on a scale that ranges from 300 to 500 (as opposed to a candidate's total raw score that is on a percentage metric). Deriving the scale score for the March 2020 NAC exam involves three steps.

Step 1: Calculate total raw scores

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

$$\text{station score} = \frac{\text{sum of a candidate's item scores}}{\text{sum of maximum possible item scores}} * 100$$

where the numerator is the sum of each candidate's scores on each item *i* for that station and the denominator is the sum of the maximum possible score for each item for that station. For example, a station with several checklist items, oral questions, and competency rating scales could result in the following score:

$$\text{station score} = \frac{1 + 0 + 1 + 1 + 4 + 0 + 3 + 2 + 3}{1 + 1 + 1 + 1 + 4 + 4 + 4 + 4 + 4} * 100 = \frac{15}{24} * 100 = 62.5$$

The station scores are then used to calculate the total raw score for each candidate using the following formula:

$$\text{total raw score} = (\text{sum of 10 station scores})/10$$

Since station scores are based on the sum of the candidate's item scores for that station, missing data needs to be taken into account so that it does not negatively impact a candidate's score. Missing data occurs when the examiner does not provide a score for an oral question or does not provide a rating for a competency for a given candidate on the score sheet. When this occurs, the station score is based on the item scores provided by the examiner.

In the above example, if the last item is missing from a candidate's score sheet, it is excluded from both numerator and denominator when calculating this candidate's station score as shown below.

$$\text{station score} = \frac{1 + 0 + 1 + 1 + 4 + 0 + 3 + 2}{1 + 1 + 1 + 1 + 4 + 4 + 4 + 4} * 100 = \frac{12}{20} * 100 = 60$$

The station score would have been 50 per cent if the missing item were treated as zero and the adjustment not applied. However, to be fair to the candidate, we exclude the missing item from the calculation of the station score and would use a station score of 60% instead.

Step 2: Linking

This step is to link through common stations the scores from the March 2020 test form to scores from previous test forms through a chain of linking steps dating back to test form 1 in 2019 that was used for setting the cut score and establishing the scale.

As described in *Section 1, Exam development*, multiple test forms are used each year for security reasons. All test forms are assembled based on the same Blueprint and test specifications, so they are as similar as possible in terms of content coverage. However, they may slightly differ in difficulty due to variations in clinical scenarios and tasks sampled on each test form.

The process of linking total scores statistically takes into account small differences in test form difficulty and adjusts total scores for the test form being linked so that all scores are on the same metric and can be compared. Linking also provides a way to apply the same pass score to candidates who

take different test forms.

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 anchor test or NEAT design. The subset of content that is presented identically is called an anchor set. The rule of thumb for determining the number of items in an anchor set for a multiple-choice exam is 20% of the total test or 20 items, whichever is greater, to ensure that the anchor set is representative of the total test in terms of content and difficulty. Since the NAC Exam is an OSCE with a small number of stations (less than 20), we use a 30% rule. The anchor set is used to statistically estimate the overall ability of candidates that took each test form and the difficulty of test form into account.

For the March 2020 NAC exam test form, an anchor set was based on three stations. A reference group of first-time test takers was used for all linking calculations. The linking calculations from this reference group is applied to all candidates to calculate each candidate's linked score. This linked score is then transformed as described in Step 3 below. For the linking steps, the Tucker observed-score method was employed (Kolen & Brennan, 2014). Full details of the method can be found in *Test equating, scaling, and linking: methods and practice* (3rd edition) authored by Kolen and Brennan (2014).

Step 3: Scale score transformation

This step is to convert the linked total scores for the March 2020 test form to scale scores that are reported to candidates and IMG programs. Once total scores are calculated and linked to the base test form, the linked scores are transformed into scale scores ranging from 300 to 500 for reporting purposes. The score scale was established using the March 2019 session results to have a mean of 400 and a standard deviation (SD) of 25. This final transformation ensures that any differences in scale score means and SDs on the current test forms can be directly compared to test form 1 of 2019. The final score transformation formula 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 2.48 based on the transformation of the March 2020 NAC exam, *intercept* is equal to 275.30 based on the transformation of the March 2020 NAC exam, and $LinkedScore_x$ is the linked score for candidate X.

