FROM ESSAY TO ESQUIRE:

UNDERSTANDING ARIZONA BAR EXAM GRADING

UBE
- MBE graded by machine at NCBE
- MEE / MPT graded by Committee on Examinations / Arizona Graders

Essay Grading
- Grading workshop
- Calibration
- "Live" Rank-Order Grading

Statistics
- Equating of MBE Scores
- Scaling of Essay Scores
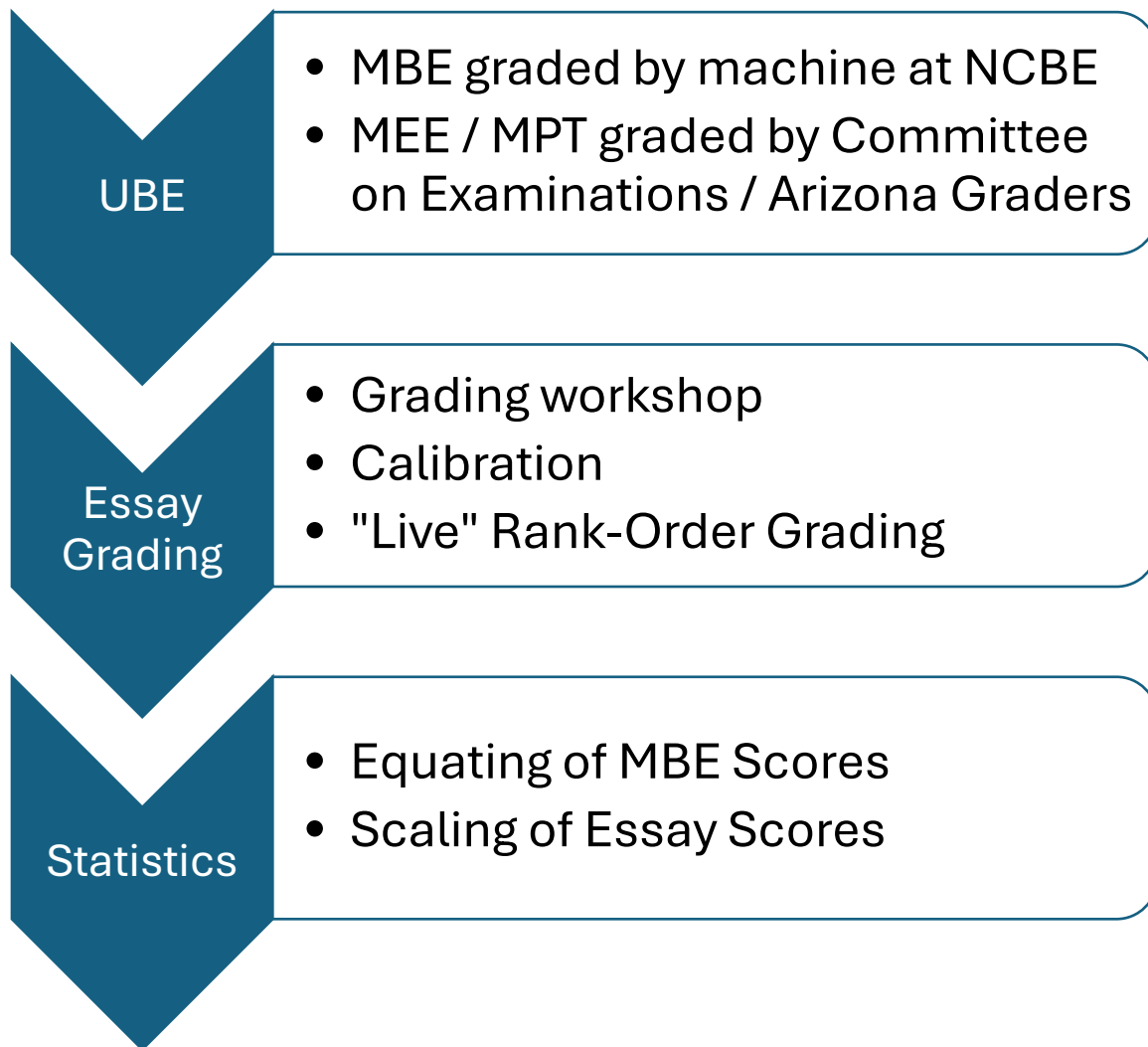
➤UBE Grading

