The Identification of High Ability Students: results of a detection process in Navarra (Spain)

JAVIER TOURÓN, CHARO REPÁRAZ & FELISA PERALTA

This article focuses on the development and analysis of a two-stage early detection process (screening and diagnostic) for high ability students carried out in the region of Navarra (Spain) on a random sample of 1,274 elementary school students. Spanish versions of the Raven Progressive Matrices (SPM), the Renzulli Scales for Rating Behavioural Charac-teristics of Superior Students (SRBCSS), and participants' academic achievement were the main variables in the initial screening phase. The Spanish edition of WISC-R, as well as the Young Children's Academic Intrinsic Motivation Inventory (Gottfried, 1986), and the Self-concept Questionnaire (Marsh, 1988), were the main variables in the diagnostic phase. The findings of the detection process show that the WISC-R results were incompat-ible with those expected in accordance with the normal distribution of IQs. Results may suggest a clear problem with scoring rules, the Spanish norms for the test, or both. These possibilities are discussed in light of the results obtained. The paramount importance of validating and developing adequate instruments for the identification of the gifted is emphasized.

Introduction

One of the most important aspects of attracting the necessary attention to the educational needs of high ability students is, undoubtedly, the identification process. We need at our disposal not only precise instruments (reliable, valid, appropriately standardized, and so on) but also procedures, whose effectiveness and efficiency are experimentally proven, since effective educational assistance for students in need— ensuring that no student is incorrectly excluded—depends on the proper functioning of the detection processes. This is precisely the main objective of the study presented in this article: it is an analysis of the accuracy of an identification process involving elementary school children in the province of Navarra in Northern Spain. Identifying high ability students is meaningful (a) when regular schooling cannot meet the demands of these students' personal development, and (b) when there are alternative programmes better suited for their specific needs. But what happens when there are no specific, established programmes? And why should programmes be established when students have not been identified? Also, what characteristics must one look for in potential candidates?

These three questions are closely related, and thus the answers interact with each other. Indeed, identification must focus not only on a specific concept of giftedness, but also on the characteristics of the programme for which individuals are identified for participation. In turn, the results of individuals assigned to an available pro-gramme will provide valuable information for a potential re-adaptation of the identification process. As Hagen (1980) points out, identification criteria should be operationalized first by determining which concept—or which dimensions—of gift-edness, will be the focus of attention. Then the indicators suitable for the chosen dimensions, based on diverse sources of information and on the instruments used to measure and gather appropriate data, should be selected. Also, prior to gathering the data, according to Hagen, one must decide how the information will be used.

There is agreement on a series of issues in most literature on identification (e.g. Hagen 1980; Tannenbaum, 1983; Feldhusen & Baska, 1989; Clark, 1992; Feld-husen & Jarwan, 1993). However, there are also many discrepancies between views and some controversies are apparent, as summarized by Hany (1993). It is, for example, clearly stated by most authors that there must exist a relationship between identification processes and the provision of educational opportunities (i.e. pro-grammes). The implication is that identification must be guided by programme requirements. In some areas, however, where there exist neither regular identifi-cation procedures nor specific programmes for the gifted—and the setting in which the current study was pursued is a case in point—a suitable point of departure must first be found. This is why it was decided to carry out a general identification process, which could serve as a catalyst in prompting a local programme develop-ment.

The question whether specific and/or general indicators should be focused in an identification process is one of the controversies found in the literature. However, involving young students (from first to third-grade elementary school), where no specific programmes exist, a general identification process, based primarily on intellectual measures can be justified (cf. Feldhusen & Baska, 1989) and also used in an effort to activate the hitherto passive Spanish educational system for the benefit of the gifted and talented. The legal and curricular framework on how to cater for diversity in the Spanish school system has been outlined elsewhere (cf. Touron, Iriarte, Repáraz & Peralta, 1998a).

This article focuses on an analysis of the effectiveness and efficiency of the chosen identification process. Any identification process implies making decisions concern-ing which students have the sought-after characteristics and which do not. To operationalize such an identification process the following assumptions must first be made, as outlined by Hany (1993):

• It is feasible to establish a set of criteria that will permit a division of students into two groups, namely those who fulfill the criteria and those who do not.

• One must also accept that there is a way to establish who is "truly" gifted (according to predetermined criteria) and who is not. This means that it is possible to measure one or more variables that are presumably closely linked to the condition of being intellectually gifted (again according to a predetermined criterion).

		'True' value of θ (diagnostic)	
		θ_0	θ_1
Screening decision	a_0	A	B
	a_1	C	D

Figure 1. Possible decisions in the classification process.

Obviously, being gifted or not is a means of classification, since giftedness, in fact, is an emerging process more than a final result (cf. Treffinger & Feldhusen, 1996). For the decision-making process, however, which necessarily implies classification, the expressions "gifted" or "not gifted" must be used, but without discarding the specific context in which it has been applied.

All classification procedures involve making right or wrong decisions. In fact, there is the possibility of making two correct decisions and two incorrect (see Figure 1). Note that the characteristic we are looking for in the students (i.e. being intellectually gifted) is denoted as *6,* which may take on only one of two possible values; that is ϑ_0 or ϑ_1. This characteristic is unknown and must be estimated through the chosen indicators (i.e. the measured variables); otherwise there will be no errors in the classification procedure. There are four possible decisions outlined in Figure 1: $e_{ij} = (a_i; \vartheta_j)$, which divide into two possible errors (cells B and C) and two correct decisions (cells A and D).