All scale scores are rounded to a whole number between 300 and 500. The reported scale scores as seen by candidates is these rounded values. For example, a passing candidate with a linked score of 83.5 would have a scale score of 482:

$$ScaleScore_x = (2.48) * (83.50) + (275.30) = 482.38 \text{ rounded to } 482$$

A failing candidate with a linked score of 34.77 would result in a scale score of 362:

$$ScaleScore_x = (2.48) * (34.77) + (275.30) = 361.53 \text{ rounded to } 362$$

Pass/Fail status

The pass score for this exam was set by a panel of 21 physicians from across Canada, representing various specialties, demographics, and years of experience supervising students and residents. The panel recommended the pass score of 398 through a rigorous standard-setting exercise in April 2019. It was subsequently approved for implementation by the NEC in May 2019. Test form 1 from March 2019 was used to establish the pass score, and a contrasting group method was used for standard setting. Full details of the standard-setting exercise can be found in the *Technical report on the standard-setting exercise for the NAC Examination*. The established pass score of 398 was used to assign each candidate either a pass or fail status.

Domain subscores

Domain subscore calculations are used to create the figure in the candidates' SIRs. For each domain subscore, the associated items are converted to a percentage ranging from 0 to 100, where the total number of score points

obtained by a candidate is divided by the maximum score points per domain, multiplied by 100.

For example, if a candidate received scores of five, seven, eight and one on a domain with associated maximum scores of 10, 10, nine and one, respectively, the total number of score points obtained by the candidate is 21 and the maximum number of score points for this domain is 30. The domain subscore is $21/30 \times 100$ or 70.0. There are three subscores (reflecting three broad domains of physician activities) that are presented to candidates in their SIRs: Assessment & Diagnosis, Management, and Communication.

As a reminder, domain subscores should not be compared with scale scores as they are reported on different scales and because they have fewer items than the scale scores, they have less measurement precision than scale scores. Domain subscores are intended to provide general feedback to candidates on their relative strengths and weakness in their performance on the NAC Examination.

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 station quality and domain subscore profiles. Results reviewed and approved by the NEC following the March 2020 administration are used in this section, excluding candidates whose status is No Standing or Denied.

Scale scores

Summary statistics and the pass rate from March 2020 session are presented in **Table 4**. The score distribution is displayed in **Figure 1**. These statistics are based on the scale scores reported to candidates. The minimum, maximum, and SD are indicators of the variation in scale scores.

Table 4: Summary statistics of scale scores for the March 2020 NAC Examination

	Candidates, No.	Min. score	Max. score	Mean score	Median score	SD	Pass rate, %
Test form 1^a	434	305.3	474.1	405.8	406.0	26.7	64

^a One special case was excluded from calculations, except the pass rate

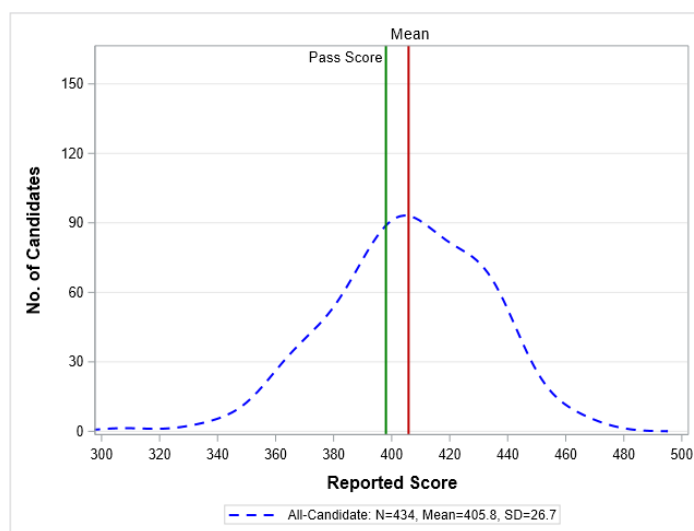


Figure 1. Score distribution for the March 2020 NAC Examination

Estimates of score reliability and classification decisions

Table 5 shows the reliability estimates, the standard error of measurement (SEM), the decision consistency and decision accuracy estimates and the associated false positives and false negatives by test form.

Cronbach alpha

Cronbach alpha was used to estimate score reliability for the NAC exam. A score reliability estimate indicates the desired consistency (or reproducibility) of exam scores across replications of measurement (Crocker & Algina, 1986; Haertel, 2006). Scores that are highly reliable are reproducible and consistent from one testing occasion to another. In other words, if the testing process was repeated with a group of test takers, essentially the same results would be obtained. This reliability estimate is described in Educational Measurement by Haertel in section 2.4.4 (Haertel, 2006). The formula for Cronbach 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 score variance for station i , and σ_X^2 is the variance of the total scores (Haertel, 2006, p. 74). As a rule, a reliability estimate greater than 0.80 on an OSCE is desirable. The reliability estimate in conjunction with the total exam standard error of measurement provides further evidence of the reliability of the candidate's scale score.

