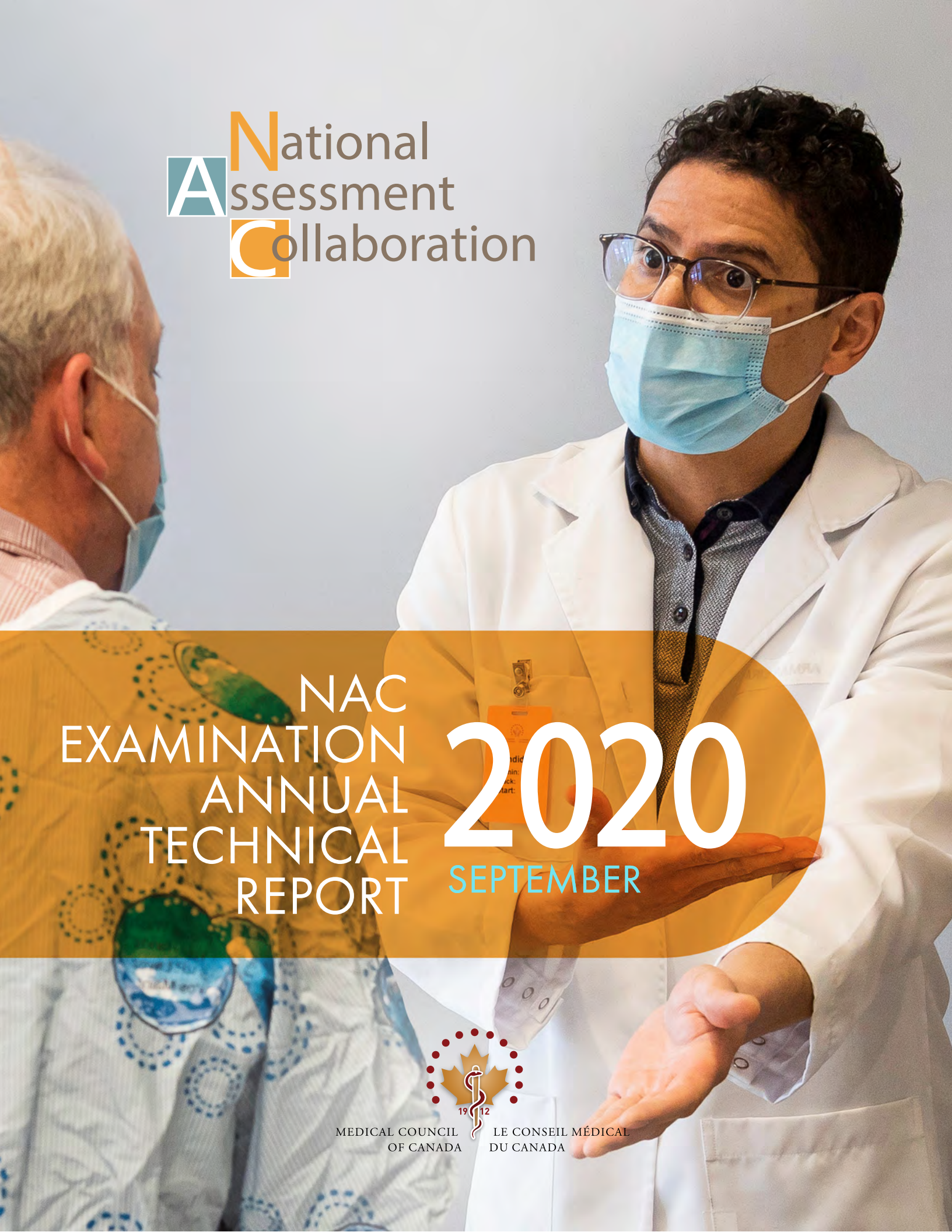
National Assessment Collaboration

NAC EXAMINATION ANNUAL TECHNICAL REPORT

2020

SEPTEMBER



MEDICAL COUNCIL OF CANADA
LE CONSEIL MÉDICAL DU CANADA

TABLE OF CONTENTS

TABLES AND FIGURES

OVERVIEW

In In response to the 2004 Report From 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 IMGs 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 NAC Examination 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 September 2020, the NAC exam was delivered in 11 sites 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 NAC Examination Committee (NEC)[1] oversaw the creation and maintenance of the NAC exam content. The NEC ensured that all content adheres to the NAC exam Blueprint, and that the overall exam content and

[1] The NEC was amalgamated with another MCC committee, the Central Examination Committee, in 2021, called the Exam Oversight Committee.

format meet NAC guidelines. In addition, the NEC approved the release of results, and adjudicated 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, as well as provide 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 NAC exam and not to extraneous differences. Additionally, these policies and procedures are necessary for high-volume testing programs, such as the NAC, for which 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 two test forms for the NAC exam that took place September 2020.

