NBME®

Guide to the

Customized Assessment Services

Create Examinations Tailored to Your Curriculum

Table of Contents

Overview and Policies for Use of Customized Assessment Services

The National Board of Medical Examiners® (NBME®) introduced the Customized Assessment Services (CAS) program in 2007 in response to the trend toward integrated preclinical curricula at many US medical schools. CAS allows faculty to build high-quality assessments tailored to local discipline-based, organ-system, problem-based or other curricular approaches. The NBME has assembled a large pool of items covering the topics commonly taught and tested in basic science coursework. These items have been permanently transferred from the United States Medical Licensing Examination® (USMLE®) pool, and therefore have undergone the same rigorous quality assurance protocols as all USMLE test items.

Test construction services are available through the secure *NBME Services Portal (NSP)*, the primary source for authorized medical school users to obtain information about their students for a variety of activities related to USMLE, NBME subject examinations, and other services.

An annual subscription fee for CAS covers:
- program setup
- support for test construction and test administration activities
- construction of all examinations for students in the preclinical curriculum
- score reports, individual student performance profiles, and item analyses for each examination

POLICIES FOR USE OF CUSTOMIZED ASSESSMENT SERVICES

Participating schools must agree to use the customized assessments in compliance with NBME policies and procedures. Moreover, participating institutions understand that NBME customized assessments are intended to complement other sources of information about the educational progress of students. The assessment results should be interpreted in light of other available information. Likewise, curriculum evaluation cannot be based on assessment results alone. The quality of teaching can and should be evaluated by frequent peer observation and student feedback, not inferred solely from the level of assessment results. Faculty and administration should not view the results of NBME customized assessments as the beginning and end of evaluation. The success of this program depends on vigorous adherence to NBME policies and protocols, including the implementation of the strictest security measures.

KEY MEDICAL SCHOOL STAFF ROLES

Listed below are the key players that support the construction and administration of the Customized Assessment Services.

The NBME Executive Chief Proctor (ECP)

The ECP, nominated by the dean, has access to NSP and has overall responsibility for ordering services online, coordinating test administration activities, and assuring that the security of the test items reviewed by faculty and seen by examinees is maintained. Specific ECP responsibilities are outlined on **page 4**.

The Test Construction Coordinator (TCC)

One faculty member should be identified to take the lead during the test construction process. This Test Construction Coordinator (TCC) should:

> **OTHER RESOURCES:** For more information about all NBME web-based examinations, please visit the NBME site (**www.nbme.org**) and go to Schools & Residency Programs.

- recruit faculty to participate in the process
- generate the test for faculty review
- coordinate the schedule for the sessions during which faculty will define specifications (blueprint) and refine the test
- serve as the primary local resource for use of the test construction software
- confirm the exam to be administered is the one finalized by faculty

The above tasks are suggested for the TCC role and may be delegated to others if appropriate. The NBME will automatically assign this role to the user who generates the test to start the **test refinement process**.

Faculty Participants

Any faculty member may participate in the test construction process; however, those who will review NBME test items must agree to terms and conditions for review of the material. These conditions will be presented in an electronic nondisclosure agreement before faculty can view any web pages that display item content. This nondisclosure agreement is shown on **page 10**.

The User Account Administrator (UAA)

The UAA creates accounts in NSP for anyone who needs to access the test construction system, score reports, or item analysis data.

ECP RESPONSIBILITIES

The ECP plays a key role in all local aspects of the Customized Assessment Services program, as outlined in the tasks below. The ECP and TCC should have a continu-ous dialogue throughout the process to achieve success-ful results. The ECP will order all services through the on-line ordering system on NSP.

The following are the primary responsibilities of the ECP:

Order a Subscription

The first step is for the ECP to order a subscription to the service and identify the estimated month and year that test construction activities will begin. The NBME must re-ceive payment for the subscription fee before the sub-scription period starts. The subscription becomes active on the first day of the month entered as the starting point for test construction activities.

As of July 1, 2014, the CAS fee structure has been up-dated to a 'pay as you go' model, as opposed to the previous model in which all administrations for the sub-scription year were paid up front. Once a subscription is ordered, invoices will be generated each time an exam administration is ordered.

Go to **CAS Fees** to review the most current fees for the service.

