Prometheus Bound: Modeling Unselected Population Performance on a Graduate Admissions Test

by John M. Boyer, PhD, MSPE; and Ruslan Kalitvianski, PhD, MSPE

I. Introduction

The well-known intelligence quotient (IQ) mean of 100 and standard deviation of 15 are characterizations of intelligence measurements of the population of adult human subjects, which, in this report, we will call the unselected population. The Greek letter μ (mu) is often used to denote the mean, and the Greek letter σ (sigma) is often used to denote standard deviation. Various high-IQ societies have entrance requirements that are at or near a certain number of standard deviations above the mean intelligence of the unselected population. For example, Mensa is sometimes called a 2σ society. Similarly, the International Society for Philosophical Enquiry (ISPE), the One-in-a-Thousand Society (OATHS), and the Triple Nine Society (TNS) are sometimes called 3σ societies because their entrance criteria are close to, but a little above, 3σ. One prominent high-IQ society, Prometheus, requires from its members a demonstrated intellectual power of at least 4σ above the mean.

Percentiles are typically used by high-IQ societies to formally characterize the intelligence levels they require for admission. For examples, Mensa requires the 98th percentile; ISPE, OATHS, and TNS select at the 99.9th percentile; Prometheus requires the 99.997th percentile. While these organizations have percentile thresholds that are proximate to 2σ, 3σ, and 4σ, respectively, most high-IQ societies focus on the percentile threshold. In addition to the above societies, examples include Intertel, Colloquy, Infinity International Society, and ePiq IQ Society, which require the 99th, 99.5th, 99.63rd, and 99.8th percentiles, respectively.

If a person earns a percentile score of P, this indicates that the person has performed as well as, or better than, P percent of the unselected population on some measure or test of intelligence. The overall performance of a sample of a population on a measure can be graphed using scores on the horizontal x-axis and, on the y-axis, the number of people in a sample of the population who obtained each score. When a randomly occurring phenomenon like intelligence is measured, the shape of the graph takes the form of a bell curve, called a normal distribution or normal curve. Given a particular score, S, on a measure, the corresponding percentile, P, can be thought of as the proportion of the area under the part of the normal curve at or to the left of the score S. To easily understand why this is so, think about scanning from left to right along the graph of all scores and adding up how many people achieved each score. The number of people with each score takes up vertical space at each score location, and the span of scores takes up horizontal space, so the sum is a measure of area. For example, given an IQ measure with μ=100 and σ=15, an IQ score of 146 corresponds to the 99.9th percentile because the sum of adults who scored at or below 146 is essentially 99.9 percent of the unselected population, based on the sample whose IQs were measured. Similarly, the area under the unselected-population normal curve for intelligence that is at or to the left of the IQ score of 146 is essentially 99.9 percent of the total area under the curve.

To facilitate meeting their entrance requirements, many high-IQ societies accept a wide range of intelligence measures, including results from certain graduate admissions tests. One challenge with using graduate admissions tests is that graduate admissions test takers as a population are significantly more intelligent than the unselected population. As a result, percentiles of area under the normal curve for intelligence of the unselected population do not mean the same thing as percentiles of area under the normal curve for intelligence of the population of graduate admissions test takers. To account for this difference, psychometricians at high-IQ societies calibrate the graduate admissions test score thresholds they accept with the IQ score or percentile thresholds that they accept. For example, Mensa states an admission threshold of the 98th percentile of the unselected population and the 95th percentile on both the Graduate Management Admission Test (GMAT) and another graduate admissions test called the Miller Analogies Test (MAT).

To join the Prometheus Society today, the only test one can currently take is the Miller Analogies Test. Curiously, the percentile corresponding to their MAT scaled-score admission threshold is the same as their stated percentile threshold relative to the unselected population. Did Prometheus Society simply "play it safe" and set a higher bar for admission?

In this report, we answer this question as follows. First, we present additional background information as well as the mathematical foundations needed to develop a model for the unselected-population normal curve that predicts MAT scaled scores. This includes a novel application of the first fundamental theorem of calculus that reduces our work by requiring the MAT scaled-score thresholds of only two high-IQ societies. Next, we present simple step-by-step computer code that searches for a normal-curve model that accurately predicts the MAT scaled-score entrance thresholds for two high-IQ societies: Colloquy and ISPE.[1] We then use the model to determine the Prometheus bound, i.e., the MAT scaled score that best fits the 4σ level of the unselected population. Finally, we use the model to estimate the IQ mean and standard deviation of the MAT-taker population, and we use it to predict the MAT scaled-score admission thresholds of various other high-IQ societies.

II. Background: The Miller Analogies Test and Entrance Thresholds for High-IQ Societies

The Miller Analogies Test (MAT) is a 60-minute, 120-item, "high-level mental ability test requiring the solution of problems stated as analogies."[2] It has been in use in the US for over 80 years for graduate-school candidate selection. The MAT is also used for admission purposes by high-IQ societies such as Mensa, Intertel, ISPE, and the Prometheus Society because it is "an efficient and effective way to sample reasoning processes and to measure verbal comprehension and analytical intelligence."[3] In point of fact, tests performed on human subjects established a high correlation between MAT scores and the Terman Concept Mastery Test (Form T), a standardized, high-ceiling, verbal IQ test used to measure adult IQs in the average to exceptionally gifted range.[4]

Up until 2004, test results for an MAT taker were reported as a raw score, along with percentile ranks within the entire group of MAT takers as well as the subgroup of the MAT taker's intended field of study. Up to eight forms of the test were in circulation at any time, and, since they varied slightly in difficulty, raw scores between forms weren't directly comparable. In 2004, Pearson addressed this problem by moving to a scaled-score basis. Scaled scores range from 200 to 600, the average being 400 and the standard deviation 25.[5]

Given this mean and standard deviation, it is natural to question the MAT entrance thresholds used by various high-IQ societies.[6] For example, one implementation of the 99.9th percentile threshold that ISPE and TNS use is a score of at least 146 on an IQ test scaled to μ=100 and σ=15. Since (146-100)/15=3.07, one can see that the threshold implements a requirement of 3.07 standard deviations above the mean. So why, then, do organizations like ISPE and TNS require only a 472 on the MAT when 400+(3.07×25)=477? Why do similar calculations reveal similar discrepancies for most other high-IQ societies?

