# Endurance Course Scoring

## United States Naval Academy

Coleman Strickland, Capt USMC

---

## Purpose

The purpose of this analysis is to identify an appropriate and accurate scoring table for the United States Naval Academy (USNA) endurance course. This analysis aims to align USNA endurance course grading metrics with current scoring tables at the United States Marine Corps (USMC) Officer Candidate School (OCS). Outputs of this project provide gender specific scoring tables to properly evaluate students who complete the USNA endurance course.

This report is organized into four different sections:

1) Previous Work: OCS Scoring Tables/MARSOT Scoring Algorithms
2) Assumptions
3) Analysis
4) Score Comparison

---

## Previous Work

**OCS Scoring Tables**

The endurance course at OCS is substantially different than the USNA endurance course; therefore, a one-to-one comparison is neither accurate nor recommended. A listing of the main differences with OCS's endurance course are below:

- Completion of the obstacle course is required prior to beginning the endurance course.
- A load bearing vest to include a rifle is required to be worn/carried for the duration of the event.
- Multiple water obstacles must be navigated throughout the course to include complete submersion and high-water wading.

For reference, the OCS scoring table is illustrated in Figure 1. Scoring values are divided across genders and feature a maximum and minimum score for males and females. A one-point deduction occurs every 30 seconds for both genders until the minimum score is achieved.

| Male Time | Female Time | Points | Male Time | Female Time | Points |
|---|---|---|---|---|---|
| 33:00 or Less | 43:00 or Less | 100% | 38:01 – 38:30 | 48:01 – 48:30 | 89% |
| 33:01 – 33:30 | 43:01 – 43:30 | 99% | 38:31 – 39:00 | 48:31 – 49:00 | 88% |
| 33:31 – 34:00 | 43:31 – 44:00 | 98% | 39:01 – 39:30 | 49:01 – 49:30 | 87% |
| 34:01 – 34:30 | 44:01 – 44:30 | 97% | 39:31 – 40:00 | 49:31 – 50:00 | 86% |
| 34:31 – 35:00 | 44:31 – 45:00 | 96% | 40:01 – 40:30 | 50:01 – 50:30 | 85% |
| 35:01 – 35:30 | 45:01 – 45:30 | 95% | 40:31 – 41:00 | 50:31 – 51:00 | 84% |
| 35:31 – 36:00 | 45:31 – 46:00 | 94% | 41:01 – 41:30 | 51:01 – 51:30 | 83% |
| 36:01 – 36:30 | 46:01 – 46:30 | 93% | 41:31 – 42:00 | 51:31 – 52:00 | 82% |
| 36:31 – 37:00 | 46:31 – 47:00 | 92% | 42:01 – 42:30 | 52:01 – 52:30 | 81% |
| 37:01 – 37:30 | 47:01 – 47:30 | 91% | 42:31 – 43:00 | 52:31 – 53:00 | 80% |
| 37:31 – 38:00 | 47:31 – 48:00 | 90% | 43:01 or more | 53:01 or more | Failure |

Figure 1: OCS Scoring Table

A critical feature highlighted with the OCS scoring table is that the distribution of times for males and females is uniform: males have a time window of 10 minutes (33 minutes to 43 minutes) to complete the event, and females have the same time window shifted by 10 minutes (43 minutes to 53 minutes). Allowing these completion windows to cover the same time horizon assumes that the male and female times have similar distributions. More specifically, according to the OCS scoring table, OCS assumes the distributions of male and female completion times have similar variances – we explore this premise further under the section "Analysis".

**MARSOT Scoring Algorithm**

The scoring algorithm implemented to score and rank midshipmen who participate in the MARSOT Screener is considerably different from using a standard scoring table for two main reasons:

1) The two screeners usually occur under considerably different seasons – fall and spring (mid-February). Using a standard scoring table across both screeners will have a significant effect on the endurance course scores as the obstacles are more difficult to maneuver in adverse weather conditions. The MARSOT screener encounters the same issue with the obstacle course. Calculating scores relative to the individual screeners eliminates these weather implications.
2) The MARSOT screener's intended focus is to highlight those participants that perform the best across all events of the screener in order to select them for summer training. With summer training opportunities being very competitive, using all means of comparison (allowing supermaxes) is critical to separating out the highest performers during each event. We establish the maximum time per gender per screener, allowing for a more flexible scoring scale to aid in rewarding high performers.

The algorithm for the MARSOT screener is outlined in Figure 2.

