Interpreting Norm-Referenced Scores

Percentile Ranks

Percentile ranks, which range from 1 to 99, are commonly used for reporting test results to students and their parents/guardians. You may interpret a percentile as the percentage of students in a norm group whose scores fall below a given student's scale score. For example, if a student's scale score converts to a national percentile rank (NP) of 71, the student scored higher than approximately 71 percent of the students in the national norm group.

Interpretation

When you discuss test results with students and their parents/guardians, you should make this important distinction: A percentile rank refers to the percentage of students in the norm group who fall below a particular point, not the percentage of items answered correctly. Also note that percentile ranks are not equal interval data. Differences between percentile ranks are larger near the ends of the range than they are in the middle. For example, the difference between percentile ranks of 5 and 10 or between 90 and 95 is much greater than the difference between percentile ranks of 50 and 55. Because the intervals between percentiles are unequal, percentiles are not suitable for statistical work such as computing averages.

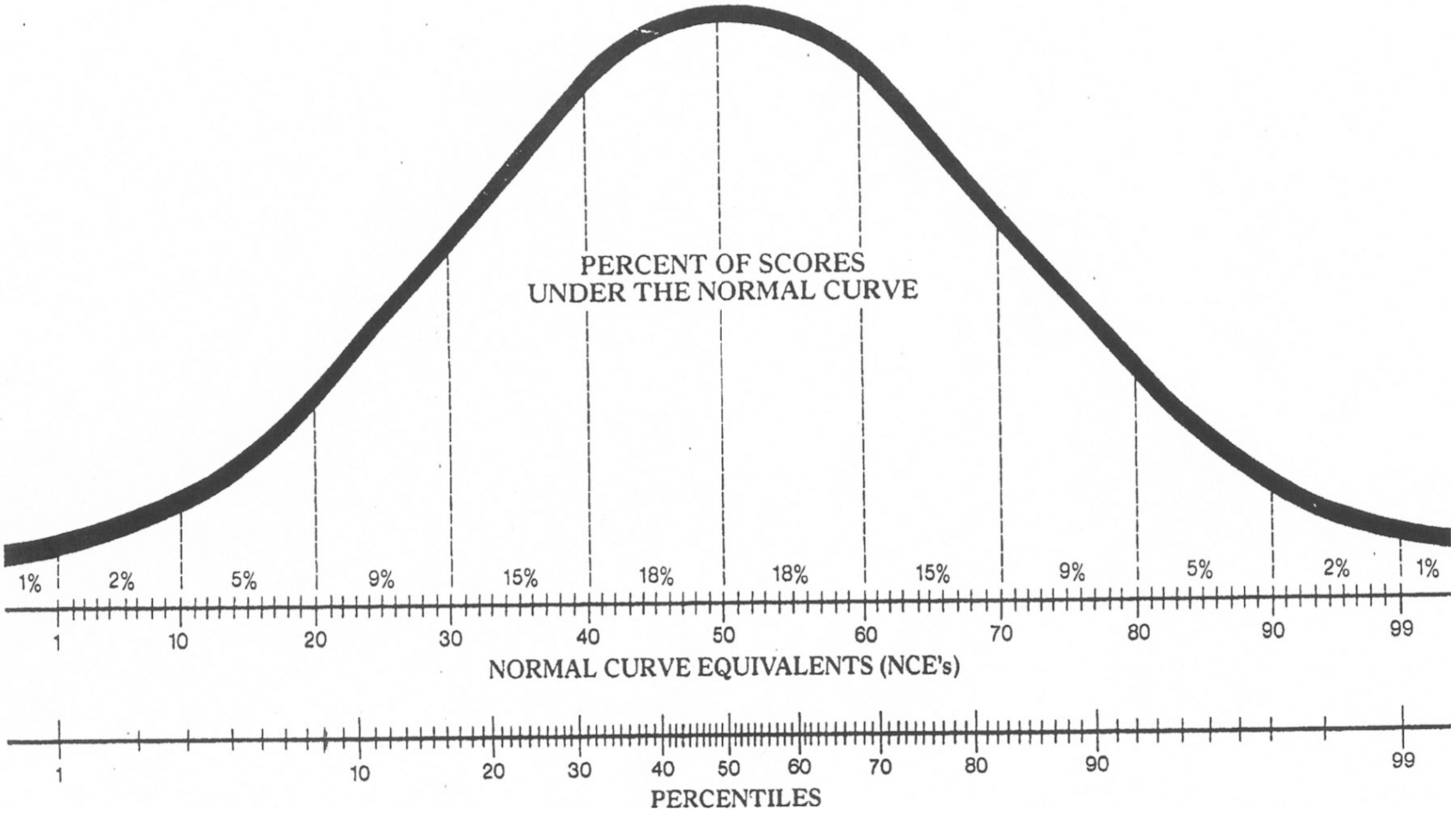
Normal Curve Equivalents

The normal curve-equivalent (NCE) scale, ranging from 1 to 99, coincides with the national percentile scale at 1, 50, and 99. The NCE was developed specifically for use in the evaluation of Title I (formerly Chapter 1) programs. It is used to aggregate "Gain Scores" from different