Due to the COVID-19 pandemic, changes to the structure of the content, format, and delivery were made to the September 2020 NAC exam to ensure the safety of all involved in the exam. A separate report is provided specifically for this session. For information on the March 2020 NAC exam, please refer to the *NAC Examination Annual Technical Report – 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 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 & Therapeutics competency. In 2015, the NEC called 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. See **Table 1** for the updated Blueprint and test specifications.

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/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 September 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 participant, not necessarily the SP's age
[c] Up to 20% must be therapeutics-specific
[d] GENDER of actual participant, not necessarily the SP's gender

Content changes due to COVID-19 protocols

Under normal circumstances, in addition to completing 10 operational stations, candidates would complete two pilot stations that did not count towards the final score. However, the pilot stations were replaced with wait stations for the September 2020 exam session.

Additionally, stations that included a physical examination, in which the candidates would normally demonstrate their physical examination skills by

placing their hands on the SP, were adjusted to a “described” or “verbalized” physical examination. Candidates were asked to tell the examiner what physical examination manoeuvre they would perform and describe what they were examining and why. Then the examiner would verbally provide physical examination findings as appropriate. The “normal” Blueprint constraints call for one station with a physical examination only (no history-taking component), and for the purposes of the September 2020 NAC, that station was removed and replaced by a combined history and physical examination station.

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 in which OSCE cases/stations were written, peer-reviewed, and approved for piloting.

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; Kane, 2013). This section highlights the test specifications and how well each test form measures the required content and skills.

The NEC works with MCC staff to select and approve the stations for a given test form. The test forms are drafted by the NAC test development officer (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, each exam sitting uses a different test form.

Table 2 shows the sampling of test specification characteristics, clinical competencies, and number of stations for each form. The "Recommended Stations" column specifies the desired number of stations for each test form for each clinical competency, discipline, gender, system, and age group. There were two test forms administered in September (Forms 2 and 3).

***Table 2:** Sampling of OSCE content by test specifications for the September 2020 NAC Examination test forms*

		Recommended stations, No.	Form 2	Form 3
CLINICAL COMPETENCY	History taking	6–7	7	7
	Physical examination	1	0	0
	Combined history taking and physical examination	2–3	3	3
	Communication skills	≥ 6	10	10
	Diagnosis	≥ 3	8	8
	Data interpretation	≥ 3	5	3
	Investigations	≥ 3	3	5
	Management[a]	≥ 3	10	8
DISCIPLINE	Medicine	2–4	6	7
	Surgery	2–4	3	2
	Psychiatry	1–2	1	1
	OB/GYN[b]	1–2	2	1
	Pediatrics	1–2	2	1
	Geriatric medicine	1–2	1	2
	Urgent care	1	5	1
GENDER[c]	Of 10 stations, no more than 60% should be female or male		F = 6 M = 4	F = 5 M = 5
SYSTEM	Respiratory	≥ 1	2	1
	Cardiovascular	≥ 1	2	2
	Gastrointestinal	≥ 1	1	1
	Musculoskeletal	2–3	4	5
	Genitourinary			
	Endocrine			
	Neurologic			
	Mental health	2–3	5	7
	Reproductive health			
	Multisystem			

[a] *Up to 20% must be therapeutics-specific*
[b] OB/GYN: Obstetrics and Gynecology
[c] *GENDER of actual participant, not necessarily the SP's gender*

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

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

[d] *Age of actual participant, not necessarily the SP's age.*

Exam format

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

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 featured checklists and rating scales.

The key features methodology gives score points to only the critical or key steps a physician must take to manage the patient's problem effectively. Both the patient interaction component and the oral question component (if applicable by station) are scored in this key featured 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 examiners on 10 operational stations. The scores from the 10 operational stations provided by each examiner were used to calculate all scores as described in the Exam Scoring section.

2. EXAM ADMINISTRATION

This section describes procedures to standardize exam administration, including candidate orientation, responsibilities of exam administration staff, Standardized Participant (SP) training, role of Chief Examiners (CEs), and examiner recruitment and training.

Exam sites and candidate numbers

The exam sites and the number of candidates for each test form in September 2020 session are depicted in **Table 3**.

Table 3: NAC candidate numbers by test form for the September 2020 administration

	SITES	Total candidates, No.	First-time test takers, No.	Repeat test takers, No.
Test form 2	AB – Calgary	671	492	179
	– Edmonton			
	BC – Vancouver			
	MB – Winnipeg			
	ON – London			
	– Ottawa			
	– Sudbury			
	– Toronto			
	QC – Montreal			
	SK – Saskatoon			
Test form 3	AB – Calgary	569	425	144
	– Edmonton			
	MB – Winnipeg			
	NS – Halifax			
	ON – London			
	– Ottawa			
	– Sudbury			
	– Toronto			
	QC – Montreal			
	SK – Saskatoon			
Total		**1240**	**917**	**323**

Abbreviations: AB, Alberta; BC, British Columbia; MB, Manitoba; NS, Nova Scotia; ON, Ontario; QC, Quebec; SK, Saskatchewan.