- After examinees complete the UBE, the MBE portion is returned to NCBE for machine grading.
- The MEE and MPT portions are uploaded by the applicant, or staff if handwritten, at the end of the exam in preparation for grading.
    - Only the essay exams are graded by the Committee on Examinations and Arizona graders
    - NOTE – Arizona graders are an anonymous pool of experienced and licensed Arizona attorneys
- A grading team is created for each essay question, consisting of: (i) at least one Committee member, and (ii) at least two experienced graders.
    - Grading teams are assembled based on members' prior experience with grading and their history of working together.
- How do we grade essay exams?
    - Many examinees fundamentally misunderstand the grading process. Although this grading process is discussed more [here](#) and [here](#), overall, we use *rank-order grading*.
    - Rank-order grading is the process of assigning a score to an exam response relative to the other exam responses for the same question.
    - What does this mean – your exam is not graded in a vacuum or solely based on NCBE's rubric.
    - Exam responses are graded based on the performance of the entire cohort's response to a particular question.
        - This means that an exam that receives a score of 6 (the top of the suggested scale) is not a perfect response; rather, a score of 6 indicates that the exam was among the best for this pool of applicants.
        - Similarly, an exam that receives a score of 1 or 2 (at or near the bottom of the suggested scale) is not necessarily a poor response; rather, a score at the bottom of the scale indicates that the exam was among the poorer responses in the pool of applicants.
    - If you are unsuccessful in taking the UBE, it *can* be helpful to go back and review the question, model answer, and NCBE rubric; however, remember that each of your essay responses was graded *relative to* the entire pool of Arizona applicants.

➢Essay Grading:

Step 1 – Understand the Question / Grading Workshop

- Immediately after the UBE is administered, NCBE provides stakeholders (graders and Committee members) with access to copies of the exam questions, the model answer, and a rubric / explanation of the points for consideration while grading.
- NCBE also provides stakeholders with examinee responses, drawn from around the country, to be reviewed to orient the grading process.
    - NCBE also offers a "Grading Fundamentals" webinar for all stakeholders. Arizona Attorney Admissions requires the webinar be reviewed before each exam grading cycle.
- Immediately after each UBE administration, NCBE hosts a live grading workshop for each essay question. Each section of the workshop is hosted by the person who took the lead in drafting the MEE or MPT question.
    - Outside of the live session, NCBE records each grading workshop session so that graders and committee members can refer back to the discussion.

➢Essay Grading: Step 2 – Calibration

- Calibration is a process by which the graders and committee members ensure that they are aligned in how the grading process will proceed.
- Each grading team participates in the calibration process for their individual question.
- Calibration involves a random set of Arizona exams that each stakeholder grades in advance of a live meeting between the stakeholders to discuss the results.
    - Shortly after the NCBE hosts its grading workshop, the Attorney Admissions team uploads a randomly selected group of exams;
    - Each stakeholder reviews the grading materials and grades the exams;
    - The grading team meets to discuss the question, the grading rubric shared by NCBE, and how everyone scored the exams;
    - The committee member(s) have access to the graders' scores in advance of the meeting to identify where the scores diverge;
    - During the live meeting, the committee member leads the discussion of the results, including where the graders diverged on a score for a particular question; and

- o As a result of the conversation, all stakeholders reach agreement on both score and key points that graders will use to differentiate exam responses.
  - o NOTE – by the end of calibration, all exams have the same score and an early score curve should start to be noticeable.
- The scores assigned during the calibration process are *not* final – the calibration exams will be re-graded during the live grading process.

