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:

$$r \text{ (reliability)} = \overline{\hspace{3cm}}$$

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 2012 National Spanish Examinations

Level 01	Vocab	Gram	Ach	Read	List	Prof	Total
Regular (N= 5377)							
Mean Raw Score	64.2	46.3	110.4	54.2	49.7	104	214.4
Standard Deviation	20.1	18.2	35.8	20.6	18.5	35.4	66.5
Reliability Coefficient	.953	.934	.966	.951	.936	.965	.980
Outside Experience (N= 1371)							
Mean Raw Score	73.3	52.6	125.9	61.6	57.7	119.3	245.2
Standard Deviation	20	18.5	36	20.6	19.9	37.3	69.1
Reliability Coefficient	.961	.937	.969	.954	.948	.970	.983
Bilingual (N= 304)							
Mean Raw Score	82.6	57.2	139.8	68.3	69.3	137.6	277.3
Standard Deviation	20.4	17.4	34.9	19.9	20.1	36.5	66.2
Reliability Coefficient	.975	.928	.970	.955	.957	.973	.983

[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= 33759)							
Mean Raw Score	70.4	53.2	123.5	61.6	53.9	115.5	239
Standard Deviation	20.0	19.8	37.3	21.6	20.1	38.3	70.6
Reliability Coefficient	.957	.946	.971	.959	.948	.972	.983
Outside Experience (N=3469)							
Mean Raw Score	81.0	60.6	141.5	70.7	65.4	136.1	277.7
Standard Deviation	17.6	19.7	34.8	20.4	21.3	38.6	68.4
Reliability Coefficient	.960	.948	.971	.960	.960	.976	.984
Bilingual (N= 2012)							
Mean Raw Score	88.3	65.5	153.8	74.2	72.6	146.7	300.5
Standard Deviation	16.7	19.4	33.5	21.9	24.2	43.3	69.8
Reliability Coefficient	.973	.950	.973	.970	.976	.984	.987

Level 2	Vocab	Gram	Ach	Read	List	Prof	Total
Regular (N= 33,322)							
Mean Raw Score	47.0	41.5	88.5	64.5	53.3	117.9	206.3
Standard Deviation	16.3	20.2	33.0	22.4	19.6	38.3	65.1
Reliability Coefficient	.915	.950	.959	.964	.944	.972	.979
Outside Experience (N= 2763)							
Mean Raw Score	58.6	48.9	107.6	70.3	63.1	133.4	240.9
Standard Deviation	19.8	20.7	36.7	22.4	21.6	40.8	71.1
Reliability Coefficient	.948	.951	.968	.968	.960	.978	.984
Bilingual (N= 2581)							
Mean Raw Score	78.7	54.7	133.3	71.2	68.8	140	273.3
Standard Deviation	20.7	17.8	35.2	22.3	21.5	40.6	67.0
Reliability Coefficient	.971	.931	.969	.968	.963	.979	.983

Level 3	Vocab	Gram	Ach	Read	List	Prof	Total
Regular (N= 26,736)							
Mean Raw Score	51.3	45.6	96.9	67.7	53.4	121.1	218
Standard Deviation	17.4	17.8	31.6	20.8	18.7	35.8	61.1
Reliability Coefficient	.927	.931	.955	.959	.938	.968	.976
Outside Experience (N= 2202)							
Mean Raw Score	62.6	50.2	112.8	70.3	60.2	130.6	243.3
Standard Deviation	20.7	19.2	36.3	20.9	19.7	37.4	65.3
Reliability Coefficient	.955	.942	.968	.962	.948	.973	.980
Bilingual (N= 2198)							
Mean Raw Score	84.6	62.9	147.4	71.7	63.8	135.5	282.9
Standard Deviation	18.5	18.6	34.4	20.9	18.8	35.9	59.1
Reliability Coefficient	.972	.942	.972	.963	.944	.971	.979

Level 4	Vocab	Gram	Ach	Read	List	Prof	Total
Regular (N= 15,607)							
Mean Raw Score	51.0	43.7	94.7	66.9	53.3	120.3	215.0
Standard Deviation	16.5	20.7	32.8	19.4	18.6	34.0	59.7
Reliability Coefficient	.918	.952	.958	.950	.938	.963	.975
Outside Experience (N= 1473)							
Mean Raw Score	58.3	47.3	105.6	69.9	61.1	131.0	236.7
Standard Deviation	18.7	21.6	36.0	19.2	20.8	36.1	65.0
Reliability Coefficient	.940	.956	.966	.953	.955	.970	.980
Bilingual (N= 2065)							
Mean Raw Score	78.8	57.4	136.2	68.7	66.8	135.5	271.8
Standard Deviation	16.6	17.9	31.0	20.4	21.5	37.9	58.0
Reliability Coefficient	.949	.933	.960	.958	.962	.974	.977

Level 5	Vocab	Gram	Ach	Read	List	Prof	Total
Regular (N= 5449)							
Mean Raw Score	52.0	45.7	97.6	68.4	49.0	117.4	215.0
Standard Deviation	16.4	19.4	31.7	19.7	17.7	33.5	57.8
Reliability Coefficient	.916	.944	.955	.954	.930	.962	.973
Outside Experience (N= 640)							
Mean Raw Score	61.9	51.1	113.1	70.4	58.5	128.9	242.0
Standard Deviation	17.7	20.9	34.7	19.7	19.2	35.1	59.9
Reliability Coefficient	.934	.952	.964	.956	.944	.968	.976
Bilingual (N= 1312)							
Mean Raw Score	80.8	63.3	144.1	69.9	64.1	134.1	278.2
Standard Deviation	16.5	17.3	30.6	19.4	20.1	35.2	55.4
Reliability Coefficient	.953	.932	.962	.954	.952	.969	.975

Level 6	Vocab	Gram	Ach	Read	List	Prof	Total
Regular (N=553)							
Mean Raw Score	54.6	47.5	102.1	72.9	54.7	127.5	229.6
Standard Deviation	17.3	20.7	34.3	18.0	18.0	32.3	59.7
Reliability Coefficient	.926	.951	.962	.948	.933	.960	.975
Outside Experience (N= 125)							
Mean Raw Score	63.7	49.5	113.2	75.2	62.0	137.2	250.4
Standard Deviation	17.8	19.6	34.1	16.1	18.7	29.9	56.3
Reliability Coefficient	.936	.944	.962	.938	.942	.957	.973
Bilingual (N= 323)							
Mean Raw Score	84.2	65.0	149.2	67.0	63.5	130.5	279.7
Standard Deviation	13.7	15.4	25.9	23.1	23.8	43.0	58.1
Reliability Coefficient	.939	.913	.948	.968	.969	.980	.977