

PrepSharp ACT Test Relative Difficulty

David Reynolds, Ross Thompson, et. al.

February 2022 (Updated April 2023)

1 Introduction

We introduce a method for comparing the relative difficulty of ACT test sections (English, Math, Reading, Science), as well as ACT test composites, across different versions the test. Our method associates a probability distribution with each test version and relative difficulty is based on a comparison of the cumulative distribution function across tests. This allows one to compute a difficulty ranking, as well as a way to quantify the magnitude of the difference in test difficulty, across different test versions. The magnitude of the difference of the test difficulty was then used to compute a difficulty score for each test section and the overall test composites.

2 Methodology

To illustrate the key ideas and computational steps underlying our relative difficulty measure, this section begins with a simple example in which we compare the difficulty across two versions of a 4 question test. Then, we extend the computational steps to allow comparison across ACT test versions.

2.1 Simple Example

To illustrate the method, let us consider a simple example in which we compare the difficulty across two versions of a 4-question test, test X and test Y .

Suppose that we observe many students take each of these tests and observe a set of percentiles for each test, as shown in Table 1.

Test Percentiles		
Score	Test X	Test Y
4	1	1
3	0.75	1
2	0.5	0.95
1	0.25	0.75
0	0	0

Table 1: This table shows the percentile associated with each possible test score (i.e., number of total correct answers). For example, on Test X , the proportion of students that scored 3 or less is 0.75.

Another way to conceptualize the observed test percentiles in Table 1 uses probability distributions. From this perspective, Table 1 presents the cumulative distribution function (CDF) for each test. Furthermore, each test is associated with a discrete probability distribution over test scores. The CDF of each distribution is defined by the observed percentiles.

The CDF of a random variable is a function that completely describes the probability distribution of the random variable. The CDF of a random variable Z is defined as,

$$F_Z(z) = P(Z \leq z), \text{ for all } z \in \mathbb{R}. \quad (1)$$

In our analogy, the random variable for a given test is the number of correct answers (i.e., the score). Therefore, the range of the random variable is the set: $\{0, 1, 2, 3, 4\}$. For Test X , if we define the score as a random variable X , then $F_X(3) = 0.75$, since the probability of scoring less than or equal to 3 is 0.75, from Table 1. To further clarify this function, we plot the CDF for our example tests below.



Figure 1: The cumulative distribution functions (CDFs) for the two example tests. We treat the score as a random variable, whose distribution is defined by observed test percentiles. The CDF for a discrete distribution is a right-continuous function defined for all real numbers. The upper limit is 1 and the lower limit is 0.

Now, let us return to the motivating question of which test is harder. On Test X , the probability of a score of 1 or less is 25%, whereas on Test Y , this probability is 75%. Further, on Test X , the probability of a score of 2 or less is 50% whereas the same probability on test Y is 95%. Finally, on Test X , the probability of a score of 3 or less is 75%, whereas the same probability on test Y is 100%. If we assume that there are no differences across the test takers that generated these distributions, then clearly test Y is the harder test since fewer correct answers are required to achieve any given percentile.

In terms of probability distributions, this increased difficulty manifests as relatively more mass for lower scores and relatively less mass for higher scores. This notion can be formalized by defining the relative difficulty as the difference between the area under the CDF across the range of scores. This can be stated mathematically as,

$$\text{Relative Difficulty}(X, Y) = \sum_{i=1}^4 F_X(i) - F_Y(i) \quad (2)$$

$$= 1.7 - 2.7 = -1. \quad (3)$$

A positive relative difficulty indicates that the left side argument (X in the expression above) is more difficult; a negative relative difficulty indicates that the left side argument is less difficult than the right hand argument. Additionally, by looking across the whole range of scores, this measure takes into account the shape of the entire distribution, rather than simply comparing the number of points required to achieve a median score, for instance.

Next, we make some minor adjustments to this framework that allow us to define relative difficulty across different versions of the ACT test.

2.2 ACT Test Comparisons

In this section, we make some minor modifications to the simple example in the prior section to allow difficulty comparisons across all versions of a given ACT test section.

As with the prior example, we use the concept of a probability distribution to compare difficulties. The key difference between the relative difficulty measure that is implemented by PrepSharp and the measure defined in equation 2 is that we consider the probability distribution that is induced by a certain version of an ACT section (for example, the English section from F11, April 2023) to be a continuous distribution.

We make this change for two reasons. First, many test versions have gaps between the number of correct answers and the scaled test score, which is mapped to a percentile. For example, 51 correct answers on the F11 ACT English section maps to a scaled score of 24, which further maps to the 77th percentile. Consider where a scaled score of 25 (81st percentile) demands 54 correct answers. By considering the test distribution to be continuous, we interpolate that those students who answered between 51 and 54 questions correctly are somewhere between the 77th and 81st percentiles, which is closer

to reality than its discrete counterpart. The other advantage of using continuous distributions is it is easier computationally to numerically integrate than to work with summations.