Use of Two-Factor Authentication

In order to access the CAS system, faculty must interact with our enhanced two-factor security. This will require an initial configuration of a cell phone, landline, or tablet. Two-factor authentication is also used to access the web-based item analysis, which is available for re-view when the score reports are posted to NSP. User in-structions are available on the "Build Exam" tab of NSP for reference.

Once an examination has been constructed and final-ized by faculty, the ECP will receive an e-mail notification that it can be requested for test administration through the online ordering system.

Test Delivery

Examinations are delivered using test administration and proctoring software developed by the NBME. Details about how to arrange for and administer an NBME web-based exam can be found on the **NBME website**.

Test Construction Activities

Those authorized to build a customized exam log in to NSP to access the test construction system using the enhanced security of two-factor authentication. The components of the test construction process include:

- creating test specifications (the blueprint)
- generating the exam and reviewing/refining it
- identifying the score categories to be reported
- regenerating the exam if necessary
- previewing the exam prior to test administration

The main menu of the customized assessment test construction system is displayed below.

Build the Test Blueprint

To create a new blueprint, the user enters the blueprint name and the desired number of items on the test. The user also has the option to build the exam by discipline or by organ system. One option offers use of pre-designed organ system blueprints to serve as "starters" to create tests for integrated courses. Descriptions of the pre-designed blueprints are available on **page 11**.

Participating faculty should review the USMLE Step 1 Content Outline available on the Build Exam tab of NSP in advance of the initial session. The test specification screens are formatted in the same hierarchical content structure used for Step 1.

During the session, faculty will filter the overall pool of items (by discipline or by organ system) to a target number that is relevant to the areas they want to test. This is done by:

- selecting the discipline(s) to be included
- reviewing the content areas, entering the minimum/maximum number of items desired for each area

- specifying the minimum/maximum number of images to be included
- indicating the types of items (patient vignettes, experimental vignettes) to be included

The blueprint can be approved once these activities have been completed.

Generate a Draft Exam
This can be done right after the blueprint is approved. This process has three parts:

- confirming the name for the examination
- providing e-mail addresses for every individual to be notified when the draft exam is ready
- initiating the exam generation process

Within one hour, NBME test construction software will generate a draft exam that consists of the target number of items provided in the test blueprint plus a 100% overage of items that can be used for substitution purposes.

Finalize the Exam
This stage is the most involved. It consists of two sections: refining the test and defining score categories. If the work for this stage cannot be finished in one session, it can be saved and accessed later.

Refine the Exam
In this section, faculty will:

- define the pacing for the exam for all examinees. Two standard pacing modes are available:
  - **regular:** 72 seconds per item, 50 items per hour, or
  - **extended:** 90 seconds per item, 40 items per hour
- review item content. A variety of attributes (such as item difficulty, score categories, and discipline/organ system tags) are provided. Faculty may also filter items in various ways. The system keeps track of items that are reviewed during the session and also allows faculty to annotate or flag them.
- use the overage pool to replace items from the draft that are a better match to course content.

Combine Your Exams The test refinement feature includes the capability to combine exams into a single test.

For example, if you are building a test that covers content across different courses, separate exams can be generated from different blueprints and then merged into one form.

Define Score Categories
The score reports include total test scaled and percent correct scores as defaults. There is also the option to define additional score categories by using system-generated categories or by creating custom categories. A category must have *at least 25 items* for reporting scores. Individual examinee feedback is generated in the form of graphic performance profiles, which highlight content areas of strength and weakness. A content category must have *at least 10 items* to appear on the profile.

After this process is complete, the exam can be locked and made available for ordering by the ECP.

Regenerate the Exam (Optional)
This feature provides the capability to make changes to an exam without losing the revisions that have already been made. It should be used when the items available for selection are unsatisfactory. The process entails dropping items, then regenerating the test to replace them with items from the same content area, if they are available.

Preview the Exam
The exam should be previewed in advance of the test administration as a final content check and to assure that the pacing is accurate. The test preview format is in the same format that an examinee sees during test delivery.

Build a Test Battery
The NBME test battery feature enables institutions to administer a series of exams using a combination of fixed-form standardized subject tests and customized examinations. This option may be useful, for example, in supplementing the content covered by a subject test with a small customized module (<25 items) that is more reflective of local assessment needs.

Test batteries are intended for administration to a single group of examinees and are handled as one test session. The examinations can be arranged in a defined test delivery sequence with the capability to set "off the clock" breaks in between exams. Once a customized exam is locked, it is available for use in a test battery.