Standard error of measurement

The SEM provides a value that can be used to construct a confidence range (for example, +/- 1 SEM and +/- 2 SEM represent 68% and 95%, respectively) within which a candidate's observed score is expected to fluctuate if the candidate was to repeat the exam over and over again. The SEM value 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 SD for the total score (square root of the variance) and ${}_a\rho_{XX'}$ is defined as the reliability estimate as shown above.

Decision accuracy and decision consistency

Estimates indicating the consistency and accuracy of pass/fail decisions are important in providing validity and reliability evidence for candidate scores on one test form with possible equivalent test forms. To this end, the NAC exam uses the Livingston and Lewis (1995) procedure. Decision consistency is an estimate of the agreement between classifications on potential parallel test forms, and decision accuracy is the estimate of agreement between the observed classifications of candidates and those based on their true score (i.e., observed score \pm measurement error). Ideally, both values should be high, such as 0.80 and above, suggesting reliable and valid pass/fail classifications.

Table 5 shows the reliability estimate, the SEM, decision consistency, and decision accuracy estimates along with associated false positive and false negative rates for March 2020 Form 1. The estimated false positive rates 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 scores but who should pass based on their true ability.

Table 5: Decision consistency, decision accuracy, reliability estimate, and SEM for the March 2020 NAC Examination

	Form 1
Decision consistency	0.78
False-positive	0.11
False-negative	0.11
Decision accuracy	0.84
False-positive	0.06
False-negative	0.09
Reliability estimate	0.70
SEM (scale score)	14.74

Abbreviation: SEM, standard error of measurement

Reliability is impacted both by the amount of variability in scores among candidates taking a particular test form and the number of items or stations included in any given exam. It is more difficult to obtain reliability estimates above 0.80 given the restricted number of stations that can be administered in any OSCE test form.

OSCE station statistics

Summary statistics for each of the OSCE stations for March 2020 are provided in **Table 6**. The percentage of missing data, average station scores or p-values, SD of station scores, and station total correlations (STCs) are presented. Please refer to *Section 3, Exam Scoring, Step 1: Calculate total raw scores*, for calculation of station scores.

P-values are the average station scores that candidates achieved on each of the stations. In general, p-values indicate station difficulty and 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. That is, comparisons of p-values across different samples of candidates do not take into account potential differences in overall candidate ability. As such, 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.

SDs indicate the general variability of scores on any given station. STCs are indicators of discrimination between low- and high-ability candidates for a given station. A low positive or negative STC (< 0.30) indicates that there is a weak or negative relationship between the station score and the overall exam score. Along with the p-values, this information is useful in flagging stations that should be reviewed by content experts and possibly removed from scoring. A moderate-to-high STC (≥ 0.30) indicates that high-ability candidates are performing well on a given station. Flagged and reviewed stations may still be included on an exam when the content is deemed relevant, important, and has been verified to be correct.

Table 6: Summary Statistics for OSCE Stations for the March 2020 NAC Examination

STATION	Form 1			
	Missing data, %	Mean p-value	SD	STC
1	0.23	0.54	0.21	0.32
2	1.11	0.48	0.19	0.43
3	-	0.47	0.21	0.37
5	0.72	0.45	0.24	0.40
6	0.06	0.60	0.27	0.27
7	0.09	0.66	0.21	0.37
9	-	0.62	0.21	0.36
10	0.21	0.57	0.21	0.31
11	-	0.49	0.20	0.28
12	0.39	0.48	0.24	0.44
Mean	0.31	0.54	0.22	0.36

Abbreviations: STC, station total correlation; SD, standard deviation

^a Stations 4 and 8 were wait stations (no encounters)

Table 6 shows the means p-values for the March 2020 administration. There were no stations flagged as being too difficult (p-value < 0.30) or too easy (p-value > 0.90). Stations with an STC of < 0.30 were reviewed for content appropriateness. All of the reviewed stations were deemed to be important and acceptable from a content perspective..