The answer is that the mean and standard deviation reported for the MAT are not based on the unselected population but rather on a sample mostly drawn from graduate-school applicants, whose average intellectual ability can safely be assumed to be above the average intelligence of the unselected population. Hence, substituting 3.07 IQ-population standard deviations for 3.07 MAT-taker standard deviations equates the two different types of standard deviations, which is an error akin to "comparing apples to oranges." Both a mean and a standard deviation have implicit *units of measure* that are based on the size of, and the points earned by, a sample *drawn from a particular population*. For example, the mean is the rate of points per person but, more specifically, *points from person in population X* per *person in population X*. The formula for standard deviation is more involved, but the same principle applies. Put simply, changing the population changes the *X*, which changes the implicit units of measure.

To join Prometheus Society, one must earn an MAT scaled score of 500,[7] which is 400+(4×25). Given that the implicit units of measure in standard deviations prevent them from being used interchangeably across different populations, it is natural to then question why Prometheus, a 4σ society, has an entrance threshold that seems to do just that. In one email on the "OATHS Yahoo!" discussion list,[8] the email author suggested that Prometheus "doesn't pull any punches." In other words, while the other high-IQ societies had accounted for the differences in populations, the email author asserted that Prometheus Society was being more stringent by using the unselected population standard deviation on the MAT-taker normal curve so that "if you qualify for Prometheus with the MAT, you really qualified." In this report, we show that this is not the case; i.e., that, by coincidence, the issue with units of measure does not significantly affect the MAT scaled-score admission threshold of Prometheus Society, i.e., the "Prometheus bound."

To account for the units-of-measure issue in standard deviations, psychometricians at high-IQ societies determine a score threshold for a graduate admissions test based on comparing scores from members who have taken the graduate admissions test as well as an unselected population test. In this report, we rely on the correctness of the thresholds set by psychometricians of two high-IQ societies, along with the first fundamental theorem of calculus, to build an accurate MAT-test normal-curve model for the unselected population that can be used to predict MAT entrance thresholds for a number of high-IQ societies. In particular, we use the model to show that the current MAT scaled-score admission threshold for Prometheus Society is essentially appropriate insofar as it is reasonably consistent with the choices of other high-IQ societies. In fact, it is perhaps even a bit low and, therefore, certainly not too high.

III. Foundations: Mathematics for Modeling a Normal Curve

A normal curve with a mean of μ and a standard deviation of σ is described by a complex mathematical formula called the *probability density function* (PDF) that is parameterized by μ and σ. The formula was developed by the mathematician Carl Friedrich Gauss to represent the probabilities of occurrence of sample outcomes of many randomly occurring phenomena. Figure 1 depicts in blue a normal curve with μ=100 and σ=15. It was generated by running the Python computer-code function `norm.pdf(x, 100, 15)` for *x*-values in the range of 40 to 160 (which is +/- 4σ). The normal curve has a "bell" shape that is taller in the middle because certain randomly occurring phenomena have outcomes that occur more often near the mean.

The *cumulative distribution function* (CDF) is related to the probability density function in that the CDF sums up all the probabilities of the outcomes from -∞ to a given outcome or measurement score of *x*. So, for example, the CDF of *x*=+∞ is 1 because the sum total of probabilities of all possible outcomes is 100%. In other words, the total area under the entire probability density function (the normal curve) is 1. Since the CDF is related to the PDF, the CDF is also parameterized by μ and σ. For clarity, these parameters are provided after a vertical bar (|) that is read as the word *given*. For the normal curve in Figure 1, the CDF(100 | 100,15) is 0.5 because 50% of the outcomes occur at or

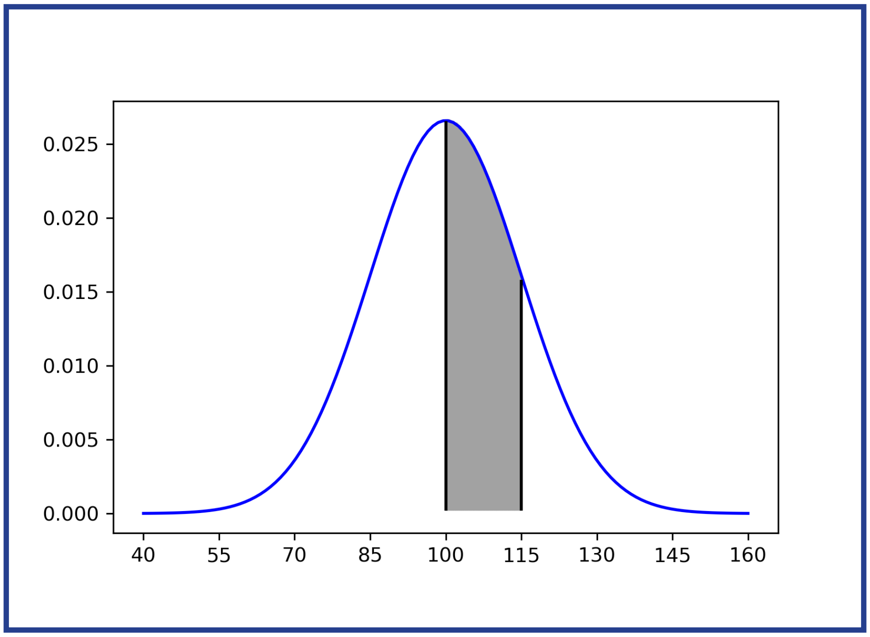

Figure 1*: A plot of the normal curve for intelligence (μ=100, σ=15) is in blue. The black vertical lines mark the horizontal lower bound and upper bound of the gray region between the mean and the first standard deviation. The area of the gray region represents the portion of outcomes that are expected to be between the mean and first standard deviation in a normal distribution.*

less than the mean of 100. More generally, the function CDF(x | μ, σ) computes the percentile of the outcome score x given a probability density function with a mean of μ and a standard deviation of σ, i.e., given PDF(x | μ, σ). For example, given the normal curve in Figure 1, CDF(130 | 100,15) is 0.97724987, so 130 is the closest IQ score to Mensa's 98th percentile threshold. As with the PDF, the CDF can be computed in Python using very similar code, e.g., running the code `norm.cdf(130, 100, 15)` produces 0.97724987.

By using subtraction, one can also compute the area under a normal curve between two scores, such as the gray region in Figure 1 above. Since CDF(115 | 100,15) is the area under the normal curve that is at or to the left of x=115, and CDF(100 | 100,15) is the area under the normal curve that is at or to the left of x=100, then CDF(115 | 100,15) minus CDF(100 | 100,15) gives the area of the gray region. It is a little more than 34.1%, which is the portion of the population sample's measured outcomes that is between the mean and the first standard deviation.

