The Spanish Journal of Psychology
2008, Vol. 11, No. 2, 400-413

ISSN 1138-7416

Variables that Predict Academic Achievement in the Spanish Compulsory Secondary Educational System: A Longitudinal, Multi-Level Analysis

Elena Martín[1], Rosario Martínez-Arias[2], Alvaro Marchesi[2], and Eva M. Pérez[2]
[1]Universidad Autónoma de Madrid (Spain)
[2]Universidad Complutense (Spain)

This article presents a study whose objective was to identify certain personal and institutional variables that are associated with academic achievement among Spanish, secondary school students, and to analyze their influence on the progress of those students over the course of that stage of their education. In order to do this, a longitudinal, multi-level study was conducted in which a total of 965 students and 27 different schools were evaluated in Language, Math and Social Science at three different times (beginning, middle and end of the period). The results show progress in all the schools and in all areas. As for the personal, student variables, the longitudinal, HLM analyses confirmed the importance of sex and sociocultural background and, distinguishing it from other studies, also the predictive capacity of meta-cognitive abilities and learning strategies on success in school. On the institutional level, the school climate and teachers' expectations of their students were the most relevant of the variables studied. The size of the school, the percentage of students who repeat grades, and the leadership of the administration also explained a portion of the variance in some areas.
Keywords: academic achievement; predictor variables; secondary education; longitudinal multi-level analysis; school effectiveness

En el artículo se presenta un estudio cuyo objetivo es identificar determinadas variables personales y de centro asociadas con el rendimiento académico de estudiantes de secundaria españoles y analizar su influencia en el progreso de los alumnos a lo largo de la esta etapa. Para ello, Se realizó un estudio multinivel longitudinal en el que se evaluó a un total de 965 estudiantes de 27 centros distintos en Lengua, Matemáticas y Ciencias Sociales, en tres momentos (inicio, mitad y final de la etapa). Los resultados mostraron progreso en el conjunto de los centros en todas las áreas. Los análisis HLM longitudinales confirmaron en el nivel personal la importancia del sexo y el nivel sociocultural y, a diferencia de otros estudios, también la capacidad predictiva de las habilidades metacognitivas y las estrategias de aprendizaje. En el nivel de escuela, el clima escolar y las expectativas del profesorado hacia los estudiantes fueron las variables más relevantes. El tamaño del centro, el porcentaje de repetidores y el liderazgo del equipo directivo explicaron también una proporción de la varianza en algunas áreas.
Palabras clave: rendimiento escolar, variables predictoras, educación secundaria, análisis mutinivel longitudinal, eficacia de la escuela

Correspondence concerning this article should be addressed to Elena Martín Ortega, Departamento de Psicología Evolutiva y de la Educación, Universidad Autónoma de Madrid, 28749 Madrid. Phone: 34 914975176. FAX: 34 914973268. E-mail: elena.martin@uam.es

How to cite the authors of this article: Martín, E., Martínez-Arias, R., Marchesi, A. and Pérez, E.M.

Worries about the academic achievement of students have been of constant concern to research on education. Studies about the effectiveness of schools have centered their interest on identifying the factors that are best able to predict achievement in order to intervene early in the process of education to curb academic failure. Fortunately, the procedures for assessing achievement have become increasingly strong, from both a theoretical and methodological perspective. The four factors for evaluation in Stufflebeam and Shinkfield's global model (1985) – context, input, process, and output – are now considered to be essential elements of any evaluation. However, this perspective has been widely expanded upon by studies of the effectiveness of schools that were conducted using other, also valuable, models that have recently begun to include longitudinal analysis (Braun, 2005; Lissitz, 2005). The findings of those studies have been accompanied by methodological developments on multi-level analysis, which allows researchers to view the structural hierarchy of data on educational from a longitudinal perspective (Goldstein, 2003; Hox, 2002; Raudenbush & Bryk, 2002; Sellström & Bremberg, 2006; Zvoch & Stevens, 2006).

Still, only few research studies have used longitudinal, multi-level analysis and among them, even fewer have followed the same students through their education. As Goldstein and Behuniak (2005) stated about what they refer to as "growth models," there have been several papers that simultaneously studied multiple cohorts, others have followed the same group of students through a longitudinal process and others still have employed mixed models.

