

The National Spanish Examinations

Reliability

NSE recognizes that test reliability is defined as the degree to which the test gives consistent results each time it is given.

In other words, reliability answers the following questions:

1. Can I depend on the test to measure the same outcomes consistently?
2. Given all the other variables being the same, will the test produce the same results again?

NSE uses the Kuder-Richardson 21 formula to calculate reliability coefficients:

$$r(\text{reliability}) = \frac{(K)(SD^2) - M(K-M)}{(SD^2)(K-1)}$$

K = the number of items in the test

SD = the standard deviation of the scores

M = the mean of the scores

The scores of reliability are judged against a perfect score of 1.00. The closer the reliability coefficient is to 1.00, the better it is. Most standardized tests usually have a reliability coefficient of .90 or above.¹

Reliability Coefficients for the 2011 National Spanish Examinations

| Level 01 | Vocab | Gram | Ach | Read | List | Prof | Total |
|------------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=4746) | | | | | | | |
| Mean Raw Score | 51.4 | 46.2 | 97.5 | 60.9 | 58.7 | 119.5 | 217.1 |
| Standard Deviation | 18.6 | 17.6 | 33.5 | 17.4 | 20.9 | 34.7 | 62.1 |
| Reliability Coefficient | 0.938 | 0.929 | 0.960 | 0.931 | 0.954 | 0.965 | 0.977 |
| Outside Experience (N=1162) | | | | | | | |
| Mean Raw Score | 62.8 | 53.1 | 115.8 | 66.4 | 66.6 | 133.0 | 248.9 |
| Standard Deviation | 20.6 | 18.5 | 36.2 | 16.1 | 20.8 | 33.6 | 63.3 |
| Reliability Coefficient | 0.955 | 0.936 | 0.968 | 0.923 | 0.958 | 0.965 | 0.979 |
| Bilingual (N=250) | | | | | | | |
| Mean Raw Score | 77.6 | 61.1 | 138.8 | 68.3 | 73.5 | 141.8 | 280.6 |
| Standard Deviation | 22.5 | 20.7 | 40.6 | 19.3 | 22.3 | 38.9 | 73.1 |
| Reliability Coefficient | 0.976 | 0.954 | 0.979 | 0.951 | 0.971 | 0.978 | 0.987 |

¹ John A. Kaufhold, Basic Statistics for Educational Research (New York: iUniverse, Inc., 2007) pp. 43-46.

| Level 1 | Vocab | Gram | Ach | Read | List | Prof | Total |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=32926) | | | | | | | |
| Mean Raw Score | 57.4 | 52.6 | 110.0 | 65.8 | 63.4 | 129.2 | 239.2 |
| Standard Deviation | 20.0 | 19.2 | 36.7 | 18.3 | 22.1 | 37.1 | 67.5 |
| Reliability Coefficient | 0.948 | 0.942 | 0.968 | 0.942 | 0.962 | 0.972 | 0.981 |
| Outside Experience (N=3365) | | | | | | | |
| Mean Raw Score | 70.2 | 61.6 | 131.8 | 71.6 | 74.1 | 145.7 | 277.5 |
| Standard Deviation | 18.9 | 19.1 | 35.5 | 16.7 | 20.4 | 34.1 | 63.0 |
| Reliability Coefficient | 0.951 | 0.945 | 0.969 | 0.936 | 0.963 | 0.971 | 0.981 |
| Bilingual (N=1896) | | | | | | | |
| Mean Raw Score | 83.1 | 66.1 | 149.2 | 71.5 | 75.4 | 146.8 | 296.1 |
| Standard Deviation | 19.1 | 18.6 | 35.0 | 19.8 | 23.8 | 40.8 | 65.2 |
| Reliability Coefficient | 0.971 | 0.945 | 0.974 | 0.958 | 0.977 | 0.981 | 0.984 |

| Level 2 | Vocab | Gram | Ach | Read | List | Prof | Total |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=32395) | | | | | | | |
| Mean Raw Score | 49.7 | 36.9 | 86.5 | 69.0 | 61.6 | 130.6 | 217.2 |
| Standard Deviation | 18.5 | 19.7 | 35.0 | 19.1 | 21.1 | 36.9 | 65.0 |
| Reliability Coefficient | 0.937 | 0.949 | 0.965 | 0.951 | 0.956 | 0.972 | 0.979 |
| Outside Experience (N=2466) | | | | | | | |
| Mean Raw Score | 61.1 | 42.4 | 103.5 | 73.1 | 70.1 | 143.2 | 246.6 |
| Standard Deviation | 20.3 | 20.9 | 37.8 | 19.8 | 22.4 | 39.2 | 69.6 |
| Reliability Coefficient | 0.952 | 0.954 | 0.970 | 0.960 | 0.968 | 0.978 | 0.983 |
| Bilingual (N=2633) | | | | | | | |
| Mean Raw Score | 79.8 | 50.7 | 130.5 | 75.0 | 75.1 | 150.1 | 280.6 |
| Standard Deviation | 20.7 | 18.1 | 35.8 | 22.6 | 24.5 | 44.3 | 68.0 |
| Reliability Coefficient | 0.972 | 0.933 | 0.970 | 0.973 | 0.979 | 0.986 | 0.984 |

