

School and College Ability Test (SCAT) Validation in Spain: Overview of the Process and Some Results

JAVIER TOURÓN
UNIVERSITY OF NAVARRA
PAMPLONA, SPAIN

In this study, we shall briefly cover some of the relevant aspects of the process of the School College Ability Test (SCAT) validation in Navarra, Spain, paying special attention to certain methodological questions of interest. The paper will begin by describing the tests to be validated and the methodology that has been followed during the process. Having established these points, we will present the results obtained, with the goal of establishing the norms of these tests to be used in Spain for the detection of verbally and mathematically talented pupils, applying the model developed for the Study of Mathematically Precocious Youth (SMPY) in 1971. Since 1979, the Center for Talented Youth (CTY) at Johns Hopkins University has continued using this model in Talent Searches, and other universities have adopted it as well, with history and results published in numerous studies (Benbow, 1991; Benbow & Lubinski, 1997; Cohn, 1991; Stanley, 1991; Stanley, George, & Solano, 1977; Stanley, Keating, & Fox, 1974; Van Tassel Baska, 1996).

SCAT Description

The School and College Ability Tests (SCAT) Series III represent a revision of the initial SCAT Series II, originally normalized and standardized in 1966 and reviewed in 1970. These tests were further developed in 1980 by the Educational Testing Service and now belong to CTY.

The SCAT measures verbal and quantitative abilities of students in grades 3-12. The tests are useful in comparing students or classes, comparing performance on the verbal and quantitative sections, estimating growth in abilities over time, and predicting success in related achievement areas. The SCAT measures the accumulation of learning rather than achievement (CTY, 1996, 1999; ETS, 1972).

The tests exist in two forms (X and Y) with three levels of difficulty: Elementary (third to fifth grades), Intermediate (sixth grade of primary education and first and second grades of secondary education) and Advanced (third and fourth grades of secondary education, and first and second grades prior to University entrance, equivalent to grades eleven and twelve in the American educational system).

SCAT tests have two separately timed parts, each consisting of 50 four-option items. Part I, verbal, uses verbal analogy items to measure the student's understanding of words. Studies indicate that verbal analogies are among the four best predictors of academic success (ETS, 1972). Part II is quantitative and tests a student's understanding of fundamental number operations through quantitative comparison items. These items require the student to compare the magnitude of two mathematical quantities and make decisions concerning them.

The questions place minimal emphasis on reading and require more resourcefulness and insight than traditional computational items.

The validation of any measurement instrument is a very complex process that implies more than a mere translation of the items. A careful review of the test content and its adaptation to the new culture is needed. For this reason, before standardizing the test, two pilot studies were conducted, involving the administration of the tests to more than 3500 students, to ensure the psychometric properties of all the forms and levels of the tests (for details of these studies see Tourón, 2000a, and Tourón, 2000b).

Sampling Procedure

Sampling is a delicate process since the goodness (representativeness and size) of the sample will greatly influence the fitness of the norms that are obtained, naturally given that the measurement instrument functions well. In order to organize the sample, it is essential to determine the characteristics of the population that is of interest. In our case we considered two strata: school type (public/private) and grade level (third to twelfth).

We determined a sample size of twice as many classrooms as would be necessary to apply a simple form of the test. The determination of the size of the sample was carried out by dividing the theoretical distribution of the possible scores into deciles, in a way that we can determine the error that would be incorporated into the estimate of the mean value for each of them. Thus, the size of the sample would be estimated from the number of subjects that should comprise each decile at each grade level to be considered. It was also clear that this size was dependent on various parameters: the variance of the population in each decile, the number of subjects of the total population in this decile, the maximum admissible sampling error, and the confidence levels adopted.

In practice, the subjects were selected from a random sample by clusters, in order to respect the natural grouping of the pupils. As such, we considered each group of the class as a cluster.

Sampling error calculations led us to a sample size of 400 students per grade level, which meant testing about 800 pupils per academic level (both forms X and Y), equivalent to a total sample of approximately 8,000 pupils: 4,560 from public centers and 3,366 from private centers, distributed among 221 classrooms (clusters) in private centers and 130 (clusters) in public centers.

The Measurement Properties of the SCAT

The analyses are based on the classical theoretical approach. Moreover, the tests have been calibrated following an approach based on Item Response Theory (IRT).

The indices of difficulty (p) and discrimination, biserial r (r_6), have been calculated for each of the items, as has the pattern of responses to each item, with the aim of determining the degree of attraction (endorsement proportion) of each of the alternatives for the groups in the highest and lowest 27%. In this way, the parameters of the items were evaluated along with the content analysis that was performed. This latter analysis proved to be particularly important during the pilot studies that were undertaken.