Several studies have been conducted about achievement during Compulsory Secondary Education (ESO) in Spain (Instituto Nacional de Evaluación y Calidad del Sistema Educativo [INECSE], 2001, 2003, 2004, 2005). However, very few of them have provided the sort of information that could identify variables to predict the results, and even fewer have been conducted using the analytical methodologies mentioned above. The objective of the present study is to understand in greater depth the complex interaction of the factors that influence academic achievement during the four years of ESO in the subjects of Math, Language and Social Science, using a longitudinal model and multi-level analysis.

We will now highlight the conclusions of the most relevant studies of predictor variables and achievement, upon which the present study will be based. The literature review will first explain general, international studies on the topic and will then delve more specifically into data from Spanish samples.

The results of international research

The most highly studied variable has been, beyond a doubt, socioeconomic and cultural status (SES), both that of students themselves, and that of their schools as a whole. In the case of students, the results have been rather consistent: a clear relationship has been found between the sociocultural context of the child's family and his or her academic achievement (OECD, 2004; Mortimore, Sammons, Stoll, Lewis y Ecob, 1988, Teddlie, Stringfield & Reynolds, 2000). Recent studies have also found a relationship between the average sociocultural status of schools and the learning of their students (Haahr, Nielsen, Hansen, & Jakobsen, 2005; OECD, 2004; Sellström & Bremberg, 2006). The predictive capacity of this variable, however, has been found to be less than the SES of the student's family. In spite of the data that has shown this to be one of the factors with the greatest influence on the students' achievement levels, the variance attributed to SES ranges and differs considerably study to study, as Sellström and Bremberg's interesting meta-analysis showed. The difference between these results obviously depends on the methodology used in the studies in question, but also on whether or not other variables are taken into account, both personal to the student and specific to the structure of the school.

One example of the weak relationship between SES and other variables is the case of private versus public schooling. Studies have been inconclusive on this topic. As early as the 1980's, Coleman and Hoffer (1987) found that in the United States, religious, private schools yielded higher levels of academic achievement than public schools did. The results of the Program of International Student Assessment study in 2003, however, showed that when SES was taken into account, the difference that usually seems to favor private schools disappears (OECD, 2004). A recent study (Corten & Dronkers, 2006) demonstrated that when one's sociocultural status is low, the difference between private and public schooling also disappears.

The influence of SES and of private versus public education on achievement also depends on the criteria being analyzed to determine that achievement. A study published in 2006 by the Educational Testing Service, conducted by Braun, Jenkins, and Grigg, showed that the influence of private versus public education is annulled when SES is factored into a multi-level analysis of reading comprehension tests, while it is maintained in the case of upper-level math classes, where public school students actually have higher levels of achievement. In the PISA studies, when both variables are incorporated into the multi-level analysis models, the influence of schools' private or public status on variance disappears.

In addition to SES and public versus private education, school environment is a factor that has finally begun to attract a great deal of attention. Numerous studies have found a relationship between the type of interaction that characterizes a serene and respectful environment, positive interpersonal relations, and high levels of learning (Haahr et al., 2005; Freiberg, 1999; OECD, 2004; Sellström & Bremberg, 2006. In these studies, it was postulated that the influence of school environment is, in part, executed by its consequences in the classroom and, in part, due to the direct manner in which students learn how to behave and approach

learning. In some studies (OECD, 2004), within the concept of school environment, the teachers' expectations of their students are included.

The procedural model found in educational institutions has also been introduced into some workplaces. Within that model, functions as diverse as level of autonomy, the administration's type of leadership, institutional organization, coordination of pedagogy and changes in the institution's culture are considered. In the PISA studies (OECD, 2004), the debate has primarily centered on the theme of autonomy as being related to high achievement levels.

Last, school size has also been shown to exhibit an influence over academic achievement. Raywid (1997) found a relationship between small schools and higher levels of learning. Similarly, Marchesi & Martínez Arias (2006) confirmed that same tendency in a Spanish PISA sample, although its proportion of the explained variance was very small.