Score Reporting

NBME customized examinations provide institutions with a tool for measuring an examinee's understanding of a series of content areas defined by each institution. Course objectives vary and customized examinations are targeted to meet the specifics of a given curriculum. The NBME neither sets nor recommends a "passing" score. Generally, customized examinations should be used in conjunction with other indicators of student performance to determine grades.

Scaled scores for the total test and content areas defined during the test construction phase are reported and computed to have a mean of 70 and a standard deviation of 8 for the scaling group* of examinees testing at your school. The scale provides a useful tool for comparing the scores of students to one another as well as identifying the relative strengths and weaknesses within the content areas defined for the exam. Feedback is based only on the performance of students from your school who took the customized examination. (Norm data based on a large, representative group of examinees are not available for customized exams, since the exams are targeted specifically to an individual curriculum.) Results should not be generalized to other cohorts, either within the institution or across other institutions.

> * A scaling group is automatically created if there are at least 25 examinees. Scaling groups can also be defined through the **Scaling Group Application**. A scaling group excludes examinees who did not take the exam under standard pacing conditions, scored >3 standard deviations below the mean, or omitted more than 10% of the items. Additionally, any examinee that is removed from the examinee group by the Chief Proctor will be removed from the scaling group.

There are two stages of score reporting:

Immediately after the Close of a Test Session – Total Test Percent Correct Scores

After the test session is ended by the test manager through the proctoring system, a roster confirmation will appear listing the examinees along with their total test percent correct scores. The percent correct mean, standard deviation, and minimum/maximum scores are also shown. The test manager may exclude any records that may be invalid (e.g., if an examinee left due to illness) and then confirm the group to be submitted for final scoring.

The individual total test percent correct scores and summary statistics may be printed, downloaded to Excel, or saved as an Adobe® Reader (PDF) file.

Within One Week of the Test Administration Date – Full Score Report

The NBME will send an e-mail notification to the ECP when the score report has been posted to NSP. Score reports are usually available within one week of the test administration date.

Three types of performance feedback are provided in the final score report:

- **Score report** – a single PDF file with the following components:
  - Score Interpretation Guide
  - Total Test and Content Area Score Descriptive Statistics
  - Total Test Scaled and Percent Correct Score Frequency Distribution
  - Roster of Scaled and Percent Correct Scores
- **Downloadable comma-separated values (CSV) files of all reported scores**
- **Individual Performance Profiles (if there were at least 25 examinees)**
- **Item analysis in two formats:**
  - Web-based: provides the capability to review full item text along with related item statistics online
  - Downloadable report: (PDF and CSV files) – includes a short keyword description of the item and individual item statistics to compare with the performance of a USMLE Step 1 reference group.

Components of the score report and information about the performance profiles and item analysis are described in the table on **page 13**.

Samples of the score report and performance feedback are provided on **pages 18-22**.

Scaling Group Application

Because scores for customized exams are locally scaled for each test administration, examinee scores cannot be directly compared across multiple administrations of the same examination. A Scaling Group Application (SGA) is available to link these administrations for comparison purposes. The SGA, accessed through NSP, is available to institutions with active subscriptions to Customized Assessment Services.

System Requirements for Test Construction

NBME Customized Assessment Services System Requirements for Test Construction Activities	
Hardware	PC with Intel P3 1GHz or higher processor
Operating System	Microsoft Windows, 7Mac OSX
Browser	Windows: IE 6, IE 7, IE 8, Firefox 3 Mac OSX: Safari 3, Firefox 3 **Note:** Higher versions of these browsers or the use of Google Chrome may compromise functionality of some features of the test construction system.
Browser Settings	• JavaScript enabled • Cookies enabled • CSS enabled
Internet Connection	Broadband (DSL, Cable, or T1)
Pop-up Blocker	Disabled
Monitor	17" or larger color monitor
Resolution	1024 x 768
Color Setting	24-bit or better
Cookies	Enabled

Nondisclosure Agreement

The screen text shown below will appear before faculty can view item content for either the exam construction phase or during review of the item analysis. By clicking on the "Agree" button that appears at the bottom of the screen, the user certifies eligibility to review the items.