The term *integral* from calculus refers to a function (such as the cumulative distribution function) that can determine the area under the curve of another function (such as the probability density function). The *first fundamental theorem of calculus* proves that an exact numeric answer for the area under the curve of a given function, f(x), between a lower bound, x=a, and an upper bound, x=b, can be obtained by evaluating its area function, F(x), at the points a and b and then performing the subtraction F(b) – F(a). This is the same operation we performed in the preceding paragraph to determine the area of the gray region horizontally bound by x=100 and x=115. In mathematical notation, the first fundamental theorem of calculus states:

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

The left side of the formula starts with the integral symbol, a stylized "S" that means "sum" up a number of values that are created based on the expression that follows. The constant parameters *a* and *b* give the lower and upper bounds of the summation. The expression that follows the integral symbol represents the length-by-width area of a rectangle. The function f(*x*) gives the rectangle height (length), and *dx* is the width. Putting it together, the process is to first create many adjacent rectangles of width *dx* along the *x*-axis from *x*=*a* to *x*=*b*. Each rectangle has a length given by the *y*-axis value for the function f(*x*) evaluated at the *x*-axis location between *a* and *b* where the rectangle is located. The sum of the areas of these rectangles approximates the area under the curve of f(*x*), except that the rectangle corners may slightly overlap the curve or leave small gaps (see Figure 2). The integral symbol expresses the operation of letting the width *dx* tend toward 0, which makes the number of rectangles tend toward infinity. As *dx* shrinks to 0, so, too, do the sizes of the overlaps and gaps of the rectangle corners, resulting in an increasingly precise measurement of the area under the f(*x*) curve.

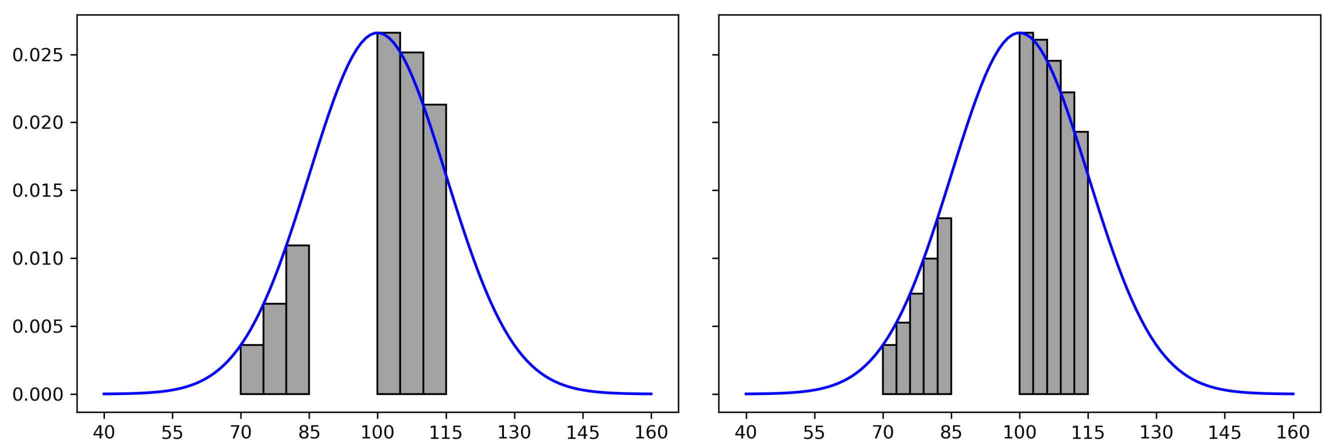

***Figure 2**: Both diagrams use a blue curve to show the function f(x) = PDF(x | 100, 15). In the left diagram, histogram rectangles show an approximation of the area under the curve in an x-axis range, such as 70 to 85 or 100 to 115. The gaps and overlaps between the rectangles and the curve illustrate amounts of imprecision (underestimation or overestimation) in the area approximation. The right-hand diagram shows that increasing the number of rectangles and decreasing their width (dx) decreases the size and total area of the gaps and overlaps, which increases the precision of the area approximation.*

The first fundamental theorem of calculus is *fundamental* because the right-hand side of the equation shows exactly how to find out what happens as *dx* approaches 0 and the number of those increasingly skinny rectangles tends to infinity. To use the right-hand side, one must find an area function, F(*x*), for a given function, f(*x*), and then use it to calculate F(*b*) – F(*a*).

In our case, the theorem is used not in this fundamental way, but rather in a novel way to guide our method of searching for a probability density function based on two known values of the cumulative distribution function. To explain this further, it is helpful to restate the theorem's formula substituting PDF(*x* | μ, σ) for f(*x*) and CDF(*x* | μ, σ) for F(*x*):

$$\int_a^b PDF(x \mid \mu, \sigma)\, dx = CDF(b \mid \mu, \sigma) - CDF(a \mid \mu, \sigma)$$

Although we do not know the μ and σ of the probability density function that would best represent the unselected population normal curve for MAT scaled scores, we do know from ISPE,[9] TNS, and OATHS that we need a μ and σ such that CDF(472 | μ, σ) is 0.999; and from Colloquy[10] we know that, for the same μ and σ, CDF(455 | μ, σ) should be 0.995. By finding μ and σ such that the specific cumulative distribution function CDF(x | μ, σ) meets these two criteria, from the equation above, we know that μ and σ also give the desired probability density function PDF(x | μ, σ) because CDF(x | μ, σ) is the area function of PDF(x | μ, σ).

IV. Search: A Model for Predicting Unselected Population Percentiles on the MAT

In this section, we present the code and results of a computerized search for the mean and standard deviation of a normal curve that meets the requirements that a score of 455 corresponds to the 99.5th percentile and a score of 472 corresponds to the 99.9th percentile. Computer algorithms must necessarily take a finite number of steps and operate over numerical representations of finite size and precision. As a result, the search algorithm we present seeks the mean and standard deviation that creates a highly precise approximation of the required correspondences at a finite, yet acceptable, level of precision. Therefore, it makes sense to first examine what level of precision should be deemed acceptable.

Insofar as this paper is first and foremost about the Prometheus bound, we adopt the level of precision articulated by the Prometheus Society on the home page of its website: “The Prometheus Society, however, discriminates at the 99.997th percentile, which equates to ‘1 in 30,000’ (four standard deviations above the norm).”[11] Of course, the three quantities in the quoted sentence are not equal, but they are high-precision approximations of one another. The 1 in 30,000 selectivity level corresponds to the 99.99666th percentile, and a selectivity level of four standard deviations corresponds to a percentile of 99.996833. Since the range of the values that Prometheus “equates” is $3.333\ldots\times10^{-6}$, we use that number as the precision target for numerical approximations that should be deemed sufficiently close, given the finite limitations of computation.