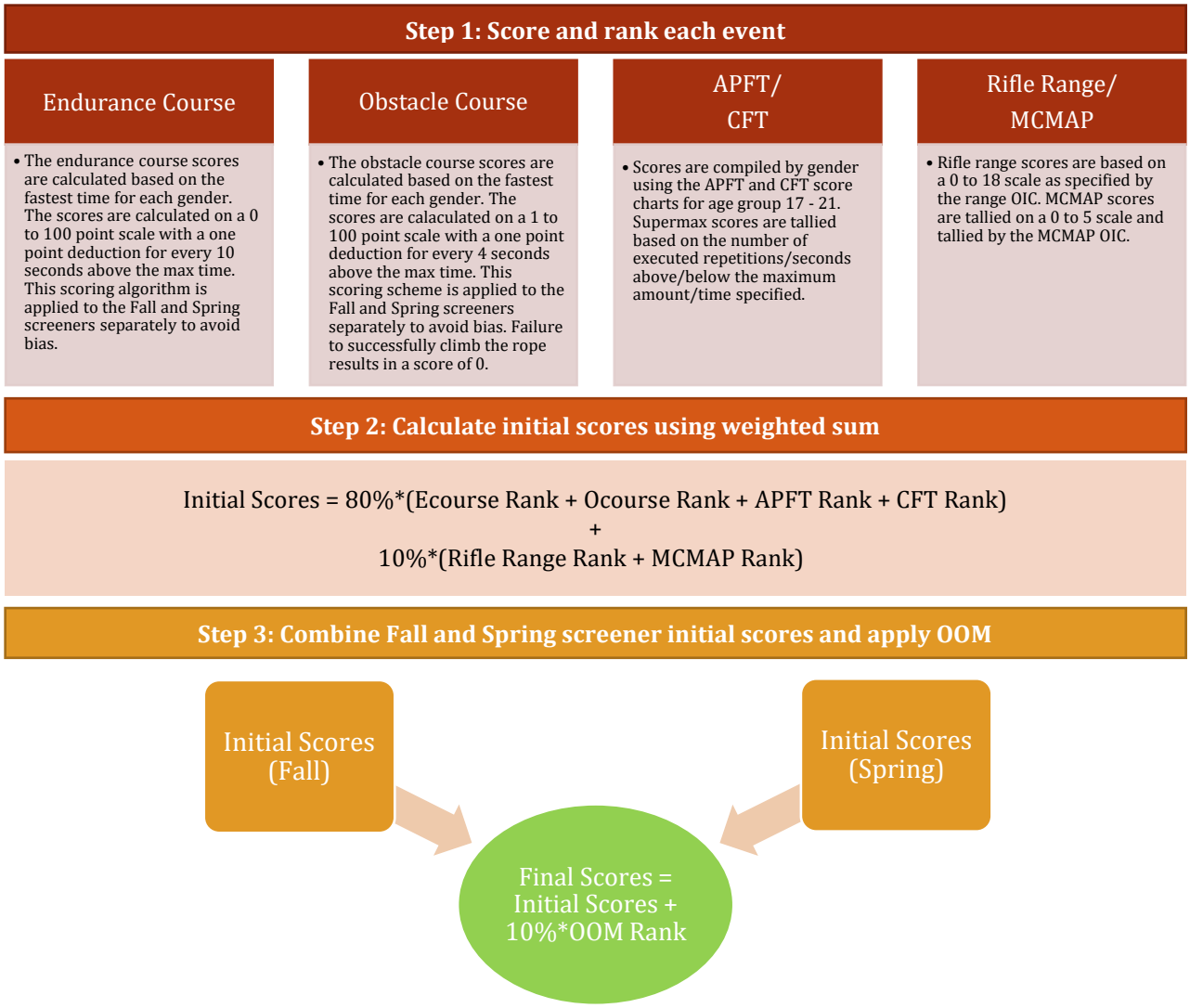


Figure 2: MARSOT Scoring Algorithm

Final scores for the MARSOT screeners are ranked and submitted to the USNA Summer Training Department for summer training slating. It is noted, the OOM ranking is applied to avoid ties in the scoring. The dominating factor for the MARSOT screener scores is performance across the four main physical events.

While the MARSOT scoring algorithm does not use a standard scoring table, we use the data gathered during the screener in order to build a quality scoring table that is similar to other USMC endurance courses. Without the use of this data, it is very difficult to construct a scoring system to align with the USNA endurance course and it's participants.

# Assumptions

Assumptions for constructing and implementing an endurance course scoring table are outlined below:

- The past four years of MARSOT screeners represent a quality distribution of endurance course completion times.
- For statistical purposes, we assume our data comes from a normal distribution.
- The participants are not expected to complete the obstacle course during the endurance course.
- The participants do not encounter significant water obstacles during the event.
- All participants run the endurance course in full NWUs without wearing extra gear.

These assumptions are critical to ensuring our analysis is as accurate and relatable as possible to other similar scoring tables across the USMC – specifically OCS.