When variables personal to the students are studied, in addition to the sociocultural status to which we have already referred, a relationship is usually found between sex and academic achievement. Girls tend to perform worse in Math and Natural Science and to perform better at Reading Comprehension and Language, the latter usually being a more considerable difference (Beaton et al., 1996; Haahr, et al. 2005; Mullis et al., 1998; OECD, 2001, 2004). However, although this generally tends to be found in studies, these results are debatable to a greater or lesser degree depending on the country. Holland, Japan, Hong Kong, China, and Hungry do not exhibit gender differences in the PISA studies. Also, in studies done by Opdenakker & collaborators (Opdenakker, Van Damme, De Fraile, Van Landeghem, & Onghena, 2002; De Fraine, Van Damme, & Onghena, 2002), results have shown greater achievement on the part of the female students in both Math and Language.

The influence of sex has also begun to be studied at the school and classroom levels by measuring the ratio of male to female students in a given class and other aspects. In the Oppendaker studies, as well as others we have cited thus far, the percentage of girls in the class is correlated with higher achievement in both Math and Language.

The influence of learning strategies on academic achievement, on the other hand, has been much less widely investigated, in spite of its theoretical importance and prevalence in international reports. One's competency at forming these strategies is recognizably liked learning to learn (Definition and Selection of Competencies [DeSeCo], 2005; EURYDICE, 2002; European Commission, 2004). The very interesting Haahr et al. report (2005) about the results of PISA and the Trends in International Mathematics and Sciencie Study (TIMMS) states that there is no relationship between the type of learning strategy employed by students (control, memorization, elaboration) and their scores on Math and Reading Comprehension tests. As the author of that study said himself, that result contradicts the widely-held belief that that type of instructional process has importance.

Studies with Spanish samples

The most important results about academic achievement in Spanish samples are without a doubt those provided by international studies (PISA, TIMMS) that include Spain and those conducted by the National Institute for Quality and Evaluation in the Education System (INECSE). However, theirs was not a case where multi-level analysis was used.

As far as what are referred to as school variables, in the PISA studies, the average SES of the school had the same tendency among Spanish students as it had in other countries, but to a lesser extent. The percentage of the explained variance between schools explained by school's average SES was 20% on average, according to the OECD, while in Spain it was only 14%. In fact, Spain is one of the 10 countries where wealth distribution is most equal across schools. On the other hand, the Marchesi & Martínez Arias study (2006) showed that the disparity between the more and less privileged social classes, reaching 114 points in the OECD data, is smaller in the case of Spanish students (only 84 points). Also, in a related study about ESO students by Marchesi, Martínez Arias & Martín (2004), it was found that the influence of SES was not equal in all contexts. The results of students in the lowest sociocultural context – in a four-level classification system – were noticeably inferior to the results of students from the three other contexts.

On the other hand, in the PISA studies that used Spanish samples, as with the majority of other countries, the differences between private and public school education disappeared when SES was introduced into the analysis (INECSE, 2004). It is interesting to consider as well that in the case of Math, the influence of the school on students' success was below the OECD average; it was the students' personal variables that explained the majority of the variance in that subject area.

With respect to personal variables, in the case of Spanish student samples, the SES of students' families was another factor that explained a high percentage of the variance, but it was significantly less than the mean from the OECD in the PISA studies (Marchesi & Martinez Arias, 2006; INECSE, 2004, 2005).

Students' sex shows the same tendencies that were found in most of the aforementioned research studies. Girls scored higher on Language tests according to the PISA 2000 data (INECSE, 2005) and their ESO evaluations (INECSE, 2003), while the boys scored higher on Math, supporting the results of both PISA and TIMMS, and the ESO evaluations done by INECSE (1997, 2003, 2004). As for Natural Science, the results vary. The TIMMS studies have found that men have better results on that subject, but in the Spanish sample in the PISA study, the differences between the sexes were not found to be significant. With regards to Social Science, the INECSE (2003) study found that girls had higher scores. Nevertheless, it is worth mentioning that in some of the autonomous communities that participated in the PISA 2003

sample, no differences were found between the sexes in achievement in Math. In The Basque Country and the region of Castilla and León, it was the case that male and female students obtained similar results. In another finding related to the influence of sex, the Spanish PISA sample showed that the total percentage of female students in the school is a predictor of higher achievement in Math, but the female students still tend to score lower than male students (Marchesi & Martínez Arias, 2006).