We begin by examining whole-number means and standard deviations because the computer code is easiest to understand, produces the most easily memorable results, and because we find that the results meet the aforementioned precision target. However, for the sake of completeness, and because the authors are Thousanders (members of ISPE), we end the section by presenting the results of a higher-resolution search down to the thousandths place.

The search algorithm begins with initializing program variables that help detect increasingly precise results as they are found:

```
best_mean_found = 0
best_sd_found = 0
best_precision_found = 1
```

The equal sign (=) in the above computer code represents an assignment of the right-hand-side value to the left-hand-side variable. This initialization represents the concept that nothing useful has yet been found before the searching begins. This is true because we expect the mean MAT scaled score of the unselected population to be at least 200 (the minimum score), because we also expect the standard deviation to be a non-zero number (since people don’t all get the same score on the MAT), and because we have a precision target that is at or below $3.333\ldots\times10^{-6}$ (which is far smaller than 1).

The search algorithm then performs a programming construct called a “loop” that analyzes all mean values within a given range. The lower bound of the range is 200. It is reasonable to set an upper bound of 400 because that is the mean for the graduate-school applicants whom we believe to have higher average intelligence than the unselected population. If we do not get a good result within this range, it will be easy to revise the search later, so we begin with Python code that looks like this:

```
for mean in range(200, 400 + 1):
    Code to run for each mean value
```

In Python, the range is interpreted as including the lower end and excluding the upper end. Since we want to include the upper end of the range, we just add 1.

Any lines of Python code indented under the "for" loop line are collectively called the "body" of the mean loop, and they are iteratively performed while setting the mean variable equal to each successive whole-number value in the range. The search algorithm does two things in the body of the mean loop. First, given the mean variable value μ, we use an additional "inner" loop to find the standard deviation σ such that CDF(455 | μ, σ) most closely matches the value 0.995. Second, a conditional logic construct called an "if" statement is used to determine whether the proximity of CDF(472 | μ, σ) is closer to 0.999 than the best mean and standard deviation tested in any prior iteration of the "mean" loop.

The additional loop to find the standard deviation is called an "inner" loop because it is inside the body of the mean loop above. We can change the code easily if our initial assumption is incorrect, so we begin with a wide range around the MAT-taker standard deviation of 25, as follows:

```
best_sd = 15-1
for sd in range(15, 40 + 1):
    curr = norm.cdf(455, mean, sd)
    best = norm.cdf(455, mean, best_sd)
    if (abs(curr-0.995) < abs(best-0.995)):
        best_sd = sd
```

The variable best_sd is initialized to a value outside of the range of analysis. It is updated in any iteration of the sd loop in which the current iteration's sd value produces a CDF result for 455 that is closer to 0.995 than was found with the prior value of best_sd. Closeness is determined using the absolute value of the difference between a CDF result and the desired CDF value of 0.995. The CDF calculation uses the mean variable value in the current iteration of the "outer" mean loop that surrounds the "inner" sd loop. In this way, every standard deviation in the range of the inner loop is tried for each mean in the outer loop's range.

In the rest of the mean-loop body, the best_sd found by the inner sd loop is used with the current iteration's mean value to determine whether the pair of values produces a CDF result for 472 that is closer to 0.999 than was found in the prior iterations of the mean loop. If so, then the variables best_precision_found, best_mean_found, and best_sd_found are updated. The Python code to perform these operations is below, which completes the search algorithm:

```
curr = norm.cdf(472, mean, best_sd)
if (abs(curr-0.999) < abs(best_precision_found)):
    best_precision_found = curr-0.999
    best_mean_found = mean
    best_sd_found = best_sd
```

Using the search algorithm above, the best model for MAT scaled scores of the unselected population is a normal curve with a mean of **370** and a standard deviation of **33**. Based on this normal-curve model, Table 1 below presents the exact percentile for the MAT scaled scores 455 and 472 and their precision results, i.e., the proximity of the percentiles to the desired percentiles of 0.995 and 0.999.

Table 1: Precision Results

MAT Scaled Score (x)	CDF(x | 370, 33)	Precision
455	0.99499896	1.04×10^{-6} *
472	0.99900228	2.28×10^{-6} *

**More accurate than the precision target of* $3.333\ldots\times10^{-6}$

In addition to being more accurate than the precision target, the PDF(x | 370, 33) normal-curve model accurately predicts the entrance threshold of the largest high-IQ society, Mensa. For the unselected population, the stated threshold is the 98th percentile, or an IQ of 130, which we know from above is the 97.724987th percentile. On the MAT, Mensa accepts the 95th percentile.[12] According to the publisher of the MAT, the 95th percentile corresponds to MAT scaled scores of 436 to 438.[13] In Table 2, we show that a score of 436 corresponds exactly to the percentile for an IQ of 130, and a score of 438 is the lowest score to exceed the 98th percentile.

Table 2: Predictive Results for Mensa

MAT Scaled Score (x)	CDF(x | 370, 33)
436	0.97724987
437	0.97883713
438	0.98032968

Given the accuracy of the Mensa prediction along with the precision at the 99.5th and 99.9th percentiles, we can now confidently proceed to predicting the Prometheus bound. As mentioned previously, Prometheus Society sets its unselected population threshold in the range of 1 in 30,000 selectivity (99.99666… percentile), an IQ of 160 (99.996833 percentile), and the 99.997th percentile. In Table 3, we see that the current Prometheus Society MAT scaled-score threshold of 500 is below the low end of the desired percentile range. The 99.995916th percentile has a selectivity of 1 in 24,390. A higher MAT scaled score of 502 has exactly the same percentile as an IQ of 160. As Table 3 shows, no other MAT scaled score is in the range set by Prometheus Society, so we assert that the Prometheus bound should be a score of 502 on the MAT.

Table 3: The Prometheus Bound

MAT Scaled Score (x)	CDF(x | 370, 33)
500	0.99995916
501	0.99996402
502	**0.99996833**
503	0.99997215

This is an interesting result because it shows that the 4σ level in the unselected population is actually *more* stringent than the 4σ level in the MAT-taker population. In the next section, we compare the two normal curves with visualizations to show the exact score beyond which the unselected population curve becomes more stringent.

Finally, for completeness, we slightly amended the code presented above to examine all means and standard deviations down to the thousandths place rather than just whole numbers. The results were a mean and standard deviation very near the whole-numbered values we've been using. The best model produced had a mean of 369.874 and a standard deviation of 33.048. This model is even more accurate, with the CDF of 455 being closer to 0.995 than 3×10^{-9}, and the CDF of 472 being closer to 0.999 than 3×10^{-10}. However, this additional accuracy has virtually no effect on the predicted Prometheus bound. The CDF(502 | 369.874, 33.048) is 0.99996806, and the score of 502 is still the only score whose CDF is in the range expressed by Prometheus Society. Since the more accurate model produces a difference without a distinction, and the model based on whole numbers had better accuracy than the precision target, we recommend using a normal curve with a mean of 370 and standard deviation of 33 to model MAT scaled scores for the unselected population.