Reliability and validity were the two global indicators of the tests that were taken into account. The first was estimated by using Cronbach's coefficient alpha, and content, concurrent, and predictive validity were analyzed.

The results that we present pertain exclusively to the X form of the test.

Mean values of SCAT levels

Table I delineates the mean values of the diverse sections of the SCAT (X form) and the comparison with the results of the American versions offered in the manual.

Table 1

Mean Values for the Original and Spanish Versions of the SCAT

Level	Grade	Original			Spanish		
		Verbal	Quant.	Total	Verbal	Quant.	Total
Elementary	3º	23.6	20.9	44.4	20.2	16.8	37.0
	4º	29.0	26.1	54.5	26.1	23.3	49.3
	5º	31.5	29.7	61.4	32.9	32.9	65.8
Intermediate	6º	23.9	23.0	47.1	26.1	25.1	51.2
	1º ESO	26.1	26.0	52.6	29.3	29.7	59.0
	2º ESO	28.7	28.2	56.8	32.2	32.0	64.2
Advanced	3º ESO	25.8	24.9	50.6	26.7	27.9	54.5
	4º ESO	27.9	26.9	54.8	29.7	31.8	61.5
	1º Bach.	30.1	28.6	58.9	30.9	32.8	63.7
	2º Bach.	32.3	30.2	62.0	32.7	34.8	67.5

At the elementary level, in all the sections, the original values are higher than those in the Spanish version in both the third and fourth grade of primary education. In the fifth grade, however, the values of the Spanish version are slightly higher than those of the original sample.

At the intermediate and advanced levels, the results of the Spanish version show mean values a little higher in all the ages and sections, except for the verbal scores obtained in first and second grades of higher secondary education, where the differences are less than one point. The progression of the mean values in the Spanish version is similar to the original, and the deviation of the mean values does not appear to be too important, although one should not expect too great a similarity when the contexts are clearly different.

Reliability (Internal Consistency, Cronbach S' a)

The last technical SCAT manual (CTY, 1999) offered values for alpha between 0.92 and 0.95 for the total score, between 0.86 and 0.92 for the verbal section, and between 0.87 and 0.93 for the quantitative section. Table 2 shows values for the Spanish version within these ranges, although closer to their lower extremes. We can, as such, consider that the reliability of the Spanish version of SCAT, expressed as internal consistency, is highly acceptable for the practical purpose of measurement. The measurement errors associated with these values vary less than three points in the verbal and quantitative sections and close to four points for the totals.

Difficulty (p) and discrimination (rb)

In Tables 3 and 4 we have presented the indices of difficulty and discrimination for the different sections of the SCAT and for the total scores. With respect to difficulty, it is easy to see how this decreases with the advancement of the grade levels, as is logical, given the increase in the age of the students. The mean difficulty (percent correct) for the elementary level varies between 0.40 and 0.66 for the verbal section and between 0.34 and 0.66 for the mathematics section. At the intermediate level, the range of difficulty for the verbal section was between 0.52 and 0.64, and for the quantitative, between 0.50 and 0.64. The advanced level presented difficulties of between 0.53 and 0.65 for the verbal section and between 0.56 and 0.70 for the quantitative. It is clear, then, that the higher levels of the tests turn out to be easier for older students but will be of use as "out-of-level" tests for younger students.

The mean discrimination indices for the sections and levels of the X form are expressed in terms of coefficients of biserial correlation (rb) in Table 4. The elementary level presents scores between 0.53 and 0.56 for the verbal section and between 0.51 and 0.56 for the quantitative part, which can be considered very satisfactory.

The same can be said for the discrimination indices at the intermediate level. They lie between 0.49 and 0.52 for the verbal section and between 0.46 and 0.52 for the quantitative section. The values for the advanced level are somewhat less, especially in the verbal section (0.41 to 0.43), compared to those for the quantitative section (0.49 to 0.50).

The data offered by the manual for the original test for the fourth to sixth grades sets the levels between 0.45 and 0.54, somewhat less than those obtained by us. For the higher levels, the scores represented in the original test lie between 0.44 and 0.56. As such, we can see that the adapted version has a discriminative capacity similar to the original test, and at some levels higher, although in the verbal section of the advanced level it is lower.