Examiner analyses

Examiner analyses are conducted routinely for each of the OSCE stations. The examiner analyses are based on the method outlined by Bartman, Smee and Roy (2013). For the examiner analyses, the following steps are followed:

Step One

For each examiner and station scored by the examiner, the average across the candidates' station scores is calculated. This average is the examiner average for that station. Then the average of the examiner averages is calculated along with the SD. Examiners that scored fewer than 10

candidates on a station are excluded from these analyses as they have observed too few candidates to be compared to other examiners. Examiners are flagged as being a “dove” if their station score is higher than three times the station SD from the station average. Examiners are flagged as being a “hawk” if their station score is lower than three times the station SD from the station average. For example, if the average across examiner averages was 72.5 and the SD across examiners was 6.5 and an examiner had an average of 50.7 (difference of 21.8, which is more than three SDs [$6.5 \times 3 = 19.5$]) then they are flagged as a hawk.

Step Two

In step two, for each examiner flagged in step one, the station distribution (histogram) for the examiner is compared to the distribution of station scores from other examiners across the country. This is a visual check to evaluate whether the examiner is providing a range of scores that looks somewhat normally distributed (not providing all high or low scores). If an examiner’s distribution looks reasonable, they are no longer flagged at this step as being either a dove or hawk.

Step Three

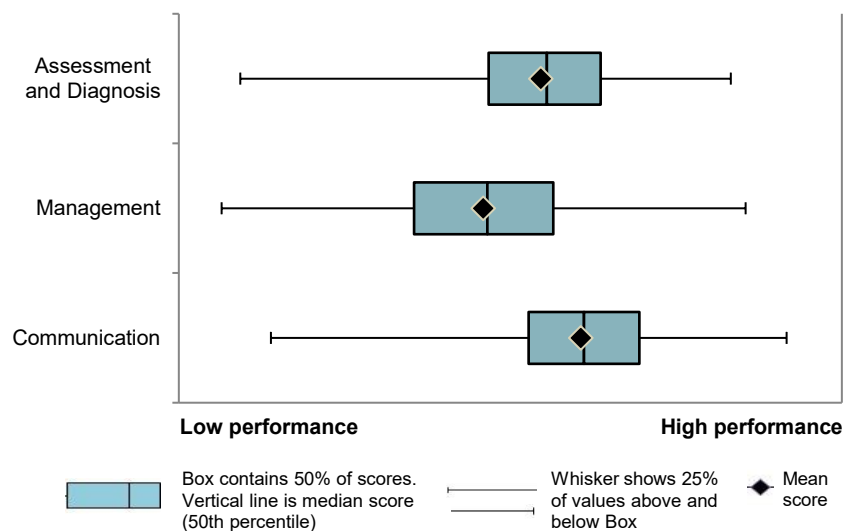
In step three, for each examiner flagged in steps one and two, the scale-score distribution (histogram) for the cohort they scored is compared with the distribution of scale scores based on the candidates across the country. This is a check that the cohort’s average scale-scores and pass rate based on all 10 examiners is higher or lower than the values across the country. In this step, we evaluate if a cohort may be higher or lower in ability that may explain a dove or hawk flag in step one. For example, an examiner may be flagged as a hawk in steps one and two, but the candidates’ scale-scores based on all 10 stations may be lower, indicating a weaker cohort. Thus, the examiner would not be flagged as a hawk at step three.

There were no examiners flagged across all three steps for March 2020.

Domain subscore profiles

The purpose of the domain subscore profile is to provide general feedback to candidates by highlighting their relative strengths and weaknesses on three broad categories of physician activities assessed by the NAC exam. A domain subscore profile is presented in the form of a graph to each candidate in the SIR (see Appendix B for a sample SIR). The graph shows the domain subscore for each of the three domains and the SEM around the domain subscore. The calculation of the domain subscores for each candidate is outlined in Section 3 of this report.

This section provides domain subscore profiles for March 2020. The range of domain subscores is shown graphically in Figure 2. The boxes for each domain indicate the range for 50% of candidates' domain subscores. The vertical line represents the median or 50th percentile domain subscore. The remaining 25% of domain subscores are shown to the right or left of the box as a line (25% to the right and 25% to the left). The mean domain subscore is indicated by the diamond.



The whiskers indicate 25% of values above and below the Box. The left and right ends of the boxes indicate the interquartile range. Box contains 50% of subscores. The vertical line inside the boxes is the median subscore (50th percentile). The diamond indicates the mean subscore

Figure 2. Domain subscore profile for the March 2020 NAC Examination

Historical comparisons

Table 7 presents candidate performance data for the total group, first-time test takers, and repeat test takers since March 2019 when the new Blueprint was implemented. Data prior to 2019 are not included as the previous NAC exam was very different in terms of Blueprint, format, scoring approach, pass score, and score scale. For historical data on the previous NAC exam prior to 2019, refer to the *2018 NAC Annual Technical Report*.