V. Analysis and Visualization: Comparing the Normal Curves

To illustrate our analysis graphically, in Figure 3 below we plot the MAT scaled-score normal curves for the MAT-taker population and the unselected population as extrapolated by our search algorithm in the previous section. All the plots below were drawn using Microsoft Excel 365 v.1812, using the `NORM.DIST() function`.

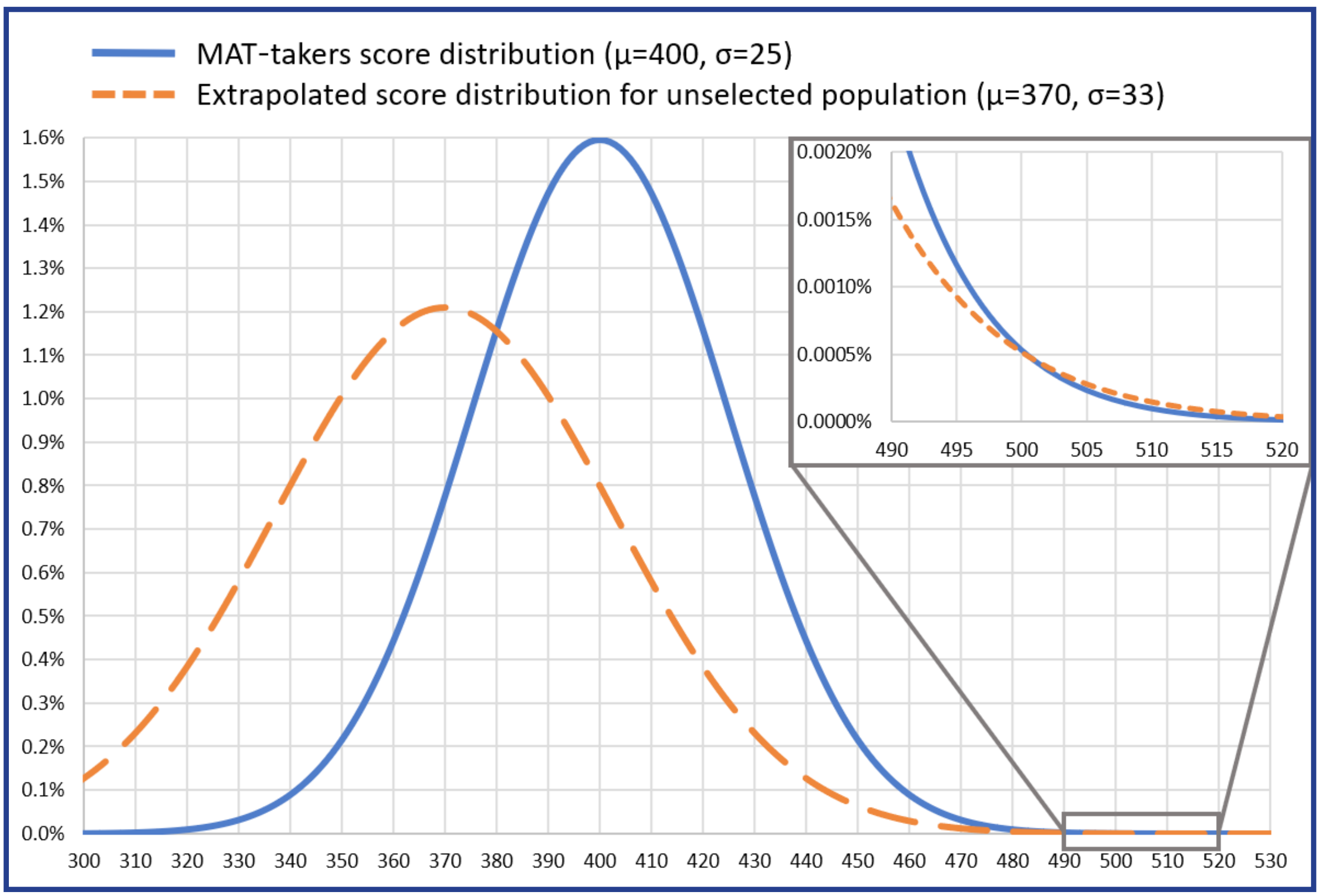

***Figure 3**: A plot of the MAT scaled-score distribution curves for the MAT norming cohort and the estimated score distribution for the unselected population. In the frame, the SS 490-520 region is enlarged.*

This first plot illustrates that the peak of the unselected-population curve is significantly to the left of that of the MAT-taker population, meaning that the MAT-taker population is, on average, considerably more intellectually select. However, the unselected-population curve is also wider, such that it converges with the MAT-taker curve near the Prometheus bound. The curves intersect just before a score of 501, which is coincidentally close, but not equal to, and not related to, the score of 502, which we recommend for the Prometheus bound based on percentile.

In Figure 4, we switch from examining the two normal curves to examining the percentiles of the two normal curves, especially in the range where they converge. For *most* of the range, the CDF value (percentile) of the MAT-taker population at a given MAT score is less than that of the unselected population. This means that it is a lesser achievement for a member of the MAT-taker population (graduate-school applicants) to achieve an MAT score than it is for a member of the unselected population. But, due to greater width of the unselected-population curve, the unselected-population CDF "catches up" with the MAT-population CDF just before the scaled score of 494, as can be seen in Figure 4.