➤ Essay Grading: Step 3 – Live Grading

- Each grading team notifies Attorney Admissions after completing the calibration process → Attorney Admissions verifies that all calibration exam responses for that question have the same score and then provides the grading team with access to the exams for live grading.
- Shortly after the end of the calibration period, the Committee and Attorney Admissions hosts a "grading weekend."
  - o Why? A grading weekend enables open communication between graders and committee members and a compressed grading period.
  - o During grading weekend, the stakeholders solely focus on the grading process and complete the scoring of as many exam responses as possible.
  - o The compressed grading period is also helpful in maintaining calibration – ensuring that both graders are continuing to use the same criteria to assign scores.
- As mentioned above, Arizona's grading teams use rank-order grading to assign final scores to the exams. In short, this means that, while scores are largely based on NCBE's grading rubric, the final score is determined by looking at a particular exam's place in the overall cohort of responses for a particular question.
    - ▪ This means that an exam that receives a score of 6 (the top of the suggested scale) is not a perfect response; rather, a score of 6 indicates that the exam was among the best for this pool of applicants.
    - ▪ Similarly, an exam that receives a score of 1 or 2 (at or near the bottom of the suggested scale) is not necessarily a poor response; rather, a score at the bottom of the scale indicates that the exam was among the poorer responses in the pool of applicants.
    - ▪ Rank-order grading enables fair and consistent grading of the MEE and MPT from administration-to-administration, regardless of differences in exam difficulty.

- During live grading:
  - Graders work with groups of exams and inform the committee member leading the team when the group has been graded;
  - In addition to grading the exams, graders also keep track of notes for each exam, as future exam responses may warrant adjusting prior exams
    - grading is a dynamic process that involves constantly referring back to prior exams to determine where a particular exam fits within the entire cohort
  - In addition to grading the pool of exams, graders also score "embedded exams" – these are live exams that *both* graders score. This is done to ensure that both graders remain calibrated and are still aligned on how the exams are being scored.
    - This is a vital step in the grading process → embedded exams serve as a method of quality control to objectively assess the grading process.
    - In the event that the embedded exam scores do not align, the committee member pauses the grading process and investigates why there is a difference between the two graders.
  - The Curve:
    - Based on the natural distribution of individuals' talent and knowledge, most questions' scores naturally segregate into a bell curve. An example curve is included below.
    - In general, graders are suggested to keep the following scores aligned: (i) 1s and 6s; (ii) 2s and 5s; and (iii) 3s and 4s. In other words, NCBE suggests that the numbers of these scores should be similar.
      - NOTE – a score of zero is used only for a blank or completely non-responsive answer.
    - An important responsibility for the committee member is to assess the curve as graders finish with their exam sets.
      - After approximately the first 90 exams, the curve should start to resemble a bell curve, with the average score around a 3.5 (the peak of the bell curve between 3 and 4).
    - Based on the state of the curve, the committee member may suggest that the grader evaluates whether any exams should be moved up or down at any potential score level.
    - In addition, the committee member ensures that the graders are making adequate use of the entire grading scale.

- o Other tools used during the grading process for quality control:
  - ▪ Any exam that is scored as a 0 or 1 is automatically reviewed by the committee member – this ensures that there is additional scrutiny for these low scores
  - ▪ Graders are instructed to use + / - scoring
    - • As mentioned above, graders are instructed to keep notes on all exam scores – the use of a + / - score enables easier score adjustments as grading proceeds
  - ▪ In addition to the committee member reviewing the curve and the embedded exams, Attorney Admissions also reviews the curve and embedded exam scores to double check the results.