Last, departing from what occurred in the general, PISA sample, in the PISA sample of Spanish students, a relationship between learning strategies and academic achievement was found. To be specific, the students that utilized memorization strategies the most had worse results (Marchesi y Martínez, 2006).

The theorists whose work was cited earlier in this paper have given us a great deal of interesting information about the influence of certain factors on the academic achievement of both male and female students. However, studies of Spanish samples have not been conducted in which those variables' influence over changes in level of achievement is explored over a long period of time or during the span of an entire stage of education, nor have they employed longitudinal, multi-level analysis. To fill those gaps in information is precisely the objective of the present study. This study attempts to concretely address the following hypotheses and key questions for research: Do personal variables and variables related to the school exhibit a different level of influence on students' progress toward academic achievement? Does the influence of those variables vary at the beginning of the educational stage and does it vary over the course of the years and progress it produces? Last, does their influence differ depending on the class?

Method

Participants

The present study was carried out in 27 Secondary Education schools in the Community of Madrid with 965 student participants who were enrolled in said schools. The schools' participation in the study was voluntary, so the sample is considered to be incidental. In each school, at the beginning of the 1997-1998 academic year, one or two first-year classes were randomly selected, depending on the school's size. The students in those classes were followed over the 4-year course of their ESO education. The initial number of participating schools was 34, with 2,039 students participating. Due to the fact that some students had to repeat grades, switched schools or dropped-out in some of the schools during the 4 years of the study, the final number of schools participating in the study was reduced to 27, with only 965 students participating.

The study was conducted in 17 private schools and 10 public schools. All of them were located in the Community of Madrid. The schools ranged in sociocultural context according to the classification criteria for the sociocultural variable explained later in this paper. 5 schools were of a high sociocultural status, 6 middle-high, 8 middle-low and 8 low. Of the participating students, 435 (45.1%) were female and 530 (54.9%) were male. At the end of the study, in the third trimester of the 2000-2001 academic year, all participants were students in their 4th year of ESO and their average age was 15.78 years old (*SD* = 0.82 years), ranging from 15 to 18 years old. 80% of the participants were 15 or 16 years old, normal ages for that grade. The average age of the male students was 15.81 years old (with a standard deviation of 0.82 years), while the mean of the female students' ages was 15.74 years old (*SD* = 0.83 years). No statistically significant differences were found between the respective ages of the male and female students. As far as sociocultural context, the distribution of the students was the following: 206 (21.3%) were of a high sociocultural status, 278 (28.7%) were medium-high, 293 (30.5%) were medium-low and 188 (19.5%) were of a low sociocultural status. The number of participating subjects varied from school to school, ranging from 11 to 76 students. The mean number of students from each school was 36.

The structural data corresponding to variables associated with the schools were provided by the school administrations, while what we refer to as the schools' functional variables were collected from a sample of 656 teachers who participated voluntarily and anonymously during the 1999-2000 academic year.

Variables and Measurement Instruments

In this study, data was collected about variables from two levels of analysis: that of the educational center and that of the students themselves. Included among the *school variables,* the following variables were measured:

Private or public, considered in this study to be a dichotomous variable (0 = private school, 1 = public school).

School size, defined by the number of groups in each grade in the first-year class of ESO, and also treated as a dichotomous variable: small (0 = with only 2 lines) and large (1 = with three or more lines of names).

Rate of students repeating grades, in other words, the percentage of students that have to repeat an academic year during their four years of ESO. This information was provided by the schools' administrations.

Percentage of girls enrolled as first years in ESO. This information, too, was provided by the schools' administrations.

Mean sociocultural status of the school: An aggregate variable defined as the average sociocultural status of the participating students at each school. The schools were classified into four groups according to their scores on this variable, set off by the quarters of the distribution (High

status: scores from 108 to 124; Medium-high status: 100-108; Medium-low status: 93-100; Low status: 86-93).

Finally, another three variables were measured – the leadership of the administration, the school environment, and the teachers' expectations – that were identified as important factors from the results of the questionnaires that were administered to the teachers, the details of which are mentioned above. A principal component analysis was performed on their data using a Varimax rotation strategy.