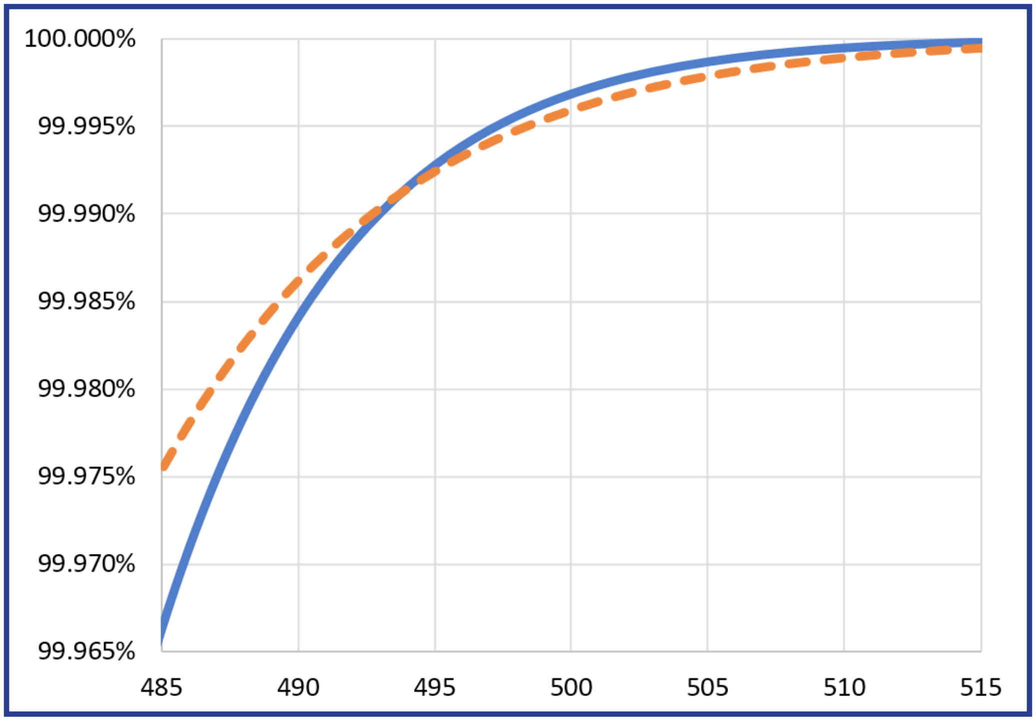

Figure 4*: The 485-515 scaled-score range is enlarged, showing convergence of both CDF curves and subsequent reversal of dominance. The blue MAT-taker curve reaches 99.996833% at a score of 500, but the dashed unselected-population curve rises to that level only at a score of 502.*

The even more counterintuitive finding is that, past that convergence point, the MAT-taker-population CDF curve dominates the unselected-population CDF curve, indicating that one who gets an MAT scaled score of 494 or higher has a higher percentile in the MAT-taker population than would be expected if the MAT were normed with the unselected population. In other words, one must select a *higher* MAT scaled score to reach a desired percentile level in the unselected population than the MAT scaled score needed to reach that same percentile in the MAT-taker population. This is precisely why the 99.997th percentile is reached at the MAT scaled score of 502 in the unselected-population-extrapolated normal curve and at 500 in the MAT-taker normal curve. A horizontal line at the

99.997th percentile on the *y*-axis in Figure 4 intersects the MAT-taker CDF curve at x=500 but would only intersect the unselected-population CDF curve at x=502.

So far, we have used the unselected-population CDF curve to extrapolate the relation between specific MAT scaled scores and specific percentiles of selectivity in the unselected population. However, the CDF curve is continuous, so it is possible to use percentile equivalency to create a linear formula that estimates IQ scores based on MAT scaled score. This is justified for two reasons that were introduced in the background (Section 2). The first is that both are normal curves for the same target population, the unselected population; and the second is the aforementioned high correlation between MAT scores and verbal IQ. We can therefore consider MAT percentiles as reasonable approximations for verbal IQ percentiles, as have many high-IQ societies, and derive a simple linear conversion function with the following general form:

IQ estimate = Mean IQ of the MAT population +
((MAT Scaled Score – 400)/25) × IQ SD of the MAT population

Let's begin by estimating the IQ of the average MAT taker. Since the estimated mean of the unselected population on the MAT is 370, the standard deviation is 33, and the mean MAT scaled score is 400, we deduce that this score is (400 - 370)/33 = 0.9091σ above the unselected population average. Thus, assuming mean IQ of 100 and standard deviation of 15, the IQ corresponding to an MAT scaled score of 400 is 100 + (15 × 0.9091) = 113.636.

Next, we can convert the MAT's σ=25 into its IQ σ equivalent: 15 × 25/33 = 11.364. This means that per each variation of 25 MAT scaled-score points, the equivalent IQ varies by 11.364 IQ points. Substituting these two values into our general formula, we obtain:

IQ estimate = 113.636 + ((MAT Scaled Score – 400)/25) × 11.364

As an illustration of the soundness of this formula, Table 4 below provides percentile and estimated IQ values for MAT scaled scores for the mean and each standard deviation of the MAT-taker population up to 4σ above mean.

Table 4: Mapping of MAT Scaled Scores, Percentiles, and Estimated IQ equivalents

MAT Scaled Score (x)	Percentile CDF(x | 370, 33)	IQ Estimate (based on Percentile)	Difference from Preceding IQ Estimate
μ + 0σ = 400	0.81834893	113.636	n/a
μ + 1σ = 425	0.95220965	125	11.364
μ + 2σ = 450	0.99232982	136.364	11.364
μ + 3σ = 475	0.99926823	147.727	11.364
μ + 4σ = 500	0.99995916	159.091	11.364

A linear conversion formula can also be determined without computing the mean and standard deviation for the MAT-taker population. A simplified formula can be expressed only in terms of the MAT scaled score and the means and standard deviations for the unselected-population IQ and MAT normal curves. Thus, via algebraic manipulation, we can factor out the specifics of the MAT-taker population as follows:

$$\begin{aligned}\text{IQ estimate} &= 100 + 15 \times (400 - 370)/33 + ((\text{MAT Scaled Score} - 400)/25) \times (15 \times 25/33)\\ &= 100 + (15/33) \times (400-370) + (15/33) \times (\text{MAT Scaled Score}) - (15/33) \times 400\\ &= 100 + (15/33) \times (\text{MAT Scaled Score} - 370)\end{aligned}$$

This reduction has the advantage of affranchising the expression from references to normative values for current *or future* MAT-taker populations.

Over time, MAT norms do evolve. So far, in this work we have assumed that the MAT average score is 400 and standard deviation is 25, based on the *Miller Analogies Whitepaper*.[14] What if a renorming were to change these values? A recently published technical memo by Pearson reports MAT percentiles based on the latest MAT norming cohort.[15] According to this document, the scaled score corresponding to the 50th percentile of this group is 396. This is a small difference, given the standard deviation, but enough to warrant examining how much it may impact conclusions we draw about the high range.