As with equation 2, we define the relative difficulty, labeled as ρ , between two tests as the area between their CDF functions. However, since we are considering these as continuous distributions, we replace the summation with an integral. The range of integration is the range of the number of possible correct answers for a given section. On the English section, for instance, this range is (0,75). Therefore, for the English section, we can define the relative difficulty between two test versions, X and Y , as,

$$\rho(X, Y) = \int_0^{75} (F_X(t) - F_Y(t))dt. \quad (4)$$

As with the prior section, a positive score indicates that the left hand argument is more difficult, 0 indicates there is no difference, and a negative score indicates that the left hand argument is less difficult.

In figure 2, we show two CDF curves for two versions of the ACT English test, versions F11 and 58E. The relative difficulty is the cumulative area of the space between these two curves. This measure indicates that test F11 is more difficult than 58E. Intuitively, this is because fewer correct answers achieves any given rank on test F11 as compared with 58E.

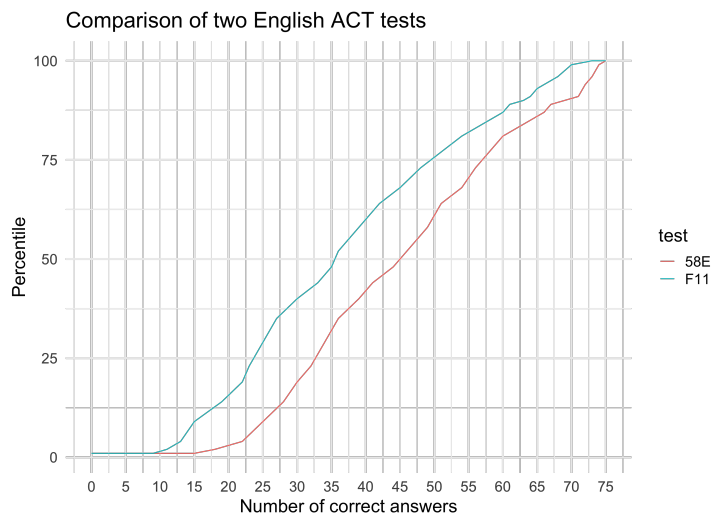


Figure 2: CDF curves for English ACT test versions 58E and F11.

In the prior example, we illustrated a case in which one test version is uniformly harder than the other. It is also worth illustrating an example of a comparison in which the CDF curves cross. In Figure 3, we take English test version 72E and compare it with test version E23. In contrast with the prior example, we cannot say that one of the test versions always requires fewer correct answers for any given rank (i.e., one of the tests is uniformly harder). Rather, for results nearer the extremes (below the 14th percentile and above the 87th), we see that E23 is the marginally harder test; however, this relationship reverses between the 14th and 87th percentiles. Our metric takes the whole range of scores into account when computing relative difficulty. Closer inspection of PrepSharp’s CDF curves can offer insight into various versions of the test’s relative difficulty for test-takers of any ability. Additionally, by changing the limits of integration, one can compare difficulty over a specified range of scores.

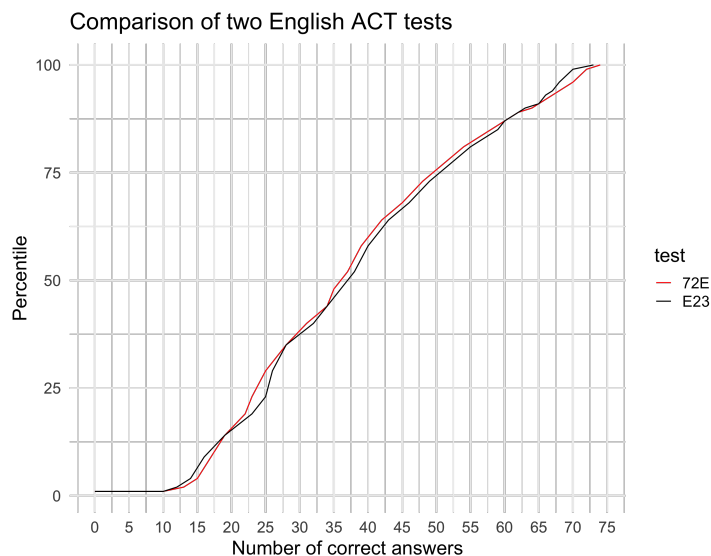


Figure 3: CDF curves for English ACT test versions E23 and 72E.