The administration's leadership. This refers to the type of procedures that the school administration follows, their organizational environment, and their pedagogical coordination. It includes items such as "In general, the administration makes appropriate decisions at the right time" and "The principal should listen to the teachers more." It was calculated by finding the mean of the scores reported by faculty on 13 items, and those items were on a 7-point, Likert scale. Cronbach's alpha coefficient for this scale was .86.

The school environment. This refers to the interpersonal relations between faculty and students and the presence of an atmosphere at school characterized by participation and respect. It includes items such as: "I think that, in general, my students appreciate me" and "It is necessary to keep the students' opinions in mind when determining the school disciplinary rules." This variable was assessed through the average scores of the faculty on a scale comprised of 7 items. Cronbach's alpha coefficient for this scale was .84.

Expectations. This refers both to the faculty's assessment of the educational institution and to their prognosis for its future based on that state – "It does not matter to me whether or not my child goes here" – and to their expectations of their students' academic achievements and failures – "The students at this school leave very well prepared." It was determined by the mean of the faculty's scores on a scale comprised of 9 items. Cronbach's alpha coefficient for this scale was .85.

Variables associated with the students

Sociodemographic variables: Sex (0 = male, 1 = female) and age

Learning strategies. Two scales from the CEAM test (Learning Strategies and Motivation Questionnaire) (Ayala, Martínez Arias, & Yuste, 2004) were used, each consisting of 10 items: strategic understanding was measured by items such as "When studying, I sometimes think about the key questions relevant to what I am reading" (Cronbach's alpha coefficient = .85) and superficial learning, or learning based on memorization, is measured by items such as "When I do not understand something, I at least memorize it" (Cronbach's alpha coefficient = .66).

Meta-cognitive abilities: Meta-cognitive abilities were measured by a test consisting of 25 items that asked students to address problems similar to those that they might be faced with in class. Items referred to four elements of meta-cognition, each represented by a different scale: meta-comprehension, verification of one's results, consciousness of the strategies one uses and consciousness of one's own comprehension (Moreno, 2002). In the case of the last dimension of meta-cognition, students were asked to indicate their degree of certainty as they answered questions about an area of concrete understanding (Language, Math, Social Science, and Natural Science). Then, from their scores on an achievement test of that knowledge, it was determined whether or not they actually demonstrated understanding, or learning. Their scores on consciousness of their own understanding referred to the extent to which their level of certainty corresponded to their actual level of competency. A total score was calculated from the participants' scores on the four scales. The Cronbach alpha coefficient of the whole test was .63.

The *students' sociocultural context* was obtained from a *Sociocultural Status Characteristics Questionnaire* about the students' families (Tiana, 2002). At the beginning of the study, the questionnaire was filled out by the students and by their parents, whose results had a correlation of nearly .80. Given that high correlation and the high rate of response among students, the students' responses were the ones used. The questionnaire includes diverse indicators of socioeconomic and sociocultural status that are typically used in large-scale assessment, both national and international: parents' level of education, employment status, access to newspapers, number of books in the home, possession of a computer at home, and several indicators of other possessions. The principal component analysis highlighted the presence of only one factor. The factorial scores were obtained, with weights assigned for a regression method. The scores were transferred to a scale with a mean of 100 and a standard deviation of 15. Using the new scores, mean scores were taken for sociocultural status for each of the participating schools, using the three quarters of the distribution as cut-off points: Level 1 was composed of schools that scored above the 75th percentile, level 2 was composed of schools that scored at or above the 50th percentile and equal or below the 75th, level 3 was comprised of the schools at or above the 25th percentile and below the 50th percentile, and level 4 was comprised of the schools that scored below the 25th percentile.

Academic results. In order to measure the results, specific tests were made for each academic area (Math, Spanish Language and Social Science) and for each year of ESO (Rivière, 2002; González Nieto, 2002; Roca, 2002). The tests were first assessed in pilot studies before being applied in order to eliminate any items that were inadequate from a psychometric perspective. These preliminary analyses of the tests were done through statistics from Classical Test Theory and Item Response Theory, using a binary model with 3 parameters, and the BILOG-MG program (Mislevy & Bock, 1990) for calibration. To verify the single-dimensionality essential to each of the tests before the TRI