According to Hany (1993), four indices may therefore be defined:

(1) Effectiveness (i.e. D/B + D)

(2) Efficiency (i.e. D/C + D)

(3) Alpha error (i.e. C/A + C)

(4) Beta error (i.e. B/B + D)

An index such as A + D/Total has also been suggested by Hany (1993). In this research, however, Effectiveness, Efficiency, alpha and beta errors are the indices and the measures of particular interest in analysing the two-stage identification process.

It would not be an exaggeration to state that, besides the evidence corresponding to the construct validity of the instruments used to obtain information about the indicators deemed appropriate, adequate standardization (i.e. norms) is one of the crucial aspects in considering the aptness of an instrument for selecting highly able students. However, since excellence and rareness are two criteria to keep in mind during the selection process also (cf. Sternberg, 1993; Sternberg & Zhang, 1995), reference to others is therefore required, and the norm of a test is its reference to the comparison group. The extent to which validity and standardization are important will become more apparent as the analysis of the identification process unfolds.

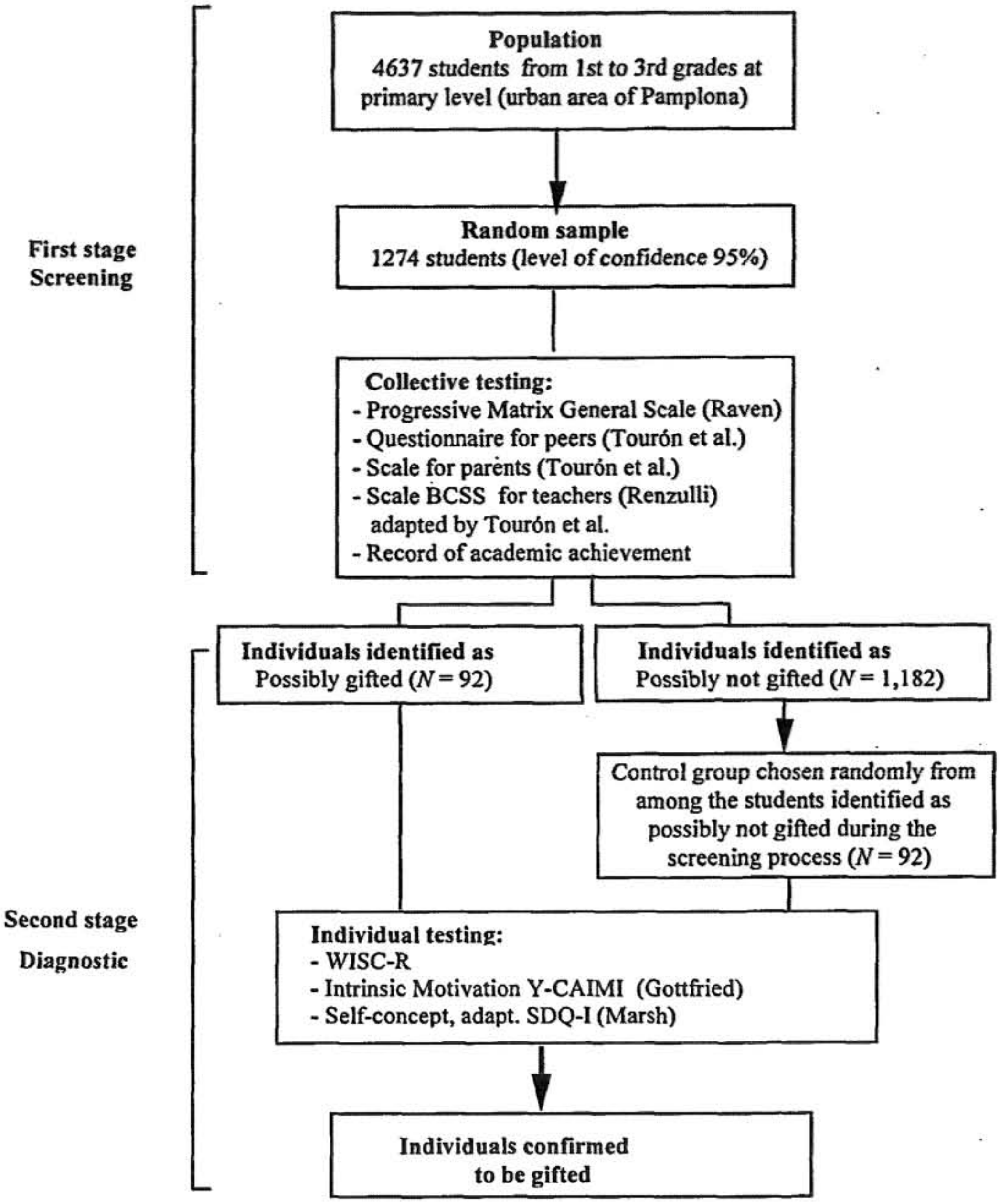

Figure 2. An outline of the adopted identification procedure (screening and diagnostic phases).

Table 1. Distribution of population and sample of elementary school children in Pamplona, classified by schools and grades. The desired sample in parentheses and the obtained sample in brackets

Type of school	Grade 1	Grade 1	Grade 3	Total
Public schools	409 (120) [119]	428 (125) [121]	480 (141) [127]	1317 (386) [367]
Private schools	1091 (321) [310]	1081 (316) [311]	1148 (337) [286]	3320 (974) [907]
School types combined	1500 (441) [429]	1509 (441) [432]	1628 (478) [413]	4637 (1360) [1274]

In this study, and based on the state of indigenous research in this field, as well as due to the scarcity of existing educational resources, we endeavour to establish a protocol of early identification by which first, second, and third-grade elementary school students in the Community of Navarra, with a high intellectual capacity, may be identified. The general character of the identification process targeted the analysis of the effectiveness and efficiency of a simple protocol of measuring instruments that would allow for a general, systematic application, and at the same time enable the progressive specificity of these processes, more in accordance with a talent search concept, so that educational planning will be facilitated (cf. Touron, Peralta & Repáraz, 1998b; Tourón, Repáraz, Peralta, Gaviria, Fernandez, Ramos & Reyero, 1998c).