The mapping between the number of correct answers on a particular section and the associated percentile was obtained from the ACT and differs across test sections [1]. This data is displayed in Appendix A. The percentile ranks associated with the scaled scores change over time and as the ACT releases new ACT Score National Ranks, PrepSharp will update its difficulty ratings.

The method described computes relative difficulty across different versions of a particular test section. The metric used can be conceptualized as the cumulative area between the curves of the CDF curves. Or, alternatively, the difference in the total area under each CDF curve. The scale of this metric depends on the range of the random variable. For each test section this range differs, since the total number of possible correct answers differs across sections. Therefore, this particular metric is only suitable for relative difficulty comparisons across versions of the same section. To illustrate, a difficulty rating of 5 (most difficult) for a particular ACT English section cannot be compared by relative difficulty to a section of Science with the same difficulty rating of 5 (most difficult).

For each test section, we rank all test versions and assign the test a score

based on its rank. We also assign a composite score to each test version which is an average of the score for each section. This simple average is appropriate since the ACT composite score is a simple average of the score on each test section.

3 Results

Using the relative difficulty measure defined in equation 4, we can compare the difficulty across different versions of each ACT test section. These comparisons can be used to categorize the test versions into difficulty levels, as well as assign difficulty ratings. To do so, we compare each test version to the CDF of the median test version. The median test version is the one whose CDF indicates that half of the tests are more difficult and half less difficult. This yields a distribution for each test section, shown in figure 4.

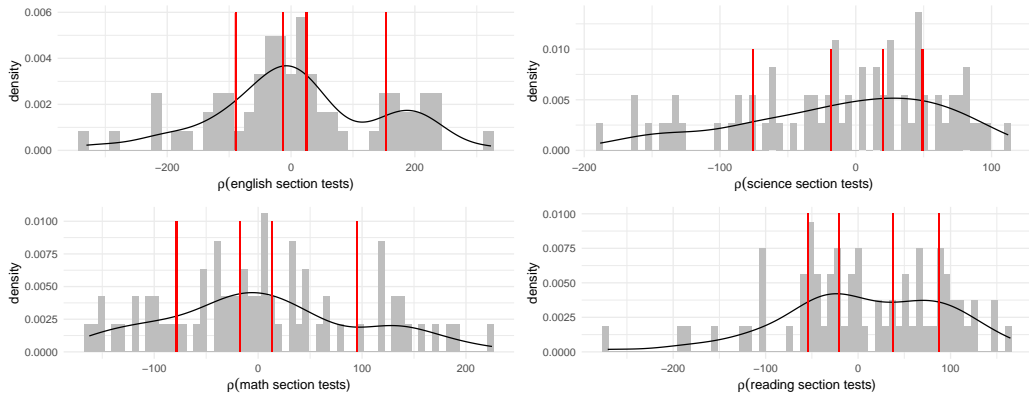


Figure 4: Histograms and density estimates for each test section of $\rho(\text{test version}_i, \text{median test})$. Each histogram includes vertical lines at the quintiles. For instance, the left-most red vertical line in each histogram represents the 20th percentile of that section’s ρ values, where the comparison is made with the median test. Therefore, the spacing of the vertical lines is narrower in ranges with relatively high observation frequency.

Included in figure 4 are vertical lines at the quintiles of each distribution. For each test section, these split the test versions into five difficulty categories,

which are based solely on each test section’s relative quintile rank of the area under its CDF curve. These difficulty categories (least difficult, less difficult, average, more difficult, and most difficult) are available in an interactive format on the PrepSharp website.

Test	English	Science	Math	Reading	Composite
16MC1	5	5	4	5	4.75
E26	5	4	5	5	4.75
F11	5	5	5	4	4.75
73C	5	5	3	5	4.50
A11	4	5	5	4	4.50
E23	5	4	5	4	4.50
16MC2	3	5	4	5	4.25
73G	3	5	4	5	4.25
B04	4	5	5	3	4.25
D06	5	4	5	3	4.25

Table 2: This table shows the hardest 10 tests by composite difficulty. The full table including all test versions is available in Appendix B.

Finally, we use the relative difficulty measure to rank each version within sections and, averaging across these, obtain a composite difficulty score for each test. In table 2, we display the top ten tests as ranked by composite difficulty score. Since some users may desire a finer level of detail (to, for instance, compare across different "more difficult" Math sections), we also provide a complete difficulty ranking of every section and composite on the PrepSharp website.

