USMLE INTERIM AND ANNUAL SCHOOL REPORT ENHANCEMENTS COMMONLY ASKED QUESTIONS

The United States Medical Licensing Examination® (USMLE®) program has redesigned and enhanced the Interim and Annual School Reports it produces for medical schools in the United States and Canada. Answers to common questions asked about these recent enhancements are available below.

What resources are available to help familiarize myself with the enhanced school reports and interpreting student performance data?

A number of resources have been created to help guide medical school faculty through the recent report enhancements. These include a sample report, two videos, and an interactive, annotated PDF.

Why did USMLE decide to redesign the Interim and Annual Score Reports distributed to medical schools?

Although the primary purpose of the USMLE exam series is to provide a standardized, independent assessment for state medical boards to determine medical license eligibility, the program also recognizes these reports, and the data within, can be a valuable component in assessing student learning and a program's curriculum.

The USMLE Interim and Annual Score Reports are created to help medical school faculty and student affairs deans make clear, appropriate, and valuable interpretations of their school's performance data compared to a national group of examinees from LCME-accredited schools.

In response to faculty feedback, a project team conducted multiple focus groups comprised of faculty from institutions across the United States. During the focus groups, it was determined that some data on the reports, particularly the content area profile bands, were prone to overinterpretation given USMLE examinations are designed to help make decisions for licensure eligibility rather than provide precise feedback on individual content areas.

Additionally, common themes also emerged from faculty feedback, namely, a desire to compare student performance across time and a clearer way to identify meaningful inferences related to content categories.

Who provided input on the new USMLE Interim and Annual Score Report design?

Focus groups were conducted before and after the redesign with various faculty representing medical schools across the United States. It was important to the USMLE program to engage directly with their stakeholders, especially reports users, so the enhancements would focus on better supporting their needs.

During the process, cognitive interviews were also facilitated, and faculty were asked to share what inferences they believed the sample reports supported. After several rounds of design, follow-up focus groups and cognitive interviewees reacted positively to the new design. The feedback received was extremely valuable and indicated that the new design supported inferences more clearly.

As medical schools begin to use these new reports, the USMLE program welcomes additional feedback.

What are the requirements for receiving each page of the USMLE Interim and Annual reports?

The following guidelines are used to determine what data and visuals are presented for each report.

Page 1

The first page of the Step 1 and Step 2 Clinical Knowledge (CK) Interim and Annual reports are produced if a school has at least one first-time examinee in the most recent cohort included in the report. However, some summary data may be excluded if a school has exactly one first-time examinee.

The Step 3 report is produced if any of the graduating years contains one or more first-time examinees.

Subsequent Summary Pages

For the Step 1 Interim and Annual report as well as the Step 2 CK Annual report, the violin plots, histogram, and content area feedback plots are included if a school had at least 20 first-time examinees in the current cohort included in the report. This prerequisite was established as the graphical summaries require a meaningful sample size to support reliable and valid inferences. Please note that any prior historical cohorts with fewer than 20 first-time examinees will be individually excluded from the results.

If the current cohort does not meet the requirement, the violin plots, histogram, and content area feedback plots will be excluded for the current and historical cohorts.

When will the next round of USMLE Interim and Annual Score Reports be released?

The USMLE program's revised school-level Interim and Annual Score Reports will be released following the schedule below.

Report	Anticipated Release Date
Step 1 2022 Interim Report (newly enhanced)	October/November 2022
Step 1 2022 Annual Report	February/March 2023
Step 2 CK 2023 Interim Report	March 2023
Step 3 2023 Annual Report	June 2023
Step 2 CK 2023 Annual Report	August/September 2023
Step 1 2023 Interim Report	October/November 2023

Do medical schools outside of the U.S. and Canada receive these reports?

All USMLE score reports for international graduates and medical schools are distributed by the Educational Commission for Foreign Medical Graduates (ECFMG), a member of Intealth. If interested, international schools should contact ECFMG for more information about what reports may be available.

What is changing on the USMLE Interim and Annual Score Reports?