Table 7: NAC Examination candidate performance data for March 2019 to March 2020

Session	First-time test takers		Repeat test takers		Total test takers	
	Candidates, No.	Pass rate, %	Candidates, No.	Pass rate, %	Candidates, No.	Pass rate, %
March 2019	342	53.5	70	61.4	412	54.9
Sept. 2019	939	67.7	354	68.6	1293	68.0
Total 2019	1281	63.9	424	67.5	1705	64.8
March 2020	315	60.6	119	71.4	434	63.6

* One special case was excluded from the calculation.

REFERENCES

- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich College.
- Federal / Provincial / Territorial Advisory Committee on Health Delivery and Human Resources. (2004, February). *Report from the Canadian Task Force on Licensure of International Medical Graduates*.
- Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Praeger.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). 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. (2014). *Test equating, scaling, and linking: Methods and practice* (3rd ed.). Springer Science + Business Media.
- 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.
- Medical Council of Canada. (2019). *Technical report on the standard setting exercise for the NAC Examination*.

APPENDIX A:

NAC Examination competency ratings



COMPETENCY RATINGS

Based on this interaction, please rate *this candidate's performance in the following competencies* as compared to a recent Canadian graduate accepted into postgraduate training (for rating scale anchors, refer to RATING SCALE CRITERIA page).

QUALITY OF HISTORY TAKING

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Acquires from the patient, family or other source a chronologic, medically logical description of pertinent events, including questioning about onset, location, duration, character, severity, etc., as appropriate to the case. Gathers information efficiently in sufficient breadth and depth to permit a clear definition of the patient's problem(s).

DIAGNOSIS

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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

MANAGEMENT

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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, counselling, follow-up); considers risks and benefits of therapy and instructs the patient accordingly. Identifies medication classes, except when specific drugs and dosages would reasonably be expected in the context of the clinical problem.

COMMUNICATION SKILLS

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Uses a patient-centred approach: establishes trust and respect and shows sensitivity to the patient's needs. Provides clear information and confirms patient's understanding: encourages questions and uses repetition and summarizing to confirm and/or reinforce understanding. Respects confidentiality when appropriate. Avoids use of jargon/slang and uses tone and vocabulary appropriate to the patient. Demonstrates appropriate non-verbal communication (e.g., eye contact, gesture, posture and use of silence).

QUALITY OF PHYSICAL EXAMINATION

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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). Demonstrates sensitivity to the patient's comfort and modesty; explains actions to the patient.

INVESTIGATIONS

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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

DATA INTERPRETATION

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

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

APPENDIX B:

NAC Examination competency descriptors



COMPETENCY DESCRIPTORS

Based on this interaction, please rate *THE QUALITY OF THIS CANDIDATE'S PERFORMANCE IN THE FOLLOWING COMPETENCIES* as compared to a recent Canadian graduate accepted into post-graduate training (for rating scale anchors, refer to RATING SCALE CRITERIA page).

UNACCEPTABLE as compared to a recent Canadian graduate accepted into postgraduate training	BORDERLINE UNACCEPTABLE as compared to a recent Canadian graduate accepted into postgraduate training	BORDERLINE ACCEPTABLE as compared to a recent Canadian graduate accepted into postgraduate training	ACCEPTABLE as compared to a recent Canadian graduate accepted into postgraduate training	ABOVE the level expected of a recent Canadian graduate accepted into postgraduate training
○	○	○	○	○

QUALITY OF HISTORY TAKING

Acquires from the patient, family or other source a chronologic, medically logical description of pertinent events, including questioning about onset, location, duration, character, severity, etc. as appropriate to the case. Gathers information efficiently in sufficient breadth and depth to permit a clear definition of the patient's problem(s).

DIAGNOSIS

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

MANAGEMENT

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, counselling, follow-up); considers risks and benefits of therapy and instructs the patient accordingly. Identifies medication classes, except when specific drugs and dosages would reasonably be expected in the context of the clinical problem.

COMMUNICATION SKILLS

Uses a patient-centered approach: establishes trust and respect and shows sensitivity to the patient's needs. Provides clear information and confirms patient's understanding; encourages questions and uses repetition and summarizing to confirm and/or reinforce understanding. Respects confidentiality when appropriate. Avoids use of jargon/slang and uses tone and vocabulary appropriate to the patient. Demonstrates appropriate non-verbal communication (e.g., eye contact, gesture, posture and use of silence).