Method

Procedure and Instruments

In the procedure followed for this study, we have clearly defined two phases: an initial screening phase, the accuracy of which will be analysed, and a second phase involving a confirming diagnostic. Figure 2 illustrates the complete identification process adopted for this research.

For the initial screening the following instruments were used: the Raven Progress-ive Matrices General Scale (Raven, Court, & Raven, 1996); the Scale for Rating Behavioural Characteristics of Superior Students (Renzulli, Smith, White, Callahan & Harman, 1976), which was adapted especially for this study (subscales of learning, motivation, creativity, and leadership). Furthermore, a record of academic achievement by subject area was used. Other instruments were used also, as indicated in Figure 2, but as they are not directly related to the aim of this article, their description is omitted here.

One of the first decisions to be made was to determine an adequate level of the Raven Progressive Matrices (RPM). Administering the Coloured Progressive Ma-trices (CPM) to several small groups of students, permitted us to detect a clear ceiling effect. We therefore decided to adopt the General Scale (SPM), which is a more difficult form. As there were no norms available for the SPM concerning these age groups, we established norms from the raw data gathered for present the study. These norms were then used as one of the student selection criteria.

The test was collectively administered in the randomly selected classrooms by the researchers and several graduate students specifically trained to perform this task. Although there is no time limit for this test we estimated an administration time of 30 minutes. It proved to be enough for the majority of the students. All the conventional administration procedures for intelligence tests were adhered to, as well as those recommended by the Spanish manual for this particular test.

Teachers were asked to use the adaptation of the SRBCSS (Renzulli *et al.,* 1976) in nominating students whom they considered to be the more brilliant in their classrooms. They also received adequate instructions on how to fill in the different subscales.

Students' academic grades, signifying their level of achievement, were gathered by the researchers from available school records.

We determined the following criteria whether or not a student should be included in the group identified as possibly gifted:

• a score of 1.5 *SD* above the mean as based on the Raven SPM local norm, close to the 95th percentile; a psychometric criterion regarding which there is, more or less, general consensus (cf. Richert, Alvino & McDonnel, 1982; Renzulli & Reis, 1991; Richert, 1991); or, alternatively,

• an outstanding academic achievement in all or some of the different curricular areas as well as being nominated by educators in more than one of the four SRBCSS subscales.

As an additional restriction, to select a student fulfilling the second criterion, but not the first, he/she must obtain a Raven score not lower than the cut-off score minus the measurement error also.

The students who fulfilled at least one of the two criteria would be considered possibly gifted, and thus become a candidate for individual assessment, which would allow us to confirm, or question, the validity of the adopted selection process.

The second diagnostic stage was performed individually with every student identified as possibly gifted, as well as on a same-sized control group of students *not* identified in the screening, but chosen randomly from amongst the individuals in the group identified as not gifted. The use of a control group is justified, as it would be impossible, in practical terms, to diagnose individually every one of the more than 1,000 students who were not selected. The results from the control group, however, will permit us to infer what would be expected in the whole non-selected group.

The variables measured were general intelligence using the WISC-R; that is the Wechsler Intelligence Scale for Children Revised (Wechsler, 1974, 1993); academic self-concept, measured by our adaptation of Marsh's (1988) Self-Description Questionnaire (SDQ-I). This test has been recently adapted to evaluate self-concept in school children of under 8 years of age (cf. Marsh, Craven & Debus, 1991), according to the multidimensional and hierarchic model proposed by Shavelson and Marsh (Shavelson, Hubner & Stanton, 1976; Shavelson & Marsh, 1986; Marsh, 1990a,b). Note that we have taken into account the results of our own validation of this scale also (Gonzalez, Tourón & Gaviria, 1994).

Motivation was measured using Gottfried's (1986, 1990) Young Children's Academic Intrinsic Motivation Inventory (Y-CAIMI), which is an inventory made up of three subscales (language, mathematics, and work at school). These aim to estimate students' academic motivation. However, there are doubts as to whether it is actually possible to use this construct with young children.

Lastly, in reference to the diagnostic stage procedures, the WISC-R emerged as being the main instrument of reference, since the self-concept and intrinsic motiv-ation instruments did not render statistical significant differences between the possibly gifted group and the control group, thus proving useless to the identification process. The WISC-R was administered mainly by the researchers (and a few tests were administer by other psychologists) to each one of the students, in full accord-ance with the instructions of the test manual. WISC-R is a well-known and respected test for individual assessment of intellectual ability, and is extensively used by practitioners in the field. As Kubinger (1998) pointed out "for both disabilities as well as high abilities, Wechsler's scales represent the most used intelligence test battery worldwide" (p. 237). It offers three scores: verbal IQ, performance IQ, and total IQ.

Using the Wechsler scales in identifying a student as intellectually gifted or not (or better: "WISC-R gifted" or not), we established as a criterion a cut-off score of 130 IQ (close to the 98th percentile) in any one of the IQs offered by the scales, and in accordance with the 1992 Spanish norms for this battery.

The Nature of the Sample

The target population for the research was elementary school students at 6 to 8 years of age (i.e. Grade 1 to Grade 3 in the Spanish school system). The sample was selected by using school levels as independent populations, from which a "protected" sample size was derived, comprising approximately 30% of the entire population of elementary school children in the urban area of Pamplona. We also considered the type of school (i.e. public or private), taking the school as our sampling unit. If a school had more than one classroom per level, they would all be included in the sample.