| Level 3 | Vocab | Gram | Ach | Read | List | Prof | Total |
|-----------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=26326) | | | | | | | |
| Mean Raw Score | 51.8 | 48.0 | 99.8 | 75.0 | 48.1 | 123.1 | 222.9 |
| Standard Deviation | 16.6 | 19.6 | 32.8 | 20.4 | 17.7 | 34.3 | 60.4 |
| Reliability Coefficient | 0.918 | 0.945 | 0.958 | 0.965 | 0.930 | 0.965 | 0.975 |
| Outside Experience (N=2342) | | | | | | | |
| Mean Raw Score | 61.1 | 53.9 | 115.0 | 78.1 | 54.9 | 133.0 | 248.0 |
| Standard Deviation | 19.6 | 21.1 | 37.3 | 20.9 | 19.6 | 36.9 | 67.4 |
| Reliability Coefficient | 0.948 | 0.954 | 0.970 | 0.970 | 0.945 | 0.972 | 0.982 |
| Bilingual (N=2352) | | | | | | | |
| Mean Raw Score | 82.0 | 67.3 | 149.2 | 78.2 | 61.7 | 139.9 | 289.1 |
| Standard Deviation | 18.4 | 18.5 | 34.5 | 21.9 | 19.7 | 38.1 | 60.8 |
| Reliability Coefficient | 0.966 | 0.945 | 0.973 | 0.974 | 0.948 | 0.976 | 0.981 |

| Level 4 | Vocab | Gram | Ach | Read | List | Prof | Total |
|------------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=15306) | | | | | | | |
| Mean Raw Score | 50.5 | 46.1 | 96.6 | 71.3 | 48.4 | 119.7 | 216.2 |
| Standard Deviation | 16.9 | 20.4 | 33.0 | 19.9 | 19.3 | 35.2 | 60.6 |
| Reliability Coefficient | 0.922 | 0.950 | 0.959 | 0.958 | 0.943 | 0.966 | 0.975 |
| Outside Experience (N=1340) | | | | | | | |
| Mean Raw Score | 58.0 | 47.8 | 105.8 | 73.0 | 54.5 | 127.5 | 233.4 |
| Standard Deviation | 19.5 | 20.8 | 35.9 | 20.4 | 21.8 | 38.2 | 65.4 |
| Reliability Coefficient | 0.945 | 0.952 | 0.966 | 0.962 | 0.957 | 0.973 | 0.980 |
| Bilingual (N=2192) | | | | | | | |
| Mean Raw Score | 76.5 | 62.3 | 138.8 | 68.0 | 56.9 | 124.9 | 263.6 |
| Standard Deviation | 16.8 | 19.0 | 32.6 | 25.5 | 25.6 | 47.2 | 66.8 |
| Reliability Coefficient | 0.946 | 0.944 | 0.965 | 0.976 | 0.972 | 0.984 | 0.982 |

| Level 5 | Vocab | Gram | Ach | Read | List | Prof | Total |
|-----------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=5782) | | | | | | | |
| Mean Raw Score | 36.9 | 51.6 | 88.5 | 67.5 | 58.4 | 125.9 | 214.4 |
| Standard Deviation | 14.8 | 21.7 | 31.6 | 18.0 | 20.8 | 34.7 | 58.1 |
| Reliability Coefficient | 0.902 | 0.956 | 0.955 | 0.942 | 0.954 | 0.966 | 0.973 |
| Outside Experience (N=760) | | | | | | | |
| Mean Raw Score | 43.0 | 53.9 | 96.9 | 68.3 | 65.8 | 134.1 | 231.0 |
| Standard Deviation | 17.9 | 22.7 | 36.1 | 17.5 | 21.2 | 34.6 | 62.0 |
| Reliability Coefficient | 0.933 | 0.962 | 0.967 | 0.938 | 0.960 | 0.968 | 0.977 |
| Bilingual (N=1149) | | | | | | | |
| Mean Raw Score | 65.8 | 69.7 | 135.5 | 65.0 | 69.5 | 134.5 | 270.0 |
| Standard Deviation | 20.1 | 18.0 | 34.2 | 20.2 | 24.4 | 41.0 | 61.2 |
| Reliability Coefficient | 0.954 | 0.944 | 0.968 | 0.954 | 0.974 | 0.979 | 0.979 |

| Level 6 | Vocab | Gram | Ach | Read | List | Prof | Total |
|-----------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Regular (N=489) | | | | | | | |
| Mean Raw Score | 43.3 | 59.9 | 103.2 | 72.5 | 66.9 | 139.4 | 242.6 |
| Standard Deviation | 15.7 | 20.4 | 30.6 | 17.1 | 21.0 | 34.4 | 56.3 |
| Reliability Coefficient | 0.909 | 0.952 | 0.952 | 0.942 | 0.959 | 0.969 | 0.972 |
| Outside Experience (N=171) | | | | | | | |
| Mean Raw Score | 43.6 | 49.8 | 93.4 | 71.0 | 67.6 | 138.6 | 232.0 |
| Standard Deviation | 16.4 | 23.1 | 34.3 | 19.5 | 22.5 | 38.0 | 60.7 |
| Reliability Coefficient | 0.918 | 0.963 | 0.963 | 0.955 | 0.966 | 0.975 | 0.976 |
| Bilingual (N=325) | | | | | | | |
| Mean Raw Score | 70.7 | 71.7 | 142.4 | 68.1 | 71.6 | 139.6 | 282.1 |
| Standard Deviation | 20.2 | 18.2 | 35.2 | 20.2 | 24.7 | 40.9 | 62.1 |
| Reliability Coefficient | 0.959 | 0.948 | 0.972 | 0.956 | 0.976 | 0.980 | 0.981 |