Total Score Distribution - MEE #
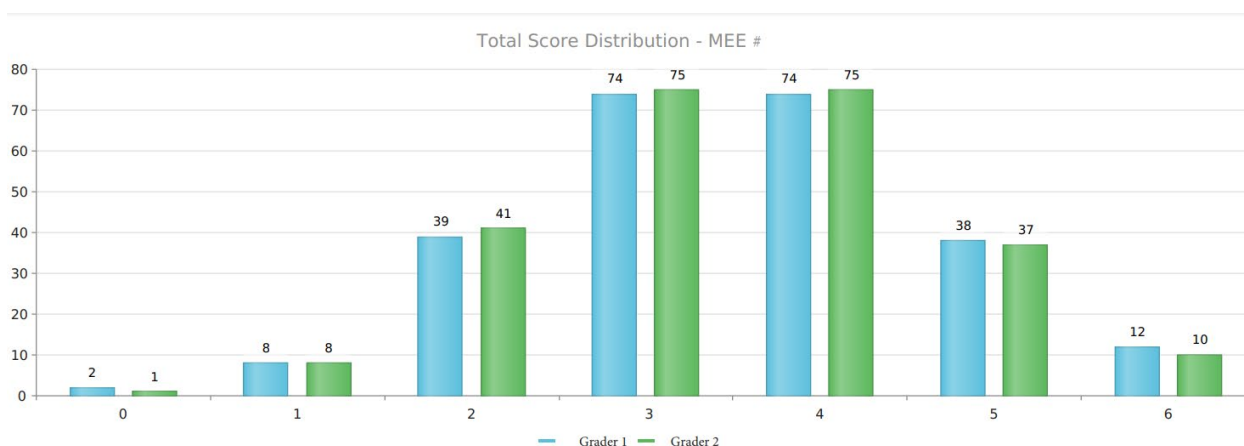
➤ Statistics

- • There are two distinct statistical processes that occur after scoring – here is a wealth of resources on these processes.
- • MBE / multiple choice:
  - o NCBE recognizes that not all MBE administrations are the same – some administrations are easier and some are harder
  - o NCBE uses equating to compensate for differences in MBE difficulty
  - o NCBE published a highly descriptive article that details the equating process.
  - o In short, the equating process relies on the following steps:
    - ▪ Each MBE administration includes embedded "equators" – prior MBE questions that represent a full MBE in terms of subject matter and statistical similarity

- Using the examinees' responses to the equators on both the current exam and prior exams, NCBE is able to adjust the MBE score to address the differences in exam difficulty and examinee performance.
- To ensure accuracy, NCBE uses at least two different psychometricians to complete the equating process
- NOTE – this is the difference between a *scaled* MBE score and a *raw* MBE score
- MEE / MPT:
  - *Cannot* use "equators" for the essay portion – the essay questions are too memorable and cannot be reused
  - NCBE uses "scaling" to address differences in average examinee proficiency → scaling not only adjusts for an upswing or downswing in proficiency, but also for any change in the difficulty of written test questions or any change in the "harshness" of graders
  - Here is a helpful NCBE article on the scaling process.
  - The scaling process uses the MBE scale to normalize the essay scores.
  - In brief, here is a summary of the scaling process:
    - First, the mean and standard deviation for ARIZONA's MBE scores is determined (see equating discussion above);
    - The mean and standard deviation for each essay response is determined; and
    - The essay response scores are converted so that they have the same mean and standard deviation as the MBE for the same group of examinees (the pool of Arizona applicants).
    - Again, it is important to realize that a raw score (e.g., a score of 4 on MEE #1) is NOT the same as a scaled score, which is the result of the scaling process.
  - For those who are interested in learning more about the scaling process, check out this article, which includes some simple data that you can use to practice understanding the scaling process.

➢A note on new graders and new Committee on Examinations Members

- The Committee is always recruiting qualified Committee members and graders. This means that Committee is always engaging with individuals who are new to the grading process. Here are the steps that the Committee follows to ensure that new stakeholders are oriented to the process:
  - New graders and new Committee members attend an orientation to become initially acquainted with the grading process;
    - New graders shadow experienced graders – this includes actively participating in NCBE grading workshops and calibration process;
    - New graders are also invited to observe the live grading process (but do not assign any final scores for essays);
  - New Committee members shadow experienced Committee members – often the Committee's Chair or Vice Chair;
    - This includes actively participating in NCBE grading workshops and the calibration process
    - New Committee members are also invited to closely observe the experienced Committee member's actions during the live grading process, including understanding score distributions and rank-order grading.
  - After completion of the grading cycle, the new grader is evaluated to determine whether they should be added to the grader pool and grade live exams during a future exam cycle.
  - Similarly, the new Committee member is also assessed to determine whether they should remain as a shadow for the next exam cycle or whether they should be able to oversee their own grader team for the next exam cycle.