The sample size for each grade was calculated on the basis of a proportion estimates formula (with a 95% level of confidence, a sample error of 4%, and p and q = 0.50) for finite populations. Although we do not work with proportions, the sample size thus calculated is not less than the one corresponding to average values. Considering that the percentage of public and private schools in Pamplona was different, we proceeded according to a stratified random sampling procedure, maintaining the same proportion that the schools presented in the population.

Table 2. Spanish norms for the Raven Progressive Matrices General Scale, for elementary school children in the urban area of Pamplona

	Raw scores (for separate grades)						
Percentiles	First	Second	Third		z	IQ	Stanines
99	43	47	50	2.33	135		
98	41–42	45–46	49	2.06	131	9	
97	40	44	48	1.88	128		
95	38	43	47	1.64	124	8	
90	35	41	45	1.28	118		
85	34	39	43–44	1.04	116	7	
80	32	37–38	42	0.84	113		
75	29	36	40–41	0.67	110		
70	28	35	39	0.52	108		
65	26–27	34	38	0.39	107	6	
60	25	32–33	37	0.25	104		
55	23	31	35–36	0.12	102		
50	22	29–30	34	0.0	100	5	
45	20	28	33	−0.12	97		
40	19	26–27	32	−0.25	96		
35	18	25	30	−0.39	94		
30	17	23–24	28	−0.52	92	4	
25	16	22	25–26	−0.67	90		
20	15	19	23	−0.84	88	3	
15	13–14	15	20–21	−1.04	84		
10	12	14	17	−1.28	80	2	
5	11	12	14	−1.64	76		
3	10	11	13	−1.88	71		
2	9	10	12	−2.06	69	1	
1	7–8	9	11	−2.33	67		

In the future, if a general procedure were to be carried out, we would also consider the fact that students are grouped in classrooms, so a correction for cluster sampling should be made. For the present purpose, however, the sampling procedure followed is accurate enough. Table 1 shows the corresponding data for the desired sample and the obtained sample, according to grade and type of school.

When the determined screening criteria were applied, the sample was made up of, on the one hand, all subjects selected during this stage of the study (i.e. the "selected group"); and, on the other hand, an equal number of students chosen through random sampling carried out by grade on the total number of students not identified by the established criteria. This group served as control group. For the control group, however, we chose more students than necessary, to make up for any possible losses (change of school, moving from the city, and so on) in relation to the size of the group of the identified students. To avoid any researcher bias in terms of expectations concerning the achievement of participating students, researchers did not know whether a particular student selected for individual diagnosis belonged to the control group or to the group of students identified as gifted.

Table 3. Measures of central tendency and dispersion pertaining to raw scores (Raven Progressive Matrices General Scale)

	Grade level		
Statistic	First	Second	Third
N	426	429	410
M	23.03	28.51	32.67
SD	8.75	9.88	9.98

All data were analysed by using SPSS 4.0 for Macintosh (Norusis, 1990) employ-ing relevant statistical procedures.

Results

Results of the Raven test show that the average performance increases with grade level (see Tables 2 and 3 for norms and the relevant central tendency and dispersion measures). If these results are compared with the norms provided in test manual to the Spanish version of Raven's Progressive Matrices, which begins with fourth-grade students, it is obvious that our results follow a similar progression in relation to the grade (age) of the school children. Also, the standard deviation in the three groups is very similar.

Table 4. Distribution parameters and the results of the Kolmogorov–Smirnov (Lillieford's) goodness of fit test, per grade level, before and after standardizing the raw scores

Grade levels	Kurtosis	*SE* (Kurtosis)	Skewness	*SE* (Skewness)	Goodness of fit (K–S Lilleford's)
			Raw score distributions		
First	−0.616	0.236	0.408	0.118	0.0863 ($p<0.0000$)
Second	−0.869	0.235	−0.235	0.118	0.0681 ($p<0.0001$)
Third	−0.669	0.240	−0.435	0.121	0.0573 ($p<0.0026$)
			Standardized score distributions		
First	−0.079	0.236	−0.008	0.118	0.0381 ($p<0.2000$)
Second	−0.112	0.235	0.017	0.118	0.0359 ($p<0.2000$)
Third	−0.091	0.245	−0.004	0.121	0.0392 ($p<0.1377$)

It must be pointed out, however, that no study focusing on these grades, for this scale, has previously been done in Spain. We must also point out that out of the total score of the test (60 points), the third-grade students manage an average perform-ance of approximately 50%, when the test, in fact, is aimed at older students. A ceiling effect may therefore be ruled out.

The goodness of fit to normal distribution of the empirical data for each grade was checked using Lilliefords' approximation to the Kolmogorov-Smirnov test. Prior to standardizing the distribution, the probabilities associated with the values of this test did not allow acceptance of the null hypothesis of normality. However, once distributions were normalized, following a traditional transformation procedure, the null hypothesis of normality could be accepted for each of the grade levels (see Table 4).

The norms of the Raven Progressive Matrices were calculated based on the standardized and normalized raw scores. We developed three norms, one for each grade, where raw scores were expressed as percentiles, *z*-scores, IQs, and stanines (see Table 2). These data are important for IQ-testing in Spain particularly, since norms for the age groups concerned have not previously been available.

The internal consistency index (Cronbach's alpha) for the entire Raven test was $\alpha = 0.91$, $\alpha = 0.93$, and $\alpha = 0.94$, for first, second and third grades respectively. All values are within a more than acceptable range.

Regarding the study of the school children's behavioural characteristics, it is necessary to first point out the reduced size of the sample with which we have been working. It is important to keep in mind that the SRBCSS instrument (Renzulli *et al.,* 1976) was answered only by students chosen by the teachers. For example, in all three grades the Leadership subscale was consistently scored higher than the other subscales (Table 5).

