# Interpreting NCE Scores

John R. Hills
*Florida State University*



In recent years a new score system for standardized tests has entered the scene. The new scores are called Normal Curve Equivalents (NCEs). They were named and introduced by G. Kasten Tallmadge. NCEs are used in reporting results to federal funding agencies. Some schools may use them internally in order to deal with only one kind of score from standardized tests. Several kinds of confusion are possible with NCE scores, and the following quiz and its explanations may help you avoid some of the pitfalls. The correct answers and brief explanations appear on page 31. This "test" may be reproduced without written permission. Please cite *Educational Measurement: Issues and Practice* as the source.

Source: Educational Measurement: Issues and Practice, Fall 1984, pp. 25-26, 31

*Editor's Note. This is the last Hills' Handy Hints, a series which has been very popular. A number of NCME members have indicated that they used one or more of the series with parent and other lay groups interested in learning more about how to interpret test scores accurately. One of the purposes of* EM *is to publish articles and features that are of appeal to persons outside our NCME membership as well as to our own members, and Hills' Handy Hints certainly have done that. The editor is greatly appreciative of the work of John Hills in producing this series. Suggestions for other useful series or special features would be welcomed.*

**Hills' Handy Hints: NCE Scores**

Please circle either Y or N (or both) for each question. Mark a response for each question as there is no penalty for guessing. This is an untimed test.

Y N 1. Miss Yamini has noticed that the highest scoring third grader who was placed in remedial instruction in reading in her school had a reading comprehension NCE score of 36. She thought the cutoff for remediation was the 25th percentile. She protested to Mr. Wommack, the principal, that an error in placement had been made. Was Miss Yamini's protest sound?

Y N 2. Miss Yamini also collected reading data on her remedial students in the fall of one year and again on the same students in the fall of the next year. She converted her data into NCE scores properly, found the means, and discovered her students had zero mean gain. She concluded that her efforts had been in vain because her students had learned nothing. Mr. Wommack comforted her by saying that her students really had improved in reading. He was sure of it. Was he right?

Y N 3. Mr. Quigley has noticed that on the *Widely Used Achievement Test* (WUAT) Sybil's performance on reading has gone up from an NCE score of 10 to an NCE score of 15, and her mathematics performance has gone up from an NCE score of 35 to an NCE score of 40. Mr. Quigley interprets this as an indication that Sybil's scores have increased equal amounts in reading and mathematics. Mr. Wommack says this interpretation is not correct, because it is much easier to improve from 35 to 40 than from 10 to 15. Is Mr. Wommack correct?

Y N 4. Mr. Quigley decided to calculate a class mean NCE score by adding up the individual NCE scores and dividing by their number. Mr. Wommack was horrified. He said you can't average NCE scores. They don't mean the same thing in different parts of the score scale. Should Mr. Wommack be horrified?

Y N 5. Mr. Pridgeon thinks that NCE scores are the best of all the derived scores for tests, and he believes that they will be used universally in the near future. He decides to be in the front of the parade, so he converts all his classroom test scores to NCEs. He transforms them to have a mean of 50 and a standard deviation of 21.06 and thus his scores are NCE scores. Is that true?

Y N 6. Mr. Quigley found that the mean NCE score in arithmetic for his third graders was 44. Out of curiosity he got Mrs. Nye to let him calculate the mean NCE score in arithmetic for her sixth graders. It was only 34. Mr. Quigley decided that Mrs. Nye's sixth graders were not performing as well in arithmetic compared to

John Hills is Professor of Educational Measurement in the Department of Educational Research, Development and Foundations of the College of Education at Florida State University.

their peers as his third graders. When Mr. Pridgeon heard about this, he said Mr. Quigley was wrong. Mr. Pridgeon had been reading *Hills' Handy Hints* about grade-equivalent scores and had discovered that for a group below the mean, grade-equivalent scores get farther below the mean year by year. The same thing probably was occurring with Mr. Quigley's and Mrs. Nye's arithmetic scores. Is Mr. Pridgeon correct?

Y N 7. Mr. Pollitz is having a conference with Eanix's parents. Eanix received an NCE score of 30 on the WUAT arithmetic test but a reading NCE score of 40. Mr. Pollitz decides that Eanix is twice as far below grade level in arithmetic as in reading. Is he correct?

Y N 8. Mr. Pollitz also noticed that Eanix's arithmetic NCE score was 30 on the WUAT arithmetic test but her geometry NCE score was 40 on the St. Louis Geometry Test (SLGT). Mr. Pollitz decided from that difference that Eanix is doing much better in geometry than in arithmetic as she is only half as far below the mean in geometry. Is this conclusion sound?