Unfortunately, the table ends where things get the most interesting for ultra-high-IQ societies, as no percentile above the 99th is reported. Fortunately, we can model the curve beyond that point ourselves. In a normal distribution, irrespective of the values of μ and σ, $CDF(\mu - 1\sigma \mid \mu, \sigma)$ corresponds to about the 16th percentile, $CDF(\mu + 1\sigma \mid \mu, \sigma)$ to about the 84th, and $CDF(\mu + 2\sigma \mid \mu, \sigma)$ is just a little below the 98th. Thus, when one knows the mean of a normal distribution and its value at one of those percentiles, one can deduce a close approximation for σ by subtraction. Let's see if this works for the MAT values reported in the table from Pearson.[16]

The 16th percentile is reported to be between 375 and 376, the 84th percentile is 420, and the 98th percentile is between 447 and 451. This gives us the following standard deviation candidates:

- 396 – 375 = 21
- 420 – 396 = 24
- ([447, 448, …, 451] – 396)/2 = [25.5, 26, …, 27.5]

All of the values for σ thus computed are different, which indicates that this distribution doesn't perfectly follow the normal distribution. A histogram of the scores in the *MAT Basics* booklet shows that the distribution is slightly skewed on the left and much closer to being normal in the right tail.[17] Since that is the range of scores in which we are interested, we will assume normality in the right tail and select $\sigma=26$ as the standard deviation that best accommodates the scaled-score range at and above the 98th percentile. This yields an MAT scaled score of 448 for the 98th percentile of the MAT-taker population.

Figure 5 shows the new normal curve for $\mu=396$ and $\sigma=26$. Although the new curve is slightly shifted left from the canonical blue curve, our extrapolated model for the unselected population remains accurate for the high range. In Table 2 of Section 4, we already used the extrapolated model to accurately map the *updated* 95th percentile scores[18] to the Mensa range (the range between 2σ and the 98th percentile). Furthermore, Figure 6 shows that the updated mean and standard deviation has virtually no impact on the CDFs in the 4σ range, as they are equal for both curves at the MAT scaled score of 500, and they match out to the sixth decimal place at the MAT scaled score of 502. Therefore, even with these slightly changed assumptions, our conclusion about the Prometheus bound stands.

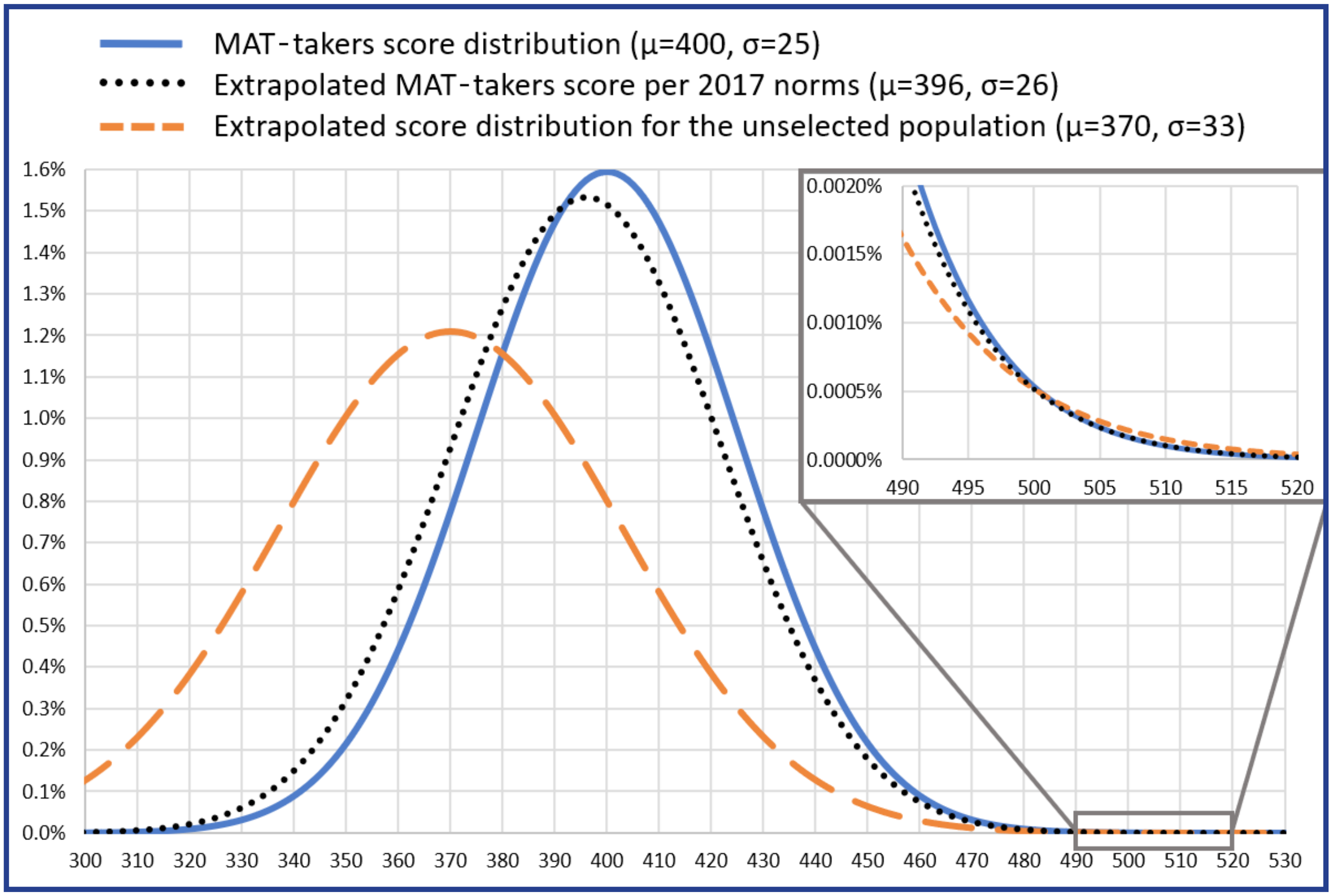

Figure 5*: A plot of the MAT scaled-score distribution curves for the MAT norming cohort, another possible MAT score distribution, and the estimated score distribution for the unselected population. In the frame, the SS 490-520 region is enlarged.*

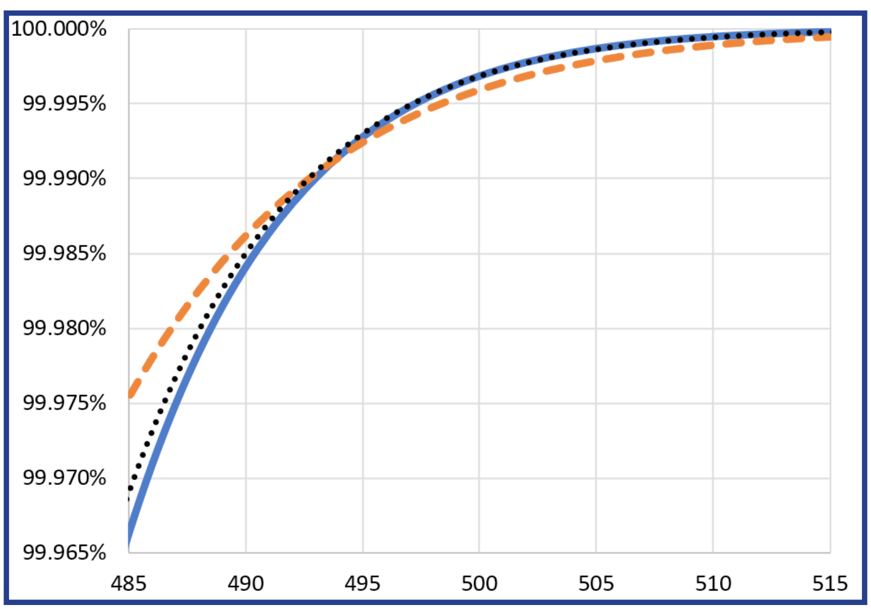

Figure 6*: The 485-515 scaled-score range is enlarged, demonstrating proximity of the* CDF(X | 400, 25) *and the* CDF(X | 396, 26) *curves.*