Table 5. Mean scores ($N = 64$), by grade, and reliability of the Scales for Rating the Behavioural Characteristics of Superior Students (Renzulli *et al.*, 1976)

Subscales	First grade ($N = 23$)	Second grade ($N = 19$)	Third grade ($N = 25$)	Cronbach's α
Learning				
M	3.93	3.86	3.67	0.84
SD	0.71	0.40	0.52	
Creativity				
M	3.44	3.46	3.32	0.74
SD	0.51	0.49	0.48	
Motivation				
M	3.63	3.66	3.39	0.73
SD	0.50	0.48	0.54	
Leadership				
M	3.98	4.05	3.86	0.79
SD	0.47	0.58	0.46	

Since leadership is a characteristic that most teachers value and emphasize, it is likely it is also one they would most easily detect in their students.

Logically, all scores should be superior to the average score of the scale (i.e. approximately three points). However, in spite of the fact that teachers were asked to evaluate students who were noticeably outstanding, scores may still not be as high as could be expected; either because teachers have not discovered in their students the characteristics pointed out by the scale (since students, in fact, may not have the targeted qualities), or because the teachers did not receive enough information from the researchers. The reliability of each subscale, however, is acceptable, and these students make up approximately 5% of the total sample.

Selection of Possibly Gifted Students

All students with an IQ of 125 or more were immediately classified as possibly gifted. However, in accordance with the alternative identification criterion, we also chose students with an IQ of 115 or higher, who were nominated by their teachers and with an academic performance of four points or more in Language or Math-ematics. The criteria were the same for students in second and third grades, except that at these levels the cut-off score of the Raven scale was 43 points (equivalent to an empirical IQ of 122) and 47 points (or an IQ of 121) respectively. Based on these cut-off scores, and other discussed criteria, students were selected for the possibly gifted group. The resulting distribution over grade levels is show in Table 6.

Note that the size of the selected groups in the three grade levels ranges between 6.5 and 8% of the total sample with which we worked. A higher percentage than the initially estimated 5% of the population were therefore selected for the second diagnostic stage of the identification process.

Table 6 also shows that intelligence is the most powerful criterion for the selection of gifted students. But it is interesting to observe that IQ in combination with other criteria also allows for the selection of individuals who would otherwise be rejected. One must keep in mind, however, that the frequencies appearing in the table (corresponding to one, two, or three criteria respectively) are not excluding the fact that only the values of the first two rows can be added (IQ, Nomi-nation + Achievement), and that these are complementary. In fact, if added together, the total appearing in the last row *(N)* is obtained.

Table 6. Distribution of students, over grade level, selected to the possibly gifted group in accordance with determined criteria.

	Grade levels		
Selection criteria	First	Second	Third
IQ	26	26	24
Teacher nomination + achievement scores	9	2	6
IQ + teacher nomination + achievement scores	11	4	10
N (selected as possibly gifted)	35	28	30
Total group	429	432	413
Percentage of total group selected as possibly gifted	8	6.5	7.3

Table 7. The two indices of Effectiveness and Efficiency, as well as the alpha and beta errors of the identification process

Index	Ratio	Index value
Effectiveness	$\frac{D}{B+D}$	0.23
Efficiency	$\frac{D}{C+D}$	0.34
Alpha error (false positives)	$\frac{C}{A+C}$	0.06
Beta error (false negatives)	$\frac{B}{B+D}$	0.77

When the possibly gifted students were selected, we also chose another sample of the same size from the total number of students not selected. Thus, we had two groups: one experimental group (i.e. those selected as a result of the screening process [*N* = 92]), and one control group (i.e. a random sample of students not selected during the screening stage [*N* =92]).

The next step in the identification process was to determine who amongst the students, initially classified as belonging to the selected group or the control group, were correctly assigned to either group. That is to say, are the possibly gifted students identified through the screening process really gifted? To answer this question self-concept, motivation, and IQ (WISC-R), were measured. Given that the differences between the groups of students selected as possibly gifted and the control group were not significant as far as the variables of self-concept and motivation were concerned, the only diagnostic criterion that we maintained was a score of at least 130 points (IQ) in one of the scales (performance or verbal), or out of the total scale of the WISC-R.

The Efficiency and Effectiveness of the Identification Process

Once the values of the diverse IQs of the WISC-R were available, we were able to reclassify students into two groups, which we labelled "normal" and "gifted" respectively. This designation of students is the "definite" classification for this study. But it is certainly not the only possible classification for these students, since every identification process is flexible and dynamic, and must adapt to the changing nature of the special educational needs of students. If there were established programmes to which these selected students could be assigned, their performance there would provide valuable information (serving as an external criterion) in considering the accuracy of the selection. Since such an external criterion is not presently available, however, the data must be analysed with the restrictions imposed by the modest development of this field of study in the region where this research took place.

		Diagnostic: NORMAL	Diagnostic: GIFTED	Total
Screening	NORMAL ($N=92$)	A 84 True negatives (91.3 %)	B 8 False negatives (8.7 %)	A+B 92
Screening	GIFTED ($N=92$)	C 61 False positives (66.3 %)	D 31 True positives (33.7 %)	C+D 92
		A+C 145	B+D 39	184

Figure 3. Results of the identification process following the reclassification procedure.

The reclassification of the students was carried out on the assumption that the WISC-R may offer a secure diagnostic of the general intellectual capacity of students selected during the screening stage, and that scores may thus be used as a criterion for this further classification process. The group defined as gifted comprised stu-dents, who presented one of the IQs of at least 130 (logically, a more restrictive criterion than used during the screening stage). Thus, for example, a student with a verbal IQ of 130 or more, even with a lower performance or general IQ, would be classified in the gifted group. If not, the student would be reassigned to the normal group. The results of this reclassification procedure are shown in Figure 3. The figure is constructed so that it shows the number of subjects in each group, based on the screening, and following the diagnostic results. It must be pointed out, however, that the results outlined in Figure 3 are those corresponding to the samples comprising 92 students (i.e. selected and control groups).