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.

For the September 2020 exam sessions, candidate orientations were online. Candidates were not given a face-to-face orientation but were given exam-day reminders and a chance to ask questions prior to the beginning of the exam.

NAC administration under COVID-19 protocols

The MCC worked closely with the examination sites throughout the summer of 2020. As a result, several changes were implemented to the delivery and administration of the NAC examination to ensure the health and safety of all exam participants.

Social distancing measures and proper usage of personal protective equipment (PPE) were enforced throughout the examination day, including all encounters. All participants wore face masks covering their nose and mouth, adhered to sanitizing protocols, and had to sign and pass a COVID-19 screening questionnaire. Hand sanitizer bottles were placed in each station and strategically placed throughout the exam sites, and support staff would sanitize doorknobs, pencils, and/or any other items/objects that would be regularly touched.

To limit the traffic of individuals at the exam sites, SP rotations were removed, sites issued staggered track arrival times, candidate catering services and sequestering were removed, orientations and training were implemented online, and groups were registered and deregistered individually to avoid contact.

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. 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 Participant 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 officers.

For the September 2020 sessions of the exam, SPs were trained, and the dry runs were conducted online.

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:

- Assisting with examiner recruitment and training, if needed
- 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
- Overseeing examiners and candidates on exam day
- Addressing, where appropriate, candidates' questions, concerns, and complaints on exam day
- 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 examiner 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 or markers must have the ability and stamina to complete the task (e.g., uncorrected hearing loss can seriously affect the ability to score an exam)

All 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:

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:

- Nonlicentiate 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
    - Collège des Médecins du Québec
    - 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
- Physicians must have recent experience supervising clerks and/or

postgraduate training year 1 residents, and/or they must have experience as an examiner at this level of training

- Physicians may be community physicians (i.e., they do not need to be faculty members if all other criteria are met)
- Physicians must be currently practising medicine in Canada; if they are a resident physician, they must be postgraduate training year 4 residents or higher or have a Certificate of the College of Family Physicians at the time of the examination
- If retired, physicians must be within three years of practising 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.

For the September 2020 sessions of the exam, the examiner orientations were modified to be completed online. Examiners were given exam day reminders and had an opportunity to ask questions prior to the beginning of the exam.

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 two lists of candidates: (1) those who fell within 10 points above and below the cut score for pass with superior performance and (2) those who fell within 10 points above and below the cut score for pass/fail. 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, or a candidate's inappropriate behaviour), the NEC may decide to remove a station from a candidate's exam, award a candidate a No Standing 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 to each candidate through their physicianapply.ca account. The Statement of Results includes the candidate's final result. For the September 2020 administration, no total score or subscores were reported. Only three categories of results were reported to candidates: pass, fail, and pass with superior performance. Due to content, format, and delivery changes that are specific to the September 2020 NAC exam, we could not statistically link the total scores to those on the previous exams (i.e., the NAC exam implemented between March 2019 and March 2020).

Establishing the standard for determining exam status

Because of the changes made to the September 2020 NAC exam, we could not apply the same previously established NAC pass score to the September 2020 cohort.

A standard-setting exercise was conducted by the MCC with a panel of 20 physicians from across the country, representing various specialties, demographics, and years of experience supervising students and residents. They met virtually on October 19–21, 2020, to establish the standards for

differentiating between “fail,” “pass,” and “pass with superior performance.” Standard setting is a process used to define an acceptable level of performance and to establish a cut score for one or more target levels of performance. The cut scores for the September 2020 NAC exam were established using the same method (Contrasting Group method) and process that were previously used in March 2019 (Cizek & Bunch, 2007). The cut scores were established using one test form and were applied all test forms through statistical linking. Please note that, because total scores were not reported to candidates, the established cut scores were not published.

The established cut scores were used to assign each candidate from the September 2020 administration either a pass or fail or pass with superior performance status. Pass indicates that a candidate’s performance on this examination met the minimum standard required for entry into a Canadian residency program. Pass with superior performance indicates that a candidate has met the minimum standard required for entry into a Canadian residency program with a stronger performance on this examination. The reporting of the new pass with superior performance was intended to help to offset the absence of a total score and help support informed decision making for stakeholders who require NAC exam results for their programs.

4. PSYCHOMETRIC RESULTS

This section provides a performance summary of candidates who took the NAC exam in September 2020, as well as estimates of reliability and classification decisions, and a summary of station quality. Results reviewed and approved by the NEC following the September 2020 administration are used in this section, excluding candidates whose status is No Standing or Denied or who missed more than one station. Candidate performance is summarized in **Table 4**.