The updated report designs now facilitate comparisons across time as well as more nuanced interpretation of overall performance.

Longitudinal data is featured throughout the reports to enable comparisons within a timeframe where test content is likely to be similar. Additionally, student performance is displayed visually using violin plots to illustrate the distribution of students scoring in a specific range. A histogram depicting the proportion of a school's most recent cohort of first-time test takers performing at various ranges relative to the national group is also provided.

The most considerable change has been made to the content area feedback plots. These plots contain three years of data and indicate the percent of a school's students scoring

statistically below, similarly to and above the national average for a particular content category.

What is the difference between a violin plot and a histogram?

A histogram shows the exact proportion of students scoring within a specified range of performance but does not convey where students performed within that score range. For example, a histogram may indicate 20% of students performed between 240 and 255. However, the histogram will not convey where most students were closer to 240, 255, or evenly spread out, for example.

A violin plot provides more granular information on the proportion of students near a specific score by showing the distribution of performance across the score scale. This additional detail, in combination with the markers for minimum passing standard and median, allows the violin plot to provide inferences beyond what can be inferred from the histogram.

Still, it's important to remember that the violin plot shows an approximation or representation of where students performed rather than indicating the exact proportion at each score point. This feature helps limit potential overinterpretations that might occur because of measurement error or cohort fluctuations.

How are examinees classified in the "lower, same, higher" content categories on the content area performance summary?

The USMLE program uses a peer-reviewed, published approach to classify examinees in the three performance levels (See Full Article Here). For each content area, this statistical approach compares whether each student scored meaningfully below or above the national content area average. Using the national content area average as the reference for comparison was suggested and supported by our focus groups. If the student did not perform meaningfully lower or higher than the average, they are classified in the "same" category. Content areas will have different cut-points because they differ in national average, size, and difficulty. Larger content areas offer greater precision and will classify more examinees in the lower and higher categories. Although our method does not use a difference in standard deviations to make classifications, the approach should yield similar results as identifying examinees about 1.0 standard deviations from the mean in most cases. However, we do not recommend using standard deviations if trying to replicate the report classifications in other situations as our statistical approach better identifies unique performers particularly when working with content areas of different length and reliability.

As discussed in the article, the approach is designed to limit unintended misinterpretations from less-reliable content area scores. The approach functions like more common statistical tests, accounting for the typical variability in each content area from sources such as measurement error. This characteristic helps identify practical differences and limits the overinterpretation of minor deviations that do not reflect differences in the underlying content area.

What constitutes a meaningful degree of change in the content area performance summary?

It's important to remember that the size of each content area—that is, the number of items in each category—differs. The number of items in a category affects the score reliability and will influence the proportion of students classified as above or below the national average. Specifically, larger categories will show more students in the lower and higher performance levels because the content category has more items to distinguish performance.

The USMLE program also recognizes that school curriculums vary. As such, meaningful change is particular to each category and school. It is recommended that you review the three years of data within a content category to see if there are consistent shifts in performance. For example, a signal of positive curricular improvement in a content area could be indicated by the number of examinees below the national average decreasing across time (or the number of examinees above the national average increasing across time). As the larger content categories provide more reliable scores, stronger inferences can be made from these categories' results.

As a reminder, the USMLE examinations are designed to measure a person's overall performance, rather than provide highly reliable data on specific content areas or curriculum. As such, it is recommended that you evaluate whether other curricular data align with your USMLE content performance summary results. USMLE items are designed to be highly integrative rather than concentrated on specific content areas and can measure more than one content category.

Why are 3-digit scores not included on the USMLE Step 1 Interim and Annual Reports?

The transition of USMLE Step 1 score reporting from a numeric score and pass/fail outcome to pass/fail only is an important first step toward facilitating broader, system-wide changes to improve and promote a more holistic review in the transition from undergraduate to graduate medical education. As part of this policy change, scores will no

longer appear on any USMLE Step 1 score reports (e.g., examinee score reports, transcripts, reports to schools). The new Step 1 Annual Report allows users to infer meaningful conclusions from the Step 1 data without the inclusion of numeric scores.