models were employed, the TESTFACT program was used (Wilson, Wood, Kandola-Downs & Gibbons, 1991), establishing Lord's criterion (1980) as the criterion of single-dimensionality. This meant that the ratio between the first and second eigenvalues had to be greater than 5. That criterion was met in all of the tests that were ultimately used. On average, each of the tests was made up of approximately 30, multiple choice items. Given the longitudinal nature of the study, as well as the need to compare the results on each subject test over three evaluations (at the beginning of the 1st year of ESO, at the end of the 2nd, and at the end of the 4th) that differed in level of complexity, it was necessary to vertically linked the tests in order to compare students' scores on three tests in each subject (AERA, APA & NCME, 1999). The procedures that were best at test equating in this case were derived from Item Response Theory (Kolen & Brennan, 2004; Hambleton & Swaminathan, 1987), and from the *non-equivalent groups with common, anchoring base items* design. There were common items in 1st and 2nd and in 2nd and 4th. The number of items in common varied according to subject, but on average, it was approximately 8. The common items were selected according to the 3-parameter logistic model and were of average difficulty. In order for a comparison to take place, a process called joint calibration was used for the three evaluations in each subject, using as a reference group, or baseline, the students' results in their 1st year of ESO. Common scores were obtained using a program BILOG-MG. The Cronbach's alpha coefficients for each of the tests are presented in Table 1.

Procedure

The study was conducted over the course of four years, beginning during the 1997-1998 academic year and concluding during the 2000-2001 academic year. During the first year, tests were administered during the month of October to determine the initial level of the students as they began their secondary education, and to identify a base line for analysis. During the remaining years of ESO (2nd and 4th) the tests were administered at the end of May.

All the questionnaires and achievement tests were administered to the students in a classroom during the school day by a test proctor who was specially trained for the task. Each of the three subject tests lasted the duration of a single class period (50 minutes). The testing was done over the course of two days and a break was always given after two hours of testing. The informed consent of participants' parents was sought before their children participated in testing.

Statistical Analyses

Various types of statistical analysis were done. Using an ANOVA repeated-measures design in SPSS 14.0, descriptive statistics were found and the slope of change in each subject area over the three repeated measures was determined. In cases where the matrices of variance-covariance did not meet the sphericity assumption, the Huynh-Feldt correction was applied with corrected degrees of freedom. Although the mean of the test results was established as being 250 with a standard deviation of 50, in the first school year, when baseline scores were taken, that result was not clearly reflected in the data. This is because the joint calibration was done with more subjects than ultimately composed the sample since in this study, when any subject missed an evaluation, their data was eliminated.

In order to study the change that transpired between the three evaluations and determine any growth, a three-level was utilized, separated by academic subject. Level 1 consisted of three evaluations of students, using as an initial value, or 0 value, their results on the evaluation taken during their 1st year of ESO. In level 2, predictor variables specific to the students were used, described above in the Variables and Measurement Instruments section. In level 3, the variables associated with the schools, also described above, were used as predictors. Linear regression models were used where the slopes and intercepts varied randomly. Also, the longitudinal strategy was selected for level 1 in order to measure gains, or growth, since it dealt with comparable and equal scores from item response theory. In other literature on this subject, models that use this type of score are not frequently used. Instead, 2-level models that use the initial evaluation itself as the predictor (see Brown, 2006 or Lissitz, 2005 for a summary) are far more common.

To analyze subjects' progress as well as the predictors, we used a program called HLM 6.0 (Raudenbush, Bryk,

Table 1
Cronbach Alpha Coefficients for the Academic Achievement Test Results

	1st ESO	2nd ESO	4th ESO
Language	.82	.72	.70
Math	.85	.73	.77
Social Science	.73	.73	.70

Note. ESO = Educación Secundaria Obligatoria (Compulsory Secondary Education)

Cheong, & Congdon, 2004). The process used to estimate the parameters was that of *maximum likelihood*. For each analysis, we began by using a nonconditional model with a base line for which the only predictor was the time of the level 1 measurement (Raudenbush & Bryk, 2002) with no covariant variables with time on that level. The times of the measurement were coded as 0, 1 and 2 and no variables used on this level were covariant with time. The intercepts as well as the slopes (rates of change) were considered to vary randomly; in order to predict them, the student variables from level 2 were used. Predictors that were not found to be statistically significant were eliminated and the variance of the coefficients was examined. In level 3, linear regression model with slopes and intercepts varying randomly was followed. We proceeded to first introduce the structural variables of the schools and then the variables associated with their actual functioning into the analysis. The results table displays the final models, after the predictors that did not turn out to be statistically significant had been removed. In these different analyses, we centered the quantitative, predictor variables around the global mean, which is the recommended procedure for an easy interpretation of the results (Hox, 2002).