Table 4: Pass rates for the September 2020 administration of the NAC Examination

			Pass rate by performance level, %	
	Candidates, No.	Pass rate, %	Pass	PWSP
Total[a]	1238	83.4	63.4	20.0
First-time test takers	916	82.2	61.9	20.3
Repeat test takers[b]	322	86.6	67.7	18.9

Abbreviation: PWSP, pass with superior performance.

[a] *Excluding 2 candidates who received a No Standing.*

[b] *Repeat test takers include candidates who previously failed the NAC exam, as well as some candidates who previously passed the NAC exam.*

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 along with 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 accurate, 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 general rule, a reliability estimate greater than 0.80 on an OSCE is desirable. The

reliability estimate in conjunction with the total exam SEM 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 ${}_{\alpha}\rho_{XX'}$ 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 Livingston and Lewis (1995) procedure. Decision consistency is an estimate of 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 ± measurement error).

Ideally, both of these values should be high, such as 0.80 and above, suggesting reliable and valid pass/fail classifications.

Table 5 shows the reliability estimates, the SEM, and decision consistency, and decision accuracy estimates along with associated false-positive and false-negative rates for September 2020 test forms. 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 test form and performance level for the September 2020 administration of the NAC Examination*

	Form 2		Form 3	
	Pass	PWSP	Pass	PWSP
Decision consistency	0.86	0.79	0.88	0.74
False-positive	0.07	0.11	0.07	0.13
False-negative	0.07	0.10	0.06	0.13
Decision accuracy	0.91	0.85	0.92	0.80
False-positive	0.03	0.11	0.03	0.14
False-negative	0.06	0.04	0.06	0.06
Reliability estimate	0.71		0.63	
SEM (scale score)	13.62		15.17	

Abbreviation: PWSP, pass with superior performance
SEM, standard error of measurement.

It should be noted that reliability is impacted both by the amount of variability in scores amongst 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 by form 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.

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 population 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-ability 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 in the September 2020 Administration of the NAC Examination*

	Form 2				Form 3			
STATION[a]	**Missing data, %**	**p-value**	**SD**	**STC**	**Missing data, %**	**p-value**	**SD**	**STC**
1	0.66	0.58	0.19	0.40	0.50	0.60	0.18	0.33
2	0.43	0.64	0.23	0.34	0.50	0.64	0.23	0.34
3	0.31	0.49	0.19	0.31	0.00	0.60	0.20	0.23
5	0.00	0.55	0.19	0.33	0.40	0.45	0.23	0.31
6	1.62	0.57	0.19	0.31	0.18	0.58	0.17	0.35
7	0.30	0.50	0.23	0.37	0.11	0.76	0.18	0.32
9	0.18	0.62	0.18	0.31	0.00	0.68	0.17	0.29
10	0.21	0.61	0.22	0.37	2.51	0.53	0.22	0.33
11	0.30	0.67	0.19	0.39	0.18	0.70	0.18	0.35
12	0.13	0.56	0.23	0.44	0.11	0.53	0.27	0.19
Mean	**0.41**	**0.58**	**0.20**	**0.36**	**0.45**	**0.61**	**0.20**	**0.30**

Abbreviations: STC, station total correlation

[a] Stations 4 and 8 are not shown because they are wait stations (no encounters)

Table 6 shows the means p-values for the September 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 < 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 mean across the candidates' station scores is calculated. This mean is the examiner mean for that station. Then the mean of the examiner means 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 with 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 mean. For example, if the mean across examiner means was 72.5 and the SD across examiners was 6.5 and an examiner had a mean of 50.7 [difference of 21.8, which is more than three SDs (6.5 × 3 = 19.5)] then that examiner is 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 with 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 Step One and Step 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 mean 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 being 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 September 2020.

5. REFERENCES

Bartman, I., Smee, S., & Roy, M. (2013, February). *A method for identifying extreme OSCE examiners*. *Clinical Teacher*, *10*(1), 27–31.

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.

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. (2022). *Technical report on the 2020 standard-setting exercise for the NAC Examination.*

APPENDIX A:
NAC EXAMINATION COMPETENCY RATINGS

National Assessment Collaboration

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
○	○	○	○	○

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
○	○	○	○	○

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

MANAGEMENT

UNACCEPTABLE	BORDERLINE UNACCEPTABLE	BORDERLINE ACCEPTABLE	ACCEPTABLE	ABOVE
○	○	○	○	○

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
○	○	○	○	○

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
○	○	○	○	○

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
○	○	○	○	○

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
○	○	○	○	○

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

APPENDIX B:
NAC EXAMINATION COMPETENCY DESCRIPTORS

National Assessment Collaboration

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).