programs, even those using different standardized tests. The NCE, unlike the NP, is an equal-interval scale and can be treated arithmetically. The difference between two successive scores on the scale has the same meaning throughout the scale. This property allows you to make meaningful comparisons among different achievement test batteries and among different tests within the same battery. You can compare NCEs obtained by different groups of students on the same test or test battery by averaging the scores for the groups. For individual students, it is best used when tracking performance over time. Below are some “rules of thumb” for interpreting educationally important differences from pre-test to post-test:

For individual students, use a difference of 11 NCE points from pre to post
For classrooms, use a difference of 7 NCE points
For schools, use a difference of 5 NCE points
For districts, use a difference of 4 NCE points

Comparison with Percentiles

An example will illustrate the difference between interpreting NCEs and interpreting percentile ranks. Consider a student who received an NCE of 53 on a reading test and an NCE of 45 on a mathematics test. One would be correct in saying that the reading score was eight points higher than the mathematics score. However, expressing the comparison as a difference in percentile units, which are not equal interval data, is inappropriate.


PERCENT OF SCORES
UNDER THE NORMAL CURVE
1%
2%
5%
9%
15%
18%
18%
15%
9%
5%
2%
1%
1
10
20
30
40
50
60
70
80
90
99
NORMAL CURVE EQUIVALENTS (NCE's)
1
10
20
30
40
50
60
70
80
90
99
PERCENTILES