It is also important to keep in mind that the Raven scores (in addition to the other criteria) permitted us to split the sample of 1,274 students into two samples of 92 students (i.e. the selected group) and 1,182 students (i.e. the non-selected group). From the non-selected group we studied a representative sample of 92 students (i.e. the control group). If applying the percentages obtained in the process to the whole sample of 1,182 students, the resulting would be as outlined in Figure 4.

		Diagnostic		
		NORMAL	GIFTED	Total
Screening	NORMAL (N = 1182)	A 1079 True negatives (91.3 %)	B 103 False negatives (8.7 %)	A + B 1182
	GIFTED (N = 92)	C 61 False positives (66.3 %)	D 31 True positives (33.7 %)	C + D 92
		A + C 1140	B + D 134	1274

Figure 4. Results of the identification process following the reclassification procedure and projected on to the whole sample.

In first considering the decisions (classifications) that appear to be correct in Tables 3 and 4, 91% of the students are categorized as normal "true negatives", which must be considered quite satisfactory. However, when turning to students classified as gifted during the screening stage, and confirmed during the diagnostic stage, the values shown represent a global percentage of "true positives" at 34%. It is logical to have a higher percentage of incorrectly classified subjects in this group, since the chosen identification process tended to be comprehensive when in reason-able doubt of giftedness.

However, there are two incorrect decisions in the identification process. One of them concerns ignoring individuals who should have been taken into account since they do have the sought qualities. These are students are the "false negatives" in the procedure. This is an important error to observe, especially when decision making implies the likely assignment to a programme or to some type of educational provision. The importance of depriving necessary help to a student who could have benefited from it is obvious. In the identification process outlined in this article, 8.7% of students have been incorrectly identified.

Note that there is yet another possible error in the process, which is less important, namely to single out as possibly gifted students who in reality are not. Individuals mistakenly selected as possibly gifted would not participate in planned programmes or other educational provision, since their status as gifted would not be confirmed. Needless to say, this error should still be avoided to prevent false expectations and undesirable experiences for both students and their parents. In the current study process, this type of error translates into 66% of "false positives". This percentage, although conspicuously high, is still within reasonable limits. Again, this would most likely be a result of the comprehensive character of the chosen identification process.

The percentages of true and false positives (34 and 66% respectively), as well as those of the true and false negatives (91 and 9%), are complementary.

These values allow the calculation of the indices of Effectiveness and Efficiency, as well as the alpha and beta errors of the identification process (see Table 7).

In accordance with the arrived-at Index values and alpha and beta errors, it would appear that the identification process was neither effective nor efficient. It produced a very high beta error (false negatives) and a very small alpha error. However, it must be remembered that in order to establish the indices, one inevitable assumption was that there also was a way of identifying those who are "truly gifted" and those who are not (keeping in mind the particular meaning of this expression in this context). For this study WISC-R was the tool for making this decision, and therefore becomes the reference point for any effectiveness or efficiency analysis. However, there appears to exist some problems with the WISC-R instrument, which we did not anticipate and that were uncovered in the process.

WISC-R and the Identification Process

A closer inspection of the results (see Figure 4) allows additional comments to explain the surprising outcome of the process. Firstly, out of the total sample of 1,274 students, there are supposedly 134 "WISC-R gifted" students representing approximately 10% of the population. However, applying the more restrictive selection criterion, and in addition considering only the total IQ, the percentage should reduce to 6.5%. But the sample of 1,274 students was randomly selected from the Pamplona population of elementary school students. If random rules apply, as they must, and considering that IQ is normally distributed, as it is, with a mean of 100 and a standard deviation of 15, no more than 2.28% of the students could be expected to have scores equal or higher than 130. Since this percentage is clearly incompatible with the percentage expected, one must ask why these results emerge, and suspend any evaluation of the presented indices. In fact, if the validity of the reference criterion (WISC-R) cannot be accepted as a tool by which to classify the students as gifted or not, in our view, it seems clear that an evaluation of the identification process outlined in this article cannot be conclusive.

We can be reasonably confident with the Raven results, however, since these are based on local norms developed for this study, from random samples, representing approximately 30% of the total local population. This test has a recognized validity and the reliability is over 0.90.

The relationship between the Raven test and the WISC-R, as estimated by means of Pearson's coefficient, ranges from 0.45 for verbal IQ to 0.55 for performance IQ. Independent of this comparison, and of what could be considered only a modest predictive validity of the Raven with regards to WISC-R (particularly the verbal section, which could be expected due to the different constructs measured; i.e. figurative versus verbal content), the main problem was the discrepancy between the expected percentage of students identified as "truly gifted" (as based on WISC-R) and the *actual* percentage. It is a fact that the number of "WISC-R gifted" students detected are unacceptable from any point of view. There are neither theoretical nor practical reasons to believe that there should exist such an unusually large number of gifted students in Pamplona. Assuming that tests were correctly scored, according to the scoring rules offered by the Spanish manual, one must wonder whether the test, the scoring rules, or both, are biased in some way. Kubinger (1998) recently revealed, from a study conducted with the German (from 1956 and its revision 1983) and American WISC-R editions, that "these scales suffer from serious psychometric shortcomings and, consequently, although it is rarely realized, the assessment of ability can be often incorrect" (p. 238). The shortcomings observed by Kubinger refer to the scoring rules and address the problem we discovered. We do not know of any similar study carried out on the Spanish version of the WISC-R, but since it contains most of the items of the American edition, we need to consider the possibility of a similar problem. However, there are also further possibilities. The norms established in 1992 were arrived at on the basis of a sample of limited size (i.e. 724 subjects for 11 age groups). This could mean that sampling error was a considerable problem, and that other biases regarding the adequate representation of the population were present also. It could perhaps also be argued that the norms from 1992 are already outdated, and therefore could cause the so-called Flynn Effect (i.e. the possible increase in IQ due to environmental effects on cognitive abilities). This effect could cause a considerable shift of points of reference with a least 15 IQ points per generation (cf. Sternberg & Kaufman, 1998, for a short description, and also Flynn, 1984, 1987).