Table 2
Internal Consistency (Cronbach's alpha) of the SCAT Form X

SCAT level	Grade	Verbal X	Quant. X	Total
Elementary	3º P	0.88	0.88	0.93
	4º P	0.89	0.88	0.93
	5º P	0.89	0.88	0.93
Intermediate	6º P	0.87	0.85	0.91
	1º ESO	0.86	0.85	0.91
	2º ESO	0.86	0.88	0.92
Advanced	3º ESO	0.82	0.88	0.91
	4º ESO	0.78	0.87	0.90
	1º Bch.	0.81	0.86	0.89
	2º Bch.	0.80	0.86	0.89

Table 3
Average Difficulty (p) of the SCAT (Form X) Verbal and Quantitative Sections

SCAT level	Grade	Verbal X	Quant. X
Elementary	3º	0.53	0.56
	4º	0.53	0.52
	5º	0.56	0.51
Intermediate	6º	0.49	0.46
	1º ESO	0.51	0.46
	2º ESO	0.52	0.52
Advanced	3º ESO	0.43	0.50
	4º ESO	0.41	0.50
	1º Bch.	0.44	0.49
	2º Bch.	0.41	0.50

Table 4
Average Discrimination (r_b) of the SCAT (form X) Verbal and Quantitative Sections

SCAT level	Grade	Verbal X	Quant. X
Elementary	3º	0.40	0.34
	4º	0.52	0.46
	5º	0.66	0.66
Intermediate	6º	0.52	0.50
	1º ESO	0.59	0.59
	2º ESO	0.64	0.64
Advanced	3º ESO	0.53	0.56
	4º ESO	0.59	0.64
	1º Bch.	0.62	0.66
	2º Bch.	0.65	0.70

Conclusion

The Spanish version of the SCAT (form X), as validated in Navarra, Spain, has very adequate psychometric properties. Accuracy of measurement, difficulty, and discrimination are within the levels commonly accepted for these types of measurement instruments.

Having the norms for ten grades and three different levels of difficulty will provide us with the possibility of performing "in level" assessment of students' verbal and quantitative abilities but also of using the tests to perform talent searches using the tests "out of level," avoiding any ceiling effect.

Identification makes sense when it is linked to proper educational intervention. It is our intention to offer programs for identified talented students who need a more challenging education and who will benefit enormously with differentiated educational opportunities, combining enrichment and acceleration in an appropriate balance.

References

- Benbow, C. P. (1991). Mathematically talented children: Can acceleration meet their needs? In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 154-165). Boston: Allyn and Bacon.
- Benbow, C. P., & Lubinski, D. (1997). Intellectually talented children: How can we best meet their needs? In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 155-169). Boston: Allyn and Bacon.

Cohn, S. J. (1991). Talent searches. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 166-177). Boston: Allyn and Bacon.

Center for Talented Youth. (1996). *School and College Ability Tests Manual: School Version*. Baltimore: The Johns Hopkins University.

Center for Talented Youth. (1999). *School and College Ability Tests Manual: Technical Version*. Baltimore: The Johns Hopkins University.

Educational Testing Service (1972). *School and College Ability Tests. Series HL Manual and Technical Report*. Princeton, NJ: Educational Testing Service.

Stanley, J. C. (1974). Intellectual precocity. In J. C. Stanley, D. P. Keating, & L.

H. Fox (Eds.), *Mathematical talent: Discovery, description, and development* (pp. 1-22). Baltimore: Johns Hopkins University Press.

Stanley, J. C. (1991). SMPY's identification, motivation, and educational facilitation model In R. E. Ciasen (Ed.), *Educating able learners: Book of readings*. Madison, WI:Madison Education Extension Programs University of Wisconsin-Madison.

Stanley, J. C., George, W. C., & Solano, C. H. (1977). *The gifted and the creative: A fifty-year perspective*. Baltimore: Johns Hopkins University Press.

Stanley, J. C., Keating, D. P., & Fox, L. H. (1974). *Mathematical talent: Discovery,description, and development*. Baltimore: Johns Hopkins University Press.

Stanley, J. C., & Stanley, B. S. K. (1986). High-school biology, chemistry, or physics learned well in three weeks. *Journal of Research in Science Teaching*, 23, 237-250.

Tobin, D., & Fox, L. H. (1980). Career interests and career education: A key to change. In L. H. Fox, L. Brody, & D. Tobin (Eds.), *Women and the mathematical mystique* (pp. 171-191). Baltimore: The Johns Hopkins University Press.

Tourón, J. (2000a). The investigación sobre alumnos de alta capacidad en Navarra. Resultados and perspectivas de futuro. [Research about high ability students in Navarra. Results and perspectives]. Faísca, 7, 3-22.

Tourón, J. (2000b, August). Expanding the Talent Search concept in Spain: The validation of the SCAT (School and College Ability Test). A comparative analysis of two pilot studies. Paper presented at the r ECHA Conference: Talent for the New Millennium, Debrecen, Hungary.

VanTassel-Baska, J. (1996). Contributions of the Talent Search concept in gifted education. In C. P. Benbow, & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 236-245). Baltimore: Johns Hopkins University Press.