VI. Conclusion

In this report, we have come to the conclusion that the Prometheus Society's admission requirement, a scaled score of at least 500 on the Miller Analogies Test (MAT), is reasonably consistent with the admission requirements of other high-IQ societies on the MAT. For most high-IQ societies' admission thresholds, the percentile associated with the MAT scaled score is lower than the percentile required on an IQ test because IQ is a measure of intelligence relative to the unselected population. The MAT, however, is administered predominantly to graduate-school applicants. A point of contention was whether Prometheus Society's admission requirement was too high, because the percentile associated with the required MAT scaled score is the same percentile as their selectivity requirement within the unselected population. However, we found that the percentiles of the two populations converge at a slightly lower MAT scaled score than 500, and after that point of convergence, the percentile pattern is reversed. For example, a scaled score of 500 on the MAT essentially corresponds to the 99.997th percentile among the population of graduate-school applicants, but a higher MAT scaled score of 502 is required to reach the same percentile of the unselected population. Therefore, we recommend that the Prometheus Society raise their admission requirement to an MAT scaled score of 502 in order to be more consistent with the admission requirements of other high-IQ societies.

To determine consistency among high-IQ societies, we used a computerized search for a normal-curve mean and standard deviation that produced the closest match of MAT scaled scores to percentiles using the admission requirements of two high-IQ societies: Colloquy and ISPE. The efficacy of using two samples to search for a normal-curve model was based on a novel application of the first fundamental theorem of calculus. The result of the search was that a normal curve with a mean of 370 and a standard deviation of 33 accurately modeled the performance of the unselected population on the MAT. Using this extrapolated normal-curve model, we were able to accurately match the entrance requirements for Mensa and show that the Prometheus Society admission threshold is not overly stringent. We were also able to estimate the IQ mean and standard deviation for the population of graduate-school applicants to be about 113.636 and 11.364, respectively. Finally, in Table 5 below, we present the results of using the extrapolated model to set consistent MAT scaled-score entrance requirements for several other high-IQ societies.

Table 5: Admission Requirement Recommendations for Other High-IQ Societies

High-IQ Society Name	Percentile Required by Society	IQ Required (σ=15)	Percentile of IQ Required	Recommended MAT Scaled Score	Percentile of MAT Scaled Score
Intertel	99	135	99.018467	447*	99.018467
Infinity International	99.63	140	99.616962	458†	99.616962
ePiq	99.8	143	99.792590	465*	99.800397
Epimetheus	99.997	160	99.996833	502*	99.996833

* *No currently set MAT scaled-score admission threshold*

† *Currently has a higher MAT scaled-score admission threshold than expected, compared to other high-IQ societies*

All the computer code used for this article is freely available at this address:
https://github.com/john-boyer-phd/Prometheus-Bound

NOTES

1. Two other 3σ societies, TNS and OATHS, use the same MAT scaled-score threshold as does ISPE.

2. Don Meagher, *Introduction to the Miller Analogies Test: Assessment Report*, Harcourt Assessment (January 2006), 2, https://images.pearsonclinical.com/images/PDF/assessmentReports/MillerWhitepaper.pdf; ten sample items are available on page 13 of the official MAT Study Guide, https://images.pearsonassessments.com/Images/dotCom/milleranalogies/pdfs/TheMATStudyGuide.pdf.

3. Meagher, 2.

4. Norman E. Wallen and Mary Lou A. Campbell, "Vocabulary and Non-verbal Reasoning Components of Verbal Analogies Tests (Miller Analogies Test and Concept Mastery Test)," *The Journal of Educational Research* 61, no. 2 (1967): 87-89, https://doi.org/10.1080/00220671.1967.10883594; Richard W. Johnson and Stewart D. North, "Use of Psychological Tests in an Administrative Staff Improvement Program," Counseling Center Reports, The University of Wisconsin-Madison, Counseling Center (August 1969), http://archive.org/details/ERIC_ED049270.

5. Meagher, 11.

6. As did occur on the ISPE private social network: https://the1000.ning.com/profiles/status/show?id=1988210%3AStatus%3A119787&commentId=1988210%3AComment%3A119788&xg_source=activity.

7. The Prometheus Society, "Join or Subscribe," http://prometheussociety.org/wp/join-or-subscribe/.

8. Message "RE: [OATHSociety] Mensa gets smart(er) about the MAT," March 17, 2008, https://groups.yahoo.com/neo/groups/OATHSociety/conversations/messages/3597.

9. International Society for Philosophical Enquiry, "Tests & Test Scores," https://www.thethousand.com/tests_test_scores.php.

10. Colloquy, "In the Spirit of Collegial Inquiry... Which Tests Qualify at Percentile 99.5?" http://colloquysociety.org/col81sco.htm.

11. The Prometheus Society, "Welcome," http://prometheussociety.org/wp/.

12. Mensa, "Qualifying Test Scores," https://www.us.mensa.org/join/testscores/qualifying-test-scores/.

13. Pearson, "Comparing MAT to GRE® Scores 2017-18," NCS Pearson, 2017, http://images.pearsonassessments.com/Images/dotCom/milleranalogies/pdfs/Comparing_MAT_GRE_Scores_2017-18.pdf.

14. Meagher, 11.

15. Pearson, 2.

16. Pearson 2017, 2.

17. Donald Meagher, Natividad Ybarra III, Tianshu Pan, Rachel Wegner, Jeffrey R. Miller, and Ariane Zamot, "MAT Basics: Test Structure and Score Interpretation," Pearson Clinical Assessment Group, 2017, https://images.pearsonassessments.com/Images/dotCom/milleranalogies/pdfs/MAT_Basics_2016_fnl.pdf.

18. Pearson. **Ω**