# Analysis

The objective for the endurance course scoring table is to follow, as reasonably as possible, the distribution of completion times across genders. The statistics outlined in Table1 feature three subcategories separated by gender: 1) All: all completion times, 2) Historic: completion times prior to Fall 2021, and 3) Fall 2021. Figure 2 provides an illustration of the time distributions of subcategories 2 and 3.

| **Male** | | | | **Female** | | | |
|---|---|---|---|---|---|---|---|
| | *1: All* | *2: Historic* | *3: Fall 2021* | | *1: All* | *2: Historic* | *3: Fall 2021* |
| Count | 285 | 721 | 180 | Count | 196 | 180 | 16 |
| Min | 16:03 | 16:03 | 18:07 | Min | 20:13 | 20:13 | 23:18 |
| Max | 39:54 | 39:54 | 37:49 | Max | 46:33 | 39:28 | 46:33 |
| Average | 24:07 | 24:12 | 23:50 | Average | 29:18 | 29:08 | 30:27 |
| Median | 23:40 | 23:46 | 23:42 | Median | 29:04 | 28:45 | 30:06 |
| Stand Dev. | 03:09 | 03:14 | 02:48 | Stand Dev. | 04:26 | 04:13 | 05:45 |

Table 1: Statistics by Gender and Subcategory

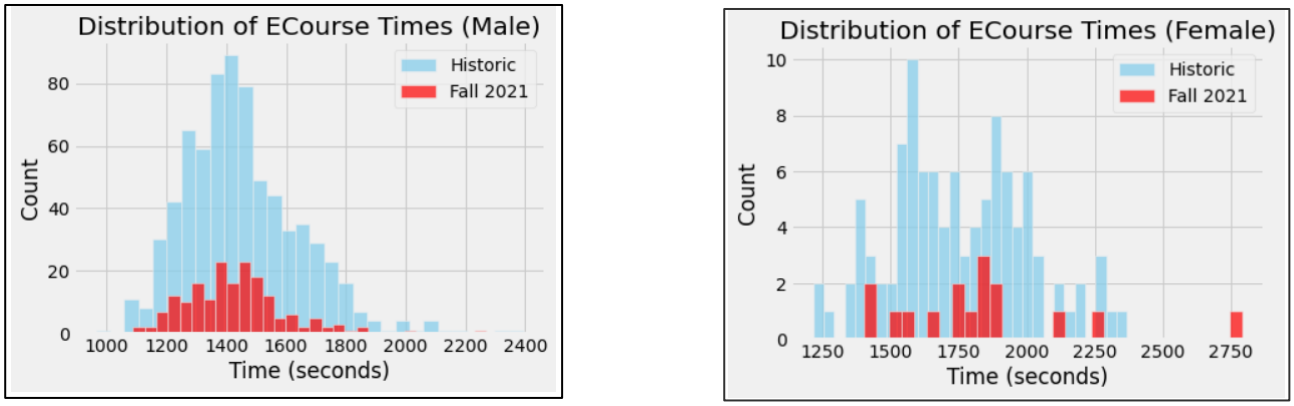


Figure 2: Distribution of Completion Times by Gender

From Table 1, it is expected the average and median completion times will be different across the genders; however, we see a significant difference in the standard deviation of completion times. The standard deviation for males is almost one and a half minutes quicker than for females in subcategory 1, and almost three minutes quicker in subcategory 3. This sharp difference in variation accounts for a large disparity between the two genders. After performing both an F-test and Levene test, we conclude there is enough evidence to support the variation among the two genders is statistically significant. We use this critical information when constructing our 1-point decrement system for the scoring tables.

We construct our scoring system using data from subcategory 2 (Historic Times) and utilize subcategory 3 (Fall 2021) to validate the results. The scoring system is set to a maximum score of 100 points and minimum score of 60 points – allowing for an average score of 80. We use the 10$^{th}$ percentile, 50$^{th}$ percentile, and 90$^{th}$ percentile as the method to establish the maximum, average (in this case median), and minimum score respectively. For example, a male participant who runs the endurance course in 23:46 (median completion time for subcategory 2), should achieve a score around 80 points.

Next, we create a 1-point decrement system to allow for a distributed point reduction based on specified completion times. To establish the 1-point decrement system, we use the middle 80% of completion times divided by the difference between the maximum and minimum points that can be achieved. This calculation results in a 1-point reduction for every next even percentile. For example, the 12$^{th}$ percentile earns a 99, the 14$^{th}$ percentile a 98, etc. Because the scoring table must be objective and not relative to percentiles in a specified data set, we calculate the average time difference between the even percentiles and use the time difference to determine when to apply the 1-point decrement.