Examinations of the National Board of Medical Examiners® (NBME®) are provided to institutions with a legitimate interest in the education of physicians. Under secure conditions, the NBME makes the examinations and/or items available for review to interested faculty ("Reviewers") who wish to evaluate the appropriateness of the items for tests for courses and clerkships or to help them interpret their students' scores. All Reviewers must certify that they recognize the importance of maintaining the complete security of this material, meet the following conditions, and agree to the following terms. The terms and conditions set forth below govern your access to NBME Examination materials.

TERMS AND CONDITIONS FOR REVIEW OF NBME EXAMINATION MATERIALS FOR CUSTOMIZED ASSESSMENT SERVICES

1. Each Reviewer acknowledges and agrees to preserve the confidentiality of the test material. Reviewers are not permitted to disclose all or any part of the contents of the test material to anyone. In addition, no reproduction of the test material may be made by any means, including, but not limited to, printing, photocopying, reconstruction through memorization, and/or dictation.
2. A Reviewer is not eligible to take part in a review if that person currently participates or plans to participate within the next three (3) years as a faculty member or preparer in any review course for which students are charged a fee. This condition applies if any student, from the Reviewer's own school or from any other school, is charged a preparation course tuition fee, other than a small charge for production of review course materials.
3. A Reviewer is not eligible to take part in a review if that person is listed as an editor, author, or coauthor of any publication which is advertised, regarded, or perceived as being sold for preparation for medical licensing examinations. Publications include materials using any medium, including, but not limited to, print, electronic media, and videotape.
4. A Reviewer who is presently enrolled in an MD or DO degree program or who is a candidate for medical licensure in the United States of America is not permitted to take part in this review.
5. A Reviewer who, within three (3) years of the date of the execution of this Non-Disclosure Agreement, becomes enrolled in an MD or DO degree program or becomes a candidate for medical licensure in the United States of America must promptly notify NBME of his or her participation in the review of NBME Examination materials.
6. Each Reviewer certifies that he or she is a faculty member in good standing at the medical school and has been authorized by the school to review the applicable test materials and that he or she has read, understands, and agrees to comply with the conditions for review of test material stated herein, including the nondisclosure provisions.
7. Each Reviewer acknowledges that the Examination Materials are owned and copyrighted by the National Board of Medical Examiners and that any unauthorized reproduction of NBME materials constitutes copyright infringement and is strictly prohibited. NBME intends to take infringers to court and reserves the right to pursue all remedies available.

By clicking on the "Agree" button below, I certify that I am authorized to review NBME test materials during this online session and **that all present understand and agree** to comply with the terms and conditions set forth above. If you do not agree to be bound by these provisions, do not proceed.

"Starter" Blueprints for Constructing Organ System Exams

Year 1: Use for curricula where there are two passes through the organ system, emphasizing normal structure and function. Use this blueprint for the *first* pass.

Year 2: Use for curricula where there are two passes through the organ system, with the second pass emphasizing abnormal structure and function. Use this blueprint for the *second* pass.

Integrated: Use for curricula where there is *one* pass through the organ system in which normal structure and function, abnormal processes, pharmacology and behavioral considerations are covered in the same block.

Blueprint Description	Target # Items	Blueprint Description (con't)	Target # Items
Cardiovascular System Year 1	50	Immune System Year 2	50
Cardiovascular System Year 2	50	Immune System All	75
Cardiovascular System Integrated	75	Musculoskeletal System Year 1	50
Central and Peripheral Nervous System Year 1	50	Musculoskeletal System Year 2	50
Central and Peripheral Nervous System Year 2	50	Musculoskeletal System All	75
Central and Peripheral Nervous System All	75	Neuroscience (includes Neuropathology and Psychopathology/Psychopharmacology)	175
Endocrine System Year 1	50	Physiology (includes Neurophysiology)	150
Endocrine System Year 2	50	Renal/Urinary System Year 1	50
Endocrine System All	75	Renal/Urinary System Year 2	50
Gastrointestinal System Year 1	50	Renal/Urinary System All	75
Gastrointestinal System Year 2	50	Reproductive System Year 1	50
Gastrointestinal System All	75	Reproductive System Year 2	50
Hematopoietic and Lymphoreticular Systems Year 1	50	Reproductive System Integrated	75
Hematopoietic and Lymphoreticular Systems Year 2	75	Respiratory System Year 1	50
Hematopoietic and Lymphoreticular Systems All	75	Respiratory System Year 2	50
Histology (includes Cell Biology)	125	Respiratory System Integrated	75
Human Genetics	50	Skin and Related Connective Tissue Integrated	75
Immune System Year 1	50		

Sample Score Report Components

Total Test Percent Correct Scores and Summary Statistics Generated at the Close of a Testing Session

NBME
National Board of Medical Examiners

Disease and Diagnosis
School ID and Name

Test Window: XX/XX/XXXX to XX/XX/XXXX

Examinees Tested: 143   Examinees Not Tested: 0   Examinees Scored: 143

Preliminary % Correct Mean: 69% Std Deviation: 8 Min: 47% Max: 87%

Note: The summary statistics and percent correct scores shown are preliminary and may be subject to change on the final score report as a result of quality control analysis.