The degrees of freedom in the analyses of each academic subject differed, since each had a slightly different number of cases.

Results

Table 2 shows the mean scores for each of the three academic subjects, on each of the three measurements.

The Huyhn-Feldt correction demonstrated that there were statistically significant differences between the measured results for all three subject areas: Language: $F(1.74, 1{,}710.12) = 658.13$, $p < .001$, $\eta^2 = 0.51$; Math: $F(1.77, 1{,}698.15) = 1{,}085.90$, $p < .001$, $\eta^2 = 0.53$; Social Science: $F(1.87, 1{,}886.42) = 1{,}282.64$, $p < .001$, $\eta^2 = 0.56$. The Bonferroni corrections we later did revealed statistically significant differences between the measurements in all three subjects as well ($p < .001$); it showed successive increase throughout students' years of ESO education. Before proceeding to apply the multi-level model to measure the change, linear and quadratic contrasts were examined and an appropriate measure for change was determined. In all three subjects, the results showed demonstrated a statistically significant, linear progression ($p < .001$), but the change was not found to be significant when viewed in the quadratic model in any of the subjects. Thus, the linear model was selected. As Figures 1, 2 and 3 show, linear change with a positive slope was standard in all of the schools evaluated, although there was some variation in lines.

In Table 3, the descriptive statistics for the variables used as predictors in the level 2 (students) and level 3 (schools) models are presented.

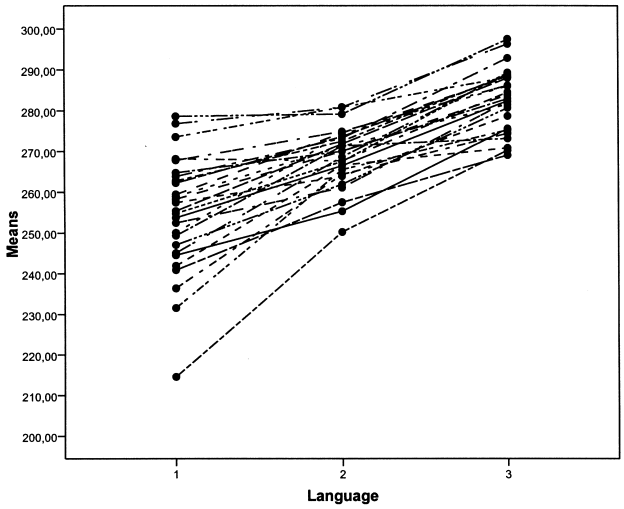

Figure 1. School means for the three measurements of Spanish Language skills.

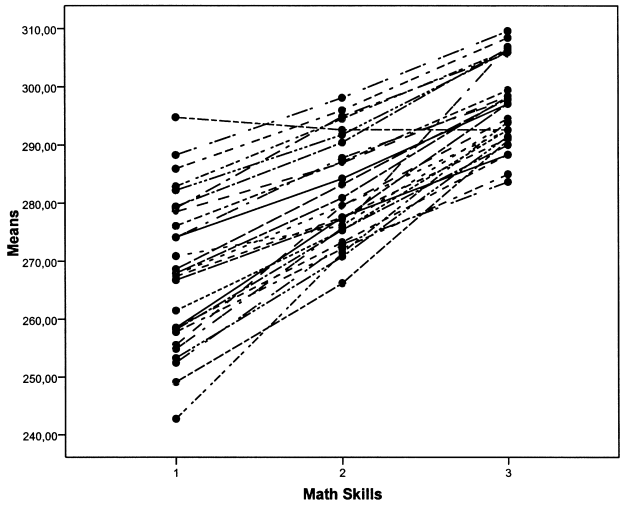

Figure 2. School means for the three measurements of Math skills.