Table 2 outlines the values for the maximum and minimum scores, along with the average time interval for the 1-point decrement system. We "smooth" the values in the table to provide a more reasonable scoring approach. For example, the exact mean time difference for females is 15.6 seconds; however, we round down to 15 to provide an easier score calculation. We note the mean time difference to apply a 1-point reduction is different for males and females, 10 seconds and 15 seconds respectively. This difference results from our previous conclusion that the variance across the genders is statistically significant;

therefore, the time difference for females must be slightly larger than for males in order to achieve a non-biased scoring table.

| | Male | Female |
|---|---|---|
| 10th Percentile = 100 points | 20:30 | 23:30 |
| 50th Percentile = 80 points | 23:50 | 28:45 |
| 90th Percentile = 60 points | 28:30 | 34:00 |
| Mean Time Difference = 1-point reduction | 10 seconds | 15 seconds |

Table 2: Scoring Metrics

Using the metrics from Table 2, and calculating the 1-point decrements based on the mean time differences, we provide the completed scoring tables for males and females in enclosure (1).

# Score Comparison

The final procedure is to apply the scoring tables to subcategory 3, Fall 2021, participants and analyze the results. Because we used data from subcategory 2 to construct the scoring table, we perform an Analysis of Variance (ANOVA) test across subcategory 2 and 3 to validate no significant statistical difference between the two subcategories – data from subcategory 2 and 3 are from the same distribution. The results in applying scoring tables across the participants for Historic and Fall 2021 completion times are illustrated in Table 3 and Figure 3.

| Male | | | Female | | |
|---|---|---|---|---|---|
| | 2: Historic | 3: Fall 2021 | | 2: Historic | 3: Fall 2021 |
| 10th percentile | 95 | 95 | 10th percentile | 92 | 88 |
| 50th percentile | 80 | 81 | 50th percentile | 78 | 77 |
| 90th percentile | 52 | 58 | 90th percentile | 51 | 59 |

Table 3: Score Comparison between Historic and Fall 2021

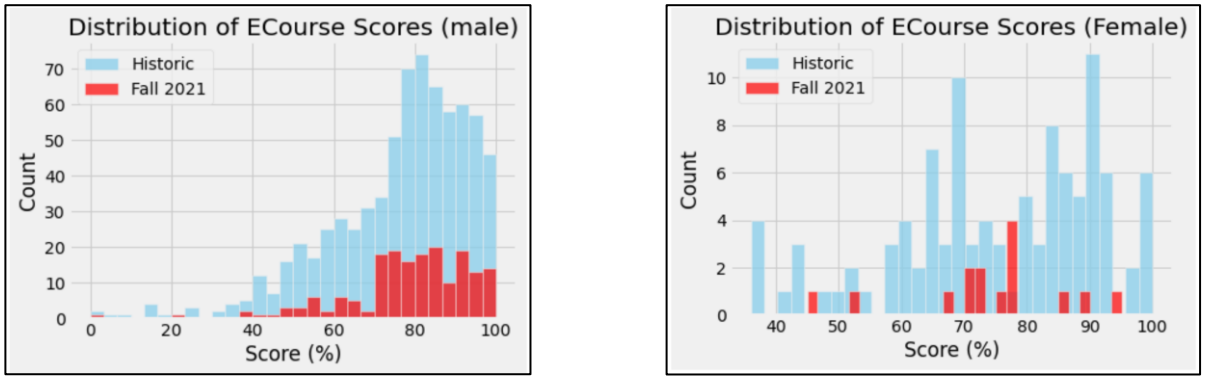


Figure 3: Distributions of Scores by Gender

As we relate the scoring percentiles from the subcategory 2, Historic, to subcategory 3, Fall 2021, we see a slight increase in score values for males and a small decrease in score values for females. Referring back to

Table 1, we note the average and median completion times for males for Fall 2021 is slightly faster than the Historic subcategory. Conversely, the average and median completion times for females for Fall 2021 is over a minute slower. These minor differences in metrics across the two groups for each gender account for the increase/decrease in scores for the two subcategories. The results from this comparison help validate that the scoring tables constructed in enclosure (1) are viable and reasonable for future application of endurance course participants.

# Conclusion

We set out to construct a standardized scoring table similar to OCS. Utilizing four years of previous MARSOT screener data, we produced a set of scoring tables to accurately score endurance course participants. The scoring tables are constructed by setting the $10^{th}$ percentile as the maximum score (100) and the $90^{th}$ percentile as the minimum score (60), with the $50^{th}$ percentile set as the average value (80). A 1-point decrement system is set utilizing the mean time difference among specified percentiles. Finally, the scoring tables are validated using the most relevant MARSOT screener data (Fall 2021) and comparing the output.

While the scoring tables are highly accurate in scoring MARSOT screener contestants, we did not validate the scoring tables against a different population of endurance course participants. One common assumption, which has not been validated, is that the MARSOT participants are, on average, in better physical condition than the rest of the Brigade; therefore, with only our data set it is difficult to predict whether the scoring tables' accuracy holds when applied to the entire Brigade of Midshipmen.