QUALITY OF PHYSICAL EXAMINATION

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.

INVESTIGATIONS

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

DATA INTERPRETATION

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

APPENDIX C: NAC Examination Statement of Results (SOR) sample



National Assessment Collaboration Examination Statement of Results

Candidate name: Aaaaa, Aaa Bbbbbbbb Cccc
Candidate code: 0123456789
Examination session: March 2020 **Your final result:** Pass
Pass score: 398 **Your total score:** 474

April 17, 2020

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

Your total score is reported as a scaled score ranging from 300 to 500 with a mean of 400 and a standard deviation of 25. The mean and standard deviation were set using the results from the March 2019 session.

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 D: NAC Examination Supplemental Information Report (SIR) sample



National Assessment Collaboration Examination Supplemental Information Report

Candidate name: Aaaaa, Aaa Bbbbbbbb Cccc
Candidate code: 0123456789 **Your final result:** Pass
Examination session: March 2020 **Your total score:** 474

This report provides you with supplemental information on your performance on the National Assessment Collaboration (NAC) Examination.

The NAC Examination assesses core abilities to apply medical knowledge, demonstrate clinical skills, develop investigational and therapeutic clinical plans, as well as demonstrate communication skills at a level expected of a medical graduate entering into postgraduate training in Canada.

The exam assesses your performance across three broad domains that reflect a physician's scope of practice as indicated in the following table. Each domain is assigned a weighting on the exam and the content weights are expressed as percentages.

Domains	Weighting (%)
Assessment and Diagnosis	70 ± 5
Management	15 ± 5
Communication Skills	15 ± 5

See p. 2 of this report for the domain definitions.

Figure 1 displays your performance in each domain. We provide your subscores along with the mean subscore of first-time takers who passed the same exam. We also provide the standard error of measurement (SEM) for each of your subscores. It represents the expected variation in your subscore if you were to take this exam again with a different set of questions covering the same domains.

Small differences in subscores or overlap between SEMs indicate that performance in those domains was somewhat similar. Overlap between the SEM and the mean score of first-time takers who passed signifies that performance is similar to the mean.

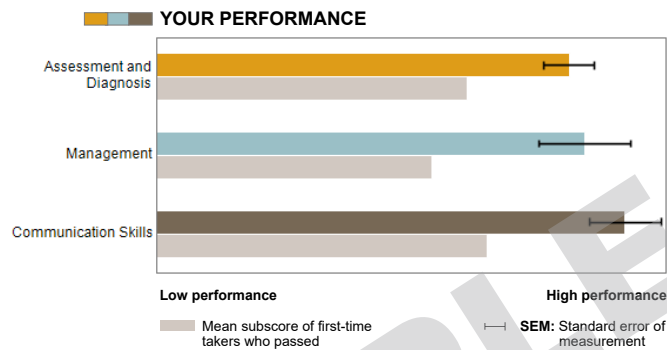
Subscores are based on less data than the total score and have less precision. Your total score and subscores cannot be compared as they are calculated differently. The pass score cannot be applied to Figure 1.

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

 mcc.ca
physiciansapply.ca
inscriptionmed.ca

APPENDIX D: NAC Examination Supplemental Information Report (SIR) sample

Figure 1: NAC Examination score profile



The following defines the three domains assessed by the exam:

- **ASSESSMENT AND DIAGNOSIS** covers the following physician activities:
 - **History Taking:** 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 problems.
 - **Physical Examination:** Elicits physical findings in an efficient logical sequence that documents the presence or absence of abnormalities, and supports a definition of the patient's problems; sensitive to the patient's comfort and modesty; explains actions to the patient.
 - **Diagnosis:** Discriminates important from unimportant information and reaches a reasonable differential diagnosis and/or diagnosis.
 - **Data Interpretation:** Interprets investigative data appropriately in the context of the patient's problems.
 - **Investigation:** Selects suitable laboratory or diagnostic studies to elucidate or confirm the diagnosis; takes into consideration associated risks and benefits.

MANAGEMENT: 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.

COMMUNICATION SKILLS: Uses a patient-centered approach; establishes trust and respect, and shows sensitivity to the patient's needs; provides clear information; confirms patient's understanding (encourages questions, and uses repetition and summarizing to confirm and/or reinforce understanding); respects confidentiality when appropriate; speaks clearly (volume and rate); avoids use of jargon/slang and uses vocabulary appropriate to the patient; demonstrates appropriate non-verbal communication (e.g., eye contact, gesture, posture and use of silence).