However, we do not consider this to be a plausible explanation in our particular case. Unfortunately, at this time, we cannot offer a conclusive explanation as to why the Spanish WISC-R norms do not agree with our actual results. No matter what the explanation might be, it is obvious that our results focus something that is well known theoretically, but not always realized practically, namely analysing the val-idity of the instruments used for measurement, and especially considering the importance of updating the norms on which such instruments operate. Since high ability is considered, at least in part, to be a matter of excellence and rarity, it is crucial to have adequate norms for reference in making decisions. Although our findings in this matter cannot be considered conclusive by any means, they do provide essential information for those involved in identifying giftedness and who carry out careful, detailed analyses, and assess the psychometric instruments they often use for such purposes. Although we did not anticipate this particular problem, we realize in hindsight the significance of our decision to accept the WISC-R more or less uncritically. It led us to discover its present problems!

Discussion and Conclusion

Out of the 1,274 students in Pamplona who made up the initial sample, we identified 92 students as possibly gifted (i.e. 7%) during the screening stage. For the diagnostic phase, designed to confirm or reclassify students tentatively identified as "possibly gifted", we involved another 92 individuals, completely at random, from the rest of the initial sample. All of the resulting 184 students were subjected to an individual testing using instruments such as WISC-R, Y-CAIMI, and SDQ-I. The Spanish adaptation of Y-CAIMI and SDQ-I, administered to the sample, confirmed that students in first, second and third grades (6 to 8 years of age) tend to evaluate themselves in an excessively positive manner, both in regard to their academic self-concept and in their learning motivation. In other words, they tend to generalize their positive attitude having not yet developed their cognitive ability to make fine discriminations as required by these instruments. This discovery made us establish the diagnostic stage mainly on IQ scores and make confirmatory or reclassification decisions on the basis of WISC-R results.

Thus, while there are important differences in the intelligence variable among students in the control group (not selected during the screening) and the selected students, such differences are not apparent in regard to self-concept and intrinsic motivation variables. Students in both groups tended to score similarly on this type of variable.

It seems appropriate, therefore, in conclusion, to evaluate the entire identification process chosen for this research, including the fact that the process also uncovered major problems with the WISC-R results. All this considered, there is no possible way, at present, to conclude whether or not the chosen process is efficient or effective; nor to make any definite statements concerning the magnitude of the possible alpha and beta errors committed. However, results may nevertheless lead to important decisions regarding psychometric norms and standardization of the tests most widely used in Spain. One important consideration also for the future of psychometric assessments, which was highlighted in following the identification process outlined in this article, is that it is most likely impossible to pursue such a process without paying much attention to the measurement instruments used. While this is important to all types of psychometric assessments, it is of particular importance in the identification of giftedness, since the whole notion of intellectual excellence and rarity is a construct largely based on valid and reliable psychometric criteria. In this research we clearly encounter this problem. We are convinced, however, that an early identification is essential to avoid losing abilities and talents for lack of attention or knowledge. But it is also essential that a careful and technical review of all the instruments involved in an identification process be made. Since this usually is carried out regularly in most countries, researchers and other professionals working in this field must take upon themselves the task of validating instruments and establishing adequate norms for use in the identification processes, so that we may be certain of an acceptable level of reliability and validity. And even if a "general identification" is acceptable with the adequate instruments, in the present situation, we are also convinced, that from this point on, it is necessary to take on the development or adaptation of test batteries that will allow the identification of specific talents, and guide in a more precise manner the nature of educational provision needed by these particular students. This work is currently in process in Pamplona. We have launched a project aimed at validating an instrument which will facilitate the identification of individuals with both verbal and mathematical talent, as well as developing guidelines that should be taken into account when establishing specific educational strategies and/or programmes. It is our hope that the instrument under validation will be free of the psychometric shortcomings discussed in this article.

Note

This article is in part based on an invited paper presented at the I Congreso Internacional de Educatión de la Alta Intelegencia, Mendoza, Argentina, August 1998. Acknowledgement The authors wish to extend their gratitude to the two anonymous referees, as well as to the Editor in chief of *High Ability Studies,* for constructive criticism and assistance in improving the article.

References

CLARK, B. (1992). Growing up gifted: Developing the potential of children at home and at school. (4th ed). Columbus, OH: Macmillan.

FELDHUSEN, J. F. & BASKA, L. (1989). Identification of the gifted. In J. F. FELDHUSEN, J. VAN TASSEL BASKA & K. SEELEY (Eds.), *Toward excellence in gifted education* (pp. 85-101). Denver, CO: Love.

FELDHUSEN, J. F. & JARWAN, F. A. (1993). Identification of gifted and talented youth for educational programmes. In K. A. HELLER, F. J. MONKS & A. H. PASSOW (Eds.), *International handbook of research and development of giftedness and talent* (pp. 233-251). Oxford, UK: Pergamon Press.

FLYNN, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1938. *Psychological Bulletin, 95,* 29-51.