Examinee Name	Examinee ID	# of Items Completed	Preliminary % Correct
Test, One	3043	144 out of 144	70%
Test, Two	3009	144 out of 144	74%
Test, Three	3022	144 out of 144	76%
Test, Four	3006	144 out of 144	62%
Test, Five	3053	144 out of 144	76%
Test, Six	3030	144 out of 144	69%
Test, Seven	3061	144 out of 144	50%

Final Score Components

Component	Description	Min. # Examinees Required to Receive
Percent Correct Score Report PDF format CSV format (score roster only)	For total test and content areas with at least 25 items: • Score Interpretation Guide • Roster of percent correct scores • Descriptive statistics: mean, standard deviation (SD), low score, high score • SEM and reliability with a minimum of 10 examinees Frequency distribution (total test only) **NOTE:** Descriptive statistics are reported for the total group. If a scaling group was defined for the test administration, then reliability and SEMs are reported based on the scaling group.	1 (based on total group*) *Includes all examinees.
Scaled Score Report PDF format CSV format (score roster only)	For total test and content areas with at least 25 items: • Score Interpretation Guide • Roster of scaled scores (mean 70, sd 8) • Descriptive statistics: mean, SD, low score, high score Frequency distribution (total test only) NOTE: Descriptive statistics are reported for the total and scaling groups. Reliability and SEMs are based on the scaling group defined for the test administration.	25 (based on scaling group taking the exam on the same test date and/or a previous administration of the same exam.*) *A scaling group excludes examinees who did not take the exam under standard pacing conditions, scored >3 standard deviations below the mean, or omitted more than 10% of the items. Additionally, any examinee that is removed from the examinee group by the Chief Proctor will be removed from the scaling group.

Final Score Components

Component	Description	Min. # Examinees Required to Receive
Item Analysis (IA) PDF and CSV formats	*Summary Info:* # items, mean item difficulty, discrimination indices by content category – compared to Step 1 reference group *Item-by-item data:* organ system identifier, Step 1 content outline category, short keyword description, vignette type, difficulty (P-value) for your examinees, Step 1 reference group, difference between item difficulties, and discrimination index **NOTE:** If a scaling group is defined for a test administration, then the IA statistics are based on the scaling group. If a scaling group is not defined, then the IA is based on the total group.	10 for item difficulty 40 for item discrimination* *Standard exemptions for scaling groups will apply.
Web-based Item Analysis Available online only. Accessible for 90 days after the score reports are posted to the *NBME* **Individual Examinee Performance Profiles** PDF format – bundled in WinZip file	For each item: • complete item text • % of examinees answering each option • difficulty (P-value) • discrimination • national P-value based on previous use of the item in USMLE and/or subject exam Graphical profile for individual examinee that shows: • total test scaled and percent correct scores • horizontal bands representing areas of strength and weakness in content areas	10 for item difficulty 40 for item discrimination (based on scaling group) 25 (based on scaling group)

Scaled Score Descriptive Statistics

National Board of Medical Examiners
Customized Examination
Scaled Score Descriptive Statistics

Page 3 of 10

School: 000000 - Generic Medical School
Examination Name: Examination Name
Scaling Group: B00000 (MM/DD/YYYY)

Test Dates: MM/DD/YYYY to MM/DD/YYYY
Order ID: B00000

Scaled score descriptive statistics, reliability and the standard error of measurement (SEM) values appear below for the total test and all content areas defined by your school.