National Percentile Rank - NP to Normal Curve Equivalent - NCE

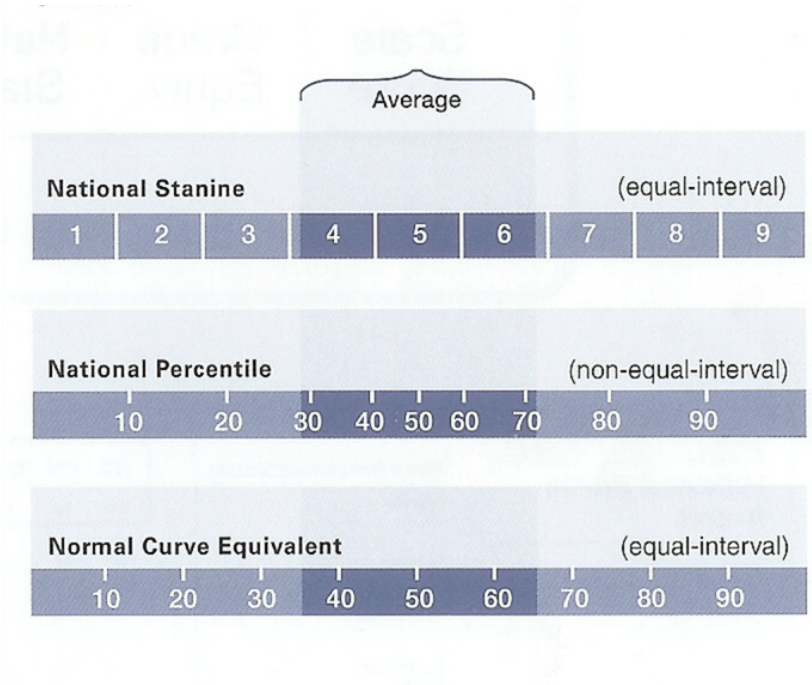
NP	NCE	NP	NCE	NP	NCE	NP	NCE	NP	NCE
1	1.0	21	33.0	41	45.2	61	55.9	81	68.5
2	6.7	22	33.7	42	45.8	62	56.4	82	69.3
3	10.4	23	34.4	43	46.3	63	57.0	83	70.1
4	13.1	24	35.1	44	46.8	64	57.5	84	70.9
5	15.4	25	35.8	45	47.4	65	58.1	85	71.8
6	17.3	26	36.5	46	47.9	66	58.7	86	72.8
7	18.9	27	37.1	47	48.4	67	59.3	87	73.7
8	20.4	28	37.7	48	48.9	68	59.9	88	74.7
9	21.8	29	38.3	49	49.5	69	60.4	89	75.8
10	23.0	30	39.0	50	50.0	70	61.0	90	77.0
11	24.2	31	39.6	51	50.5	71	61.7	91	78.2
12	25.3	32	40.1	52	51.1	72	62.3	92	79.6
13	26.3	33	40.7	53	51.6	73	62.9	93	81.1
14	27.2	34	41.3	54	52.1	74	63.5	94	82.7
15	28.2	35	41.9	55	52.6	75	64.2	95	84.6
16	29.1	36	42.5	56	53.2	76	64.9	96	86.9
17	29.9	37	43.0	57	53.7	77	65.6	97	89.6
18	30.7	38	43.6	58	54.2	78	66.3	98	93.3
19	31.5	39	44.1	59	54.8	79	67.0	99	99.0
20	32.3	40	44.7	60	55.3	80	67.7		

National Stanines

Stanines are standard scores based on a scale of nine equal units that range from a high of 9 to a low of 1. In general, stanines of 1 through 3 are considered below average, 4 through 6 average, and 7 through 9 above average.

Interpretation

Because stanines are single digits, they are not likely to be confused with the percentage of test items answered correctly. Moreover, since stanines are units on an equal-interval scale, they can be readily compared. Their major disadvantage is their lack of precision. For example, a student with a National Stanine score of 5 could have a National Percentile as low as 41 and as high as 59.



Scale Scores

The scale score is the basic score for TerraNova. It is used for deriving various other normative scores to describe test performance. The scale scores are units of a single, equal-interval scale applied across all levels of TerraNova regardless of grade or time of year of testing. These scores are expressed in numbers that range from 0 to 999.

The equal-interval property of scale scores makes them especially appropriate for various statistical purposes. For example, scale scores (unlike National Percentiles) can be added, subtracted, and averaged across test levels. Such computations permit direct comparisons between classes, schools, or entire districts. The year-to-year growth of individual students or groups in content areas can also be tracked using scale scores.

Because the test content areas are "scaled" separately, the scale scores for one area cannot be compared with the scale scores for another. For example, a scale score of 468 on a language test would not have the same meaning as the same scale score on a mathematics test. If you look just at the student's scale scores, you will not gain any information about that student's performance on a particular test area relative to other areas. You can get useful information, however, by comparing the student's performance on a test in scale score units with the average performance for the group on that test or by converting the scale score to a derived score such as a normal curve equivalent.

TERRANOVA 3®

DATA RECOGNITION

DRC

CORPORATION