FLYNN, J. R. (1987). Massive gains in 14 nations: what IQ tests really measure. *Psychological Bulletin, 101,* 171-191.

GONZALEZ, M. C , TOURÓN, J., & GAVIRIA, J. L. (1994). Validation del cuestionario de autodescripción de Marsh (SDQ-I) en el ámbito español [Validation of Marsh's Self-Description Questionnaire in a Spanish setting]. *Revista de Investigación Educativa, 24,* 7-26.

GOTTFRIED, A. E. (1986). *Children's Academic Intrinsic Motivation Inventory.* Odessa, FL: Psycho-logical Assessment Resources.

GOTTFRIED, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82*(3), 525-538.

HAGEN, E. (1980). *Identification of the gifted.* New York: Teachers College Press.

HANY, E. A. (1993). Methodological problems and issues concerning identification. In K. A. HELLER, F. J. Mönks & A. H. PASSOW (Eds.), *International handbook of research and development of giftedness and talent* (pp. 209-232). Oxford, UK: Pergamon Press.

KUBINGER, K. D. (1998). Psychological assessment of high ability: Worldwide-used Wechsler's intelligence scales and their psychometric shortcomings. *High Ability Studies,* 9(2), 237-251.

MARSH, H. W. (1988). *Self -Description Questionnaire: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept. A test manual and a research monograph.* San Antonio, TX: Psychological Corporation.

MARSH, H. W. (1990a). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2,* 77-172.

MARSH, H. W. (1990b). The structure of academic self-concept: The Marsh/Shavelson Model. *Journal of Educational Psycology, 82*(4), 623-636.

MARSH, H. W., CRAVEN, R. G., & DEBUS, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83*(3), 377-392.

NORUSIS, M. J. (1990). *SPSS®. Base system user's guide.* Chicago, IL: SPSS, Inc. RAVEN, J. C , COURT, J. H., & RAVEN, J. (1996). *Raven Matrices Progresivas. Escalas Color (CPM), General (SPM) y Superior (APM).* Manual (segunda edición ampliada). Madrid, Spain: TEA Ediciones.

RENZULLI, J. S., & REIS, S. M. (1991). The Schoolwide Enrichment Model: A comprehensive plan for the development of creative productivity. In N . COLANGELO & G. E. DAVIS (Eds.), *Handbook of gifted education* (pp. 111-141). Boston, MA: Allyn & Bacon.

RENZULLI, J. S., SMITH, L. H., WHITE, A. J., CALLAHAN, C. M., & HARMAN, R. K. (1976). *Scales for Rating the Behavioral Characteristics of Superior Students.* Mansfield Center, CT: Creative Learning Press.

RICHERT, E. S. (1991). Rampant problems and promising practices in identification. In N . COLANGELO & G. A. DAVIS (Eds.), *Handbook of gifted education* (pp. 81-96). Boston, MA: Allyn & Bacon.

RICHERT, E. S., ALVINO, J., & MCDONNEL, R. (1982). *The national report on Identification: Assess-ment and recommendations for comprehensive identification of gifted and talented youth.* Sewell, NJ: Educational Information and Resource Center for US Department of Education.

SHAVELSON, R. J, & MARSH, H. W. (1986). On the structure of self-concept. In R. SHWARZER (Ed.), *Anxiety and Cognitions.* (pp. 305-330). Hillsdale, NJ: Lawrence Erlbaum.

SHAVELSON, R. J., HUBNER, J. J., & STANTON, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*(3), 407-441.

STERNBERG, R. J. (1993). Procedures for identifying intellectual potential in the gifted: A perspec-tive on alternative "Metaphors of Mind". In K. A. HELLER, F. J. MONKS & A. H. PASSOW (Eds.), *International handbook of research and development of giftedness and talent* (pp. 185-207). Oxford, UK: Pergamon Press.

STERNBERG, R. J., & KAUFMAN, J. C. (1988). Human abilities. *Annual Review of Psychology, 49,* 479-502

STERNBERG, R. J. & ZHANG, L (1995). What do we mean by giftedness? A pentagonal implicit theory. *Gifted Child Quarterly, 39(2),* 88-94.

TANNENBAUM, A. J. (1983). *Gifted children: Psychological and educational perspectives.* New York: Macmillan.

TOURÓN, J., IRIARTE, C., REPÁRAZ, CH., & PERALTA, F . (1998a). Diversity and school curriculum: The response of the Spanish educational system to the needs of academically highly able pupils. *High Ability Studies, 9(2),* 165-180.

TOURÓN, J., PERALTA, F., & REPARAZ, CH. (1998b). La *Superdotación intelectual: Modelos, identificacion y estrategias educativas. [Intellectual giftedness: Models, identification and educa-tional strategies*]. Pamplona, Spain: EUNSA.

TOURÓN, J., REPARAZ, CH., PERALTA, F., GAVIRIA, J. L., FERNÁNDEZ, R., RAMOS, J. M., & REYERO, M. (1998c). *Identificacion del talento verbal y matemdtico: Description de un proyecto de validation [Verbal and mathematical talent identification: Description of a validation project*]. Paper presented at the Congreso International sobre Respuestas Educativas para alumnos Superdotados y Talentosos. Zaragoza, Spain, 8-11 July.

TREFFINGER, D. J., & FELDHUSEN, J. F . (1996).Talent recognition and development: Succesor to gifted Education. *Journal for the Education of the Gifted, 19(2),* 181-193.

WECHSLER, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised.* New York: Psychological Corporation.

WECHSLER, D. (1993). WISC-R. *Escala de inteligencia de Wechsler para niños-revisada [Manual of the Wechsler Intelligence Scale for Children-Revised*]. Madrid, Spain: TEA Ediciones.