The reliability coefficient estimate refers to a score's expected consistency. An examination score is reliable to the extent that administration of a different random sample of items from the same content domain would result in little or no change in an examinee's rank order in the group. Reliability is affected by the homogeneity of the items and of the examinees, as well as by the length of the examination. Measurement error is present in all test scores, and the SEM provides an index of the imprecision of scores. Like the standard error for a laboratory study, the SEM is expressed on the same scale as the test scores and can be used to construct confidence intervals around the scores. For example, if an examinee receiving a scaled score of 60 was tested repeatedly with similar exams, 95% of the scores received should fall between 52 and 68 (60 plus/minus two times an SEM of 4).

The reliabilities and SEMs are computed based on the scaling group of examinees. Examinees who did not take the exam under standard timing conditions, scored more than 3 standard deviations below the mean, omitted more than 10% of the items, and other examinees who were removed at the request of your school were excluded from these calculations.

ID	Content Area	N Items	Reliability	SEM	Total Group (N=100)				Scaling Group (N=99)			
					Mean	SD	Low	High	Mean	SD	Low	High
	Total	100	0.72	3	70	8	41	85	70	8	48	84
1	Content Area 1	40	0.41	5	70	8	40	86	70	8	43	82
2	Content Area 2	25	0.4	6	70	8	43	84	70	8	51	86
*	Content Area 3	17	0.32	4	70	8	32	85	70	8	34	82
*	Content Area 4	12	0.4	6	70	8	51	82	70	8	53	83

* This content area does not appear on the roster but may appear on the individual performance profile.

MM/DD/YYYY

Scaled Score Roster

National Board of Medical Examiners
Customized Examination
Roster of Scaled Scores

School: 000000 - Generic Medical School
Examination Name: Examination Name
Scaling Group: B00000 (MM/DD/YYYY)

Test Dates: MM/DD/YYYY to MM/DD/YYYY
Order ID: B00000

ID	Name	Total Test Scaled Score	Content Area ID* (Scaled Scores) 1	2
00001	EXAMINEE A	65	81	83
00002	EXAMINEE B	72	73	71
00003	EXAMINEE C	53	57	53
00004	EXAMINEE D	65	62	72
00005	EXAMINEE E	85	81	78

Total Test Scaled Score Frequency Distribution

National Board of Medical Examiners
Customized Examination
Total Test Scaled Score Frequency Distribution (Based on All Examinees Tested)

School: 000000 - Generic Medical School
Examination Name: Examination Name
Scaling Group: B00000 (MM/DD/YYYY)

Test Dates: MM/DD/YYYY to MM/DD/YYYY
Order ID: B00000

Total Test Scaled Score	Frequency Count	%	Cumulative Frequency Count	%	Total Test Scaled Score	Frequency Count	%	Cumulative Frequency Count	%
41	1	1	1	1	67	10	6	75	47
45	1	1	2	1	69	10	6	85	53
46	1	1	3	2	70	5	3	90	56
50	1	1	4	2	71	13	8	103	64
53	1	1	5	3	72	6	4	109	68
55	1	1	6	4	74	7	4	116	72
56	4	2	10	6	75	10	6	126	78
57	6	4	16	10	76	6	4	132	82
59	6	4	22	14	78	5	3	137	85
60	2	1	24	15	79	7	4	144	89
61	7	4	31	19	80	5	3	149	93
62	6	4	37	23	81	5	3	154	96
63	10	6	47	29	83	3	2	157	98
65	4	2	51	32	84	3	2	160	99
66	14	9	65	40	87	1	1	161	100

Percent Correct Score Descriptive Statistics

National Board of Medical Examiners
Customized Examination
Percent Correct Score Descriptive Statistics

School: 000000 - Generic Medical School

Examination Name: Examination Name

Test Dates: MM/DD/YYYY to MM/DD/YYYY

Order ID: B00000

Percent correct score descriptive statistics appear below for the total test and all content areas defined by your school. If a scaling group was identified for this order or an item analysis was produced, reliability and the standard error of measurement (SEM) values will also appear. If your order has only one examinee, descriptive statistics will only appear for the total test and content areas with at least 25 items.

The reliability coefficient estimate refers to a score's expected consistency. An examination score is reliable to the extent that administration of a different random sample of items from the same content domain would result in little or no change in a candidate's rank order in the group. Reliability is affected by the homogeneity of the items and of the candidates, as well as by the length of the examination. Measurement error is present in all test scores, and the SEM provides an index of the imprecision of scores. Like the standard error for a laboratory study, the SEM is expressed on the same scale as the test scores and can be used to construct confidence intervals around the scores. For example, if an examinee receiving a percent correct score of 60 was tested repeatedly with similar exams, 95% of the scores received should fall between 52 and 68 (60 plus/minus two times an SEM of 4).