*Answers on page 31*

**Hills' Handy Hints**
*Continued from page 26*

**Answers**
1.N, 2.Y, 3.N, 4.N, 5.N, 6.N, 7.Y, 8.Y and N

According to Dr. Tallmadge (personal communication, April 27, 1983) NCEs were to provide an equal-interval scale that would have essentially the same meaning for any nationally normed achievement test at any grade level and would be intrinsically meaningful to school teachers, parents, and other nonpsychometricians. NCEs were to be normalized, and they would be assumed to be equal-interval scales. Because they were to be tied to national percentile norms distributions, they would be as similar across tests as the publishers' norming samples. NCEs were put on a 100 point scale so that they would be most meaningful to unsophisticated audiences who are uncomfortable with a scale that has negative scores or that tops out at 80. In a sense they are very like T scores except they have a standard deviation of 21.06 (instead of 10) so that the scale extends from 0 to 100 instead of from 20 to 80.

1. NCE scores and percentile ranks correspond only at 3 points, the first, the 50th, and the 99th percentiles (or NCE scores). Actually, the NCE score of 36 corresponds to a percentile rank of 25 which is widely used as a cutoff for remedial instruction.

2. Mr. Wommack is correct. Properly derived NCE scores show a score's relationship to representative national norms. If Miss Yamini used the appropriate norms for the testing dates of her class, zero mean gain in NCE scores signifies that her group improved just as much as the norms group improved. Maintaining the same NCE score does not mean "no growth"; it means "normal growth."

3. Because NCE scores are based on normal distributions, they do *not* have the property of percentile ranks that a score difference represents different amounts of change at different places in a distribution. Mr. Wommack would be correct for percentile ranks, but not for NCE scores.

4. NCE scores are like the other standard scores (z, T, and stanine). Because they are based on the normal curve (unlike percentile ranks), we feel comfortable in averaging them.

5. NCE scores have a mean of 50 and a standard deviation of 21.06 *for the scores of a representative national sample which have, or are transformed to have, a normal distribution.* Simply converting a distribution of scores to a mean of 50 and standard deviation of 21.06 does not make the raw scores into NCE scores. A representative national sample is essential. Mr. Pridgeon does not have that.

6. Mr. Pridgeon is wrong again. NCE scores do not behave like grade-equivalent scores in this respect. An NCE score of 34 is just as far below grade level in grade 3 as in grade 6 or any other grade.

7. According to the inventor of NCEs, because NCEs are based on nationally representative groups whose scores have been normalized, they have equal intervals. Because they have equal intervals, a score of 30 is twice as far from the mean as a score of 40, so Mr. Pollitz is correct.

8. Mr. Pollitz may or may not be correct this time, and the reason gets to one of the important and perhaps controversial aspects of NCE scores. If both the WUAT and the SLGT were normed on representative samples of the national population of students at Eanix's grade level, and if the national groups' scores are normally distributed, and if they are normally distributed because they come from an equal interval scale, then Mr. Pollitz' conclusion is sound. But, if any one or combination of those assumptions is wrong, then the conclusion may not be sound.

How could one of those assumptions be wrong? First, the tests may not have been normed at the time of year when Eanix was tested. They may not have both been normed at the same time of the year. If either varied by more than a few months from the time of year when Eanix was tested, the different performances in arithmetic and geometry may be due to different growth rates in those subjects for students in Eanix's grade level, and/or different norming dates for the two tests.

Second, if the WUAT and the SLGT were published by different publishers, differences in NCE scores may reflect differences in the sampling procedures used to obtain representative samples of the national population. It is well known that two publishers attempting to get representative samples of the nation's children may get samples that differ widely. (See, e.g., the *User's Manual of the Anchor Test Study of Selected Standardized Reading Achievement Tests,* by Loret, P.G., Seder, A., Bianchini, J.C., & Vale, C.A. Washington, D.C.: U.S. Government Printing Office, 1973, page 5.)

Third, not everyone is convinced that normalizing of score distributions necessarily produces equal-interval scores. If normalizing does not produce equal-interval scores, Mr. Pollitz's decision would be unsound for that reason.

Your choice of Y or N depends on what you believe about how representative different publishers' norms are, what differences might be due to time intervals between norming and Eanix's testing, and whether normalizing score distributions in cases such as this results in equal-interval scales. ■

## 1985 ELECTIONS

The Nominations Committee nominates the following candidates for the 1985 elections.

**For Vice President**
Peter W. Airasian
*Boston College*

Richard Jaeger
*University of North Carolina, Greensboro*

The vice president, at the end of a 1-year term, automatically succeeds to the office of President.

**For Board of Directors**
Sylvia T. Johnson
*Howard University*

Vana H. Meredith
*South Carolina Department of Education*

Dean Nafziger
*San Diego Unified School District*

Richard Valencia
*University of California, Santa Cruz*

Directors serve a 3-year term; two are to be elected this year.

Nominations can also be made by NCME members. The bylaws state: "Other nominations may then be submitted by the written request of at least ten active or emeritus members, provided such nominations are received before December 1." Send nomination petitions to NCME, 1230 Seventeenth St., N.W., Washington, D.C. 20036. Ballots will be mailed to members late December 1984.