When applied to the entirety of the scale data, the difficulty rating indices generated by equation 4 do not always align fully with the expectations, experiences, and intuitions of test prep professionals and/or their students. It became apparent that, in the vast majority of cases, the primary reason behind such a discrepancy is that students who prepare for the ACT or SAT, especially under the guidance of a test prep instructor, are considerably more likely to score at or above the 70th percentile. As a result, in such test prep environments, it may be more useful to rely on difficulty ratings generated

by applying the same cdf method detailed above but only to a subset of the data corresponding to the scores earned by the top 30 percent of test takers (e.g., a composite score of 24 and above in the case of the ACT). It was found that the results of such a subset adjustment correlate more strongly with the experiences of test prep professionals and their students, and, consequently, the approach was adopted by PrepSharp for its published difficulty ratings and rankings.

A ACT Score National Ranks

Scaled Score	English	Math	Reading	Science
36	100	100	100	100
35	99	99	98	99
34	96	99	96	98
33	94	98	94	97
32	93	97	91	96
31	91	96	89	95
30	90	95	87	93
29	89	93	84	92
28	87	91	82	90
27	85	89	80	88
26	83	85	77	86
25	81	81	75	83
24	77	77	72	78
23	73	73	68	72
22	68	68	63	65
21	64	65	57	60
20	58	62	52	54
19	52	58	46	48
18	48	53	41	41
17	44	47	36	34
16	40	38	32	28
15	35	25	27	22
14	29	14	23	17
13	23	6	17	12
12	19	2	12	8
11	14	1	7	5
10	9	1	3	3
9	4	1	2	1
8	2	1	1	1
7	1	1	1	1

6	1	1	1	1
5	1	1	1	1
4	1	1	1	1
3	1	1	1	1
2	1	1	1	1
1	1	1	1	1

B Difficulty Scores

Test	English	Science	Math	Reading	Composite
16MC1	5	5	4	5	4.75
E26	5	4	5	5	4.75
F11	5	5	5	4	4.75
73C	5	5	3	5	4.50
A11	4	5	5	4	4.50
E23	5	4	5	4	4.50
16MC2	3	5	4	5	4.25
73G	3	5	4	5	4.25
B04	4	5	5	3	4.25
D06	5	4	5	3	4.25
72E	5	4	2	5	4.00
72G	4	5	4	3	4.00
73E	5	5	2	4	4.00
74F	5	1	5	5	4.00
B02	5	3	5	3	4.00
C03	5	5	3	3	4.00
Z04	5	3	4	4	4.00
Z15	5	4	4	3	4.00
18MC4	4	2	5	4	3.75
74C	4	2	5	4	3.75
55C	4	3	3	4	3.50

A10	3	2	4	5	3.50
B05	5	4	4	1	3.50
3MC	4	3	1	5	3.25
71E	3	5	1	4	3.25
72C	3	3	3	4	3.25
72F	4	2	4	3	3.25
74H	3	1	5	4	3.25
A09	3	2	5	3	3.25
C01	4	3	4	2	3.25
C02	3	1	5	4	3.25
Z08	5	1	5	2	3.25
16MC3	2	5	4	1	3.00
64E	2	2	3	5	3.00
65E	2	2	3	5	3.00
71A	2	4	4	2	3.00
71C	2	4	2	4	3.00
71G	4	4	1	3	3.00
E25	5	1	4	2	3.00
F07	4	1	5	2	3.00
2MC	3	2	1	5	2.75
5MC	2	5	2	2	2.75
60E	2	3	1	5	2.75
63C	2	3	3	3	2.75
65D	4	3	3	1	2.75
67A	2	5	2	2	2.75
69F	2	5	3	1	2.75
70C	1	5	3	2	2.75
D05	2	4	4	1	2.75
4MC	3	4	2	1	2.50
65C	4	2	3	1	2.50
66F	3	4	2	1	2.50
70G	1	3	5	1	2.50

71H	3	2	3	2	2.50
D03	3	2	4	1	2.50
61B	2	1	1	5	2.25
61C	3	1	2	3	2.25
61D	3	2	2	2	2.25
61E	2	2	2	3	2.25
61F	1	1	2	5	2.25
63D	2	1	3	3	2.25
63E	1	3	1	4	2.25
63F	1	3	1	4	2.25
68C	1	4	1	3	2.25
1MC	1	1	2	4	2.00
56B	4	1	1	2	2.00
66C	1	4	1	2	2.00
67C	2	4	1	1	2.00
69A	1	2	2	3	2.00
52C	1	3	2	1	1.75
67F	1	1	3	2	1.75
70A	1	3	1	2	1.75
57B	3	1	1	1	1.50
59F	1	1	2	2	1.50
68G	1	3	1	1	1.50
58E	1	1	2	1	1.25
68A	1	2	1	1	1.25

References

- [1] ACT. *ACT Score National Ranks*. URL: <https://www.act.org/content/dam/act/unsecured/documents/MultipleChoiceStemComposite.pdf>. (accessed: 06.01.2023).