If a scaling group was identified for this order, the reliability and SEM below are computed based on the scaling group of examinees excluding examinees who did not take the exam under standard timing conditions, examinees scoring more than 3 standard deviations below the mean, examinees who omitted more than 10% of the items, and other examinees that were removed at the request of your school. If a scaling group was not identified for this examination and an item analysis was produced, then the reliability and SEM are computed based on the examinees used for the item analysis. If a scaling group was not identified, had too few examinees, or there were not enough examinees for an item analysis, the reliability and SEM were not computed.

| ID | Content Area | N Items | SEM | Reliability | Total Group (N=100) | | |
					Mean	SD	Low	High
	Total Test	100	5	0.7	81.2	7.3	57	96
1	Content Area 1	40	7	0.43	86.1	8.4	56	100
2	Content Area 2	25	7	0.41	81.7	8.9	50	100
*	Content Area 3	17	9	0.37	83.8	11.9	25	100
*	Content Area 4	12	8	0.4	78.1	12.6	47	100

* This content area does not appear on the roster but may appear on the individual performance profile.

MM/DD/YYYY

Percent Correct Score Roster

National Board of Medical Examiners
Customized Examination
Roster of Percent Correct Scores

School: 000000 - Generic Medical School

Examination Name: Examination Name

Test Dates: MM/DD/YYYY to MM/DD/YYYY

Order ID: B00000

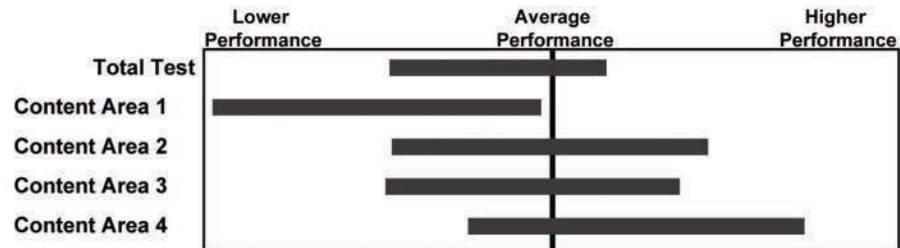| ID | Name | Total Test % Correct Score | Content Area ID* (% Correct Scores) | |
			1	2
00001	EXAMINEE A	76	96	96
00002	EXAMINEE B	84	88	82
00003	EXAMINEE C	66	72	64
00004	EXAMINEE D	78	76	86
00005	EXAMINEE E	93	96	89

Examinee Performance Profile – Page 1

NBME®

National Board of Medical Examiners
Customized Examination
Performance Profile

School: 000000 - Generic Medical School
Examination Name: Examination Name

Test Date: MM/DD/YYYY

ID: 00001
Name: EXAMINEE A   **Total Test Scaled Score:**   65
Scaling Group: MM/DD/YYYY   **Total Test Percent Correct Score:**   76

NBME customized examinations provide medical schools with a tool for measuring students' understanding of a series of content areas defined by each school. Your total test scaled and percent correct scores on this customized examination appear above. Total test scaled scores are scaled to have a mean of 70 and a standard deviation of 8 for the scaling group of examinees selected by your school. The total test percent correct score represents the percentage of items answered correctly on the examination and has been rounded to the nearest whole number.

The performance profile on the following page is provided to aid in self-assessment. It summarizes your performance onthe total test and in the content areas defined by your school. The vertical line in the center of the profile indicates the average level of performance in each content area for the scaling group of examinees selected by your school that completed this customized assessment. Areas of relative strength and weakness are indicated by the positioning of the performance bands. The width of each band reflects the precision of measurement for the associated content area;narrower bands indicate greater precision. An asterisk at either end of the performance band indicates that your performance band extends beyond the displayed portion of the scale. Small differences in relative position of bands should not be overinterpreted. If two bands overlap, performance in the associated content areas should be interpreted as similar. Please note that items may contribute to the calculation of scores in more than one content area. As a consequence, caution should be used when interpreting differences in performance across content areas.

The total test scaled score and this performance profile should not be compared to those from other NBME examinations. The scaling group that your school selected for this administration can be compared to other administrations if this same exam that used the same scaling group.

Examinee Performance Profile – Page 2

NBME®

National Board of Medical Examiners
Customized Examination
Performance Profile

School: 00000 - Generic Medical School
Examination Name: Examination Name

Test Date: MM/DD/YYYY

ID: 00001
Name: EXAMINEE A   **Total Test Scaled Score:**   65
Scaling Group: MM/DD/YYYY   **Total Test Percent Correct Score:**   76

	Lower Performance	Average Performance	Higher Performance
Total Test			
Content Area 1			
Content Area 2			
Content Area 3			
Content Area 4			

Web-based Item Analysis

Detailed Item Analysis Summary Page

NBME®

National Board of Medical Examiners
Customized Examination
Item Analysis Summary Report

School: 000000 - Generic Medical School   **Test Dates:**   MM/DD/YYYY to MM/DD/YYYY

Examination Name: Examination Name   **N Examinees:** 100

This report summarizes information provided in the online item analysis (IA) for this customized exam administered to your examinees during the test dates shown above. Mean item difficulty (p-value or "diff" in the online IA) and discrimination (r-value or "disc" in the online IA) for the content areas defined/selected by your school to be reported as subscores on this customized exam are summarized in the table below.

Due to the instability of test statistics for categories with a small number of items, mean item difficulties and discriminations are not provided for categories with fewer than 10 items. Mean item difficulties and discriminations for your school are based on performance of the current group of examinees testing during the dates specified above. Mean item difficulties are also provided based on performance of a Step 1 reference group (first-time takers from LCME-accredited schools). Since statistics based on Step 1 performance were computed when items appeared in examinations administered in different contexts, caution is advised when interpreting these item statistics. Your school results should not be generalized to other cohorts, either within your school or across other schools.

Score Category	Number of Items	Mean Item Difficulty		Mean Item Discrimination
		School	Step 1	School
Total Test	**100**	**.83**	**.78**	**.23**
Content Area 1	40	.78	.77	.33
Content Area 2	25	.84	.76	.24
Content Area 3	17	.85	.86	.27
Content Area 4	12	.88	.83	.23

Detailed Item Analysis Item Statistics Page

NBME®

**National Board of Medical Examiners
Customized Examination
Item Analysis Summary Report**

School: 000000 - Generic Medical School
Examination Name: Examination Name

Test Dates: MM/DD/YYYY to MM/DD/YYYY
N Examinees: 100

This report is provided to help you interpret the performance of your examinees testing during the dates shown above on each multiple-choice item included in this customized exam. The first table provides information for all items (total test) included in this customized exam. In the subsequent tables, information is provided for each item included within each content area defined/selected by your school to appear as subscores. For each item listed in the tables below, the Organ System, Step 1 Outline category, a Description, and Vignette type are presented. The item difficulty (p-value or "diff" in the online IA) and discrimination (item-total biserial correlation or "disc" in the online IA) indices based on the group of examinees from your school testing during the dates specified above are also provided. Examinees not taking the exam under standard timing conditions, examinees scoring more than 3 standard deviations below the mean, examinees who omitted more than 10% of the items, and other examinees specified by your school are excluded from this item analysis. Item difficulty based on performance of a Step 1 reference group is provided for gross comparative purposes only. The difference between the local school and the Step 1 item difficulties has also been computed. Since the Step 1 values were computed when items appeared in examinations administered in different contexts, caution is advised when comparing historical item statistics and interpreting differences. Your school results should not be generalized to other cohorts, either within your school or across other schools. To aid in interpretation of the data, summary information listing the total number of items, the mean item difficulties, difference, and discrimination is also provided at the end of each table.

Item Analysis for Total Test Sorted by Item Difficulty (Ascending Order)

Item #	Organ System	Step 1 Outline	Description	Vignette	Item Difficulty (p-value)			Item Discrimination (Biserial Correlation)
					School	Step 1	Difference	School
79	general principles	01.08.03.02 viral replication	hepatitis B	Experimental vignette	.31	.40	-.09	.20
39	renal/urinary system	09.02.01.02.01 glomerular disorders	lupus erythematosus, systemic	Patient vignette	.31	.57	-.26	.51
75	renal/urinary system	09.02.04.01 renal failure	necrosis, tubular, acute	Patient vignette	.32	.44	-.12	.13
10	general principles	01.07.04.01 antibacterials	Adverse effect, gentamicin	Patient vignette	.37	.46	-.09	.47

NBME®   Page 2 of 17   MM/DD/YYYY