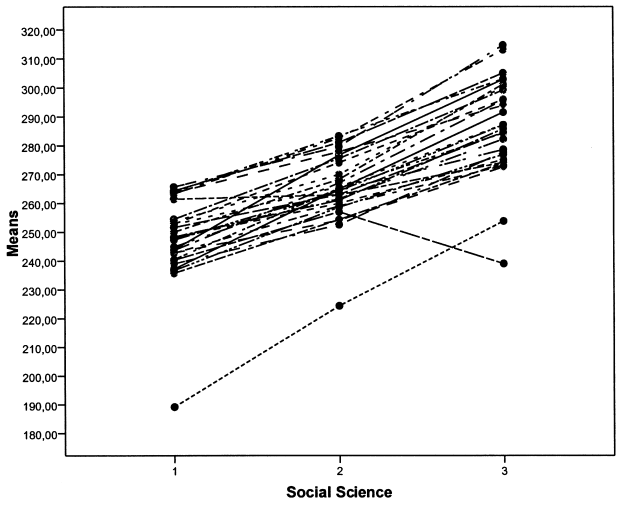

Figure 3. School means for the three measurements of Social Science skills.

Table 2
Descriptive Statistics for the Academic Results

	1st ESO *M* (*SD*)	2nd ESO *M* (*SD*)	4th ESO *M* (*SD*)
Language	257.13 (30.10)	270.77 (18.79)	284.55 (20.78)
Math	269.44 (24.46)	282.80 (17.99)	296.95 (16.07)
Social Science	249.86 (28.54)	267.85 (22.74)	290.05 (25.81)

Note. ESO = Educación Secundaria Obligatoria (Compulsory Secondary Education)

Table 3
Descriptive Statistics for the Predictor Variables in the Models

	M	*SD*
School variables		
Rate of grade repeating	24.76	13.10
Sociocultural context	100.47	9.18
Percentage of female students	49.28	12.62
Leadership of the administration	62.30	5.293
Environment	39.68	4.26
Expectations	43.99	6.76
Student variables		
Sociocultural context	103.09	14.47
Meta-cognitive abilities	11.84	2.27
Strategic understanding	71.32	12.69
Memorization-based learning	42.21	15.34

Results of the multi-level analyses.

Before starting to model the predictor variables for levels 1 and 2, the nonconditional models were estimated for the three subject areas, which included only one level 1 predictor, the time of the measurement (coded as 0, 1 and 2 for the three measurements). This type of nonconditional model was proposed by Raudenbush and Bryk (2002).

Table 4 displays the components of variance and the proportions of the between-schools variance (level 3) for the students' initial status and slope of change results. Those proportions appear in parentheses.

It may be observed that the percentage of variance between schools during the initial measurement was very similar in all three subjects, and was very close to the percentages found in Spanish data in international, transverse studies (about 20%). The percentage of variance between schools in slopes of change differed a great deal according to subject area, and that variance was considerably higher in Social Science than in Math.

Tables 5, 6 and 7 show the results of the final models in all three levels for Spanish Language, Math and Social Science, respectively. Only the predictor variables found to be statistically significant ($p < .05$) or near the conventional levels of significance ($p < .07$) are included.

These results demonstrate that, at the individual level of the students, two variables are predictors of academic achievement in all three subject areas: meta-cognitive abilities and sociocultural status. The variables associated with learning strategies were also shown to be related to academic results. In some cases – Language and Social Science – what was found to be important was the ability to strategically select relevant information and use it to create a plan of action. In others – Language and Math – an inverse relationship was

Table 4
Components of Variance in the Nonconditional Models

	Intercept Level 2	Intercept Level 3	Slope Level 2	Slope Level 3	% variance between schools (Intercept)	% variance between schools (Slope)
Language	497.14	118.32	74.51	11.38	23.80	13.45
Math	371.97	96.28	51.25	4.55	20.56	8.15
Social Science	444.18	131.20	63.01	26.16	22.80	29.30

Table 5
Three-Level Model: Spanish Language

Fixed effects	Coefficient	*SE*	*t*	*df*	*p*-value
Student variables (Level 2)					
Sex	5.31	1.00	5.31	959	.000
Sociocultural context	0.29	0.06	4.78	959	.000
Strategic understanding	0.14	0.04	3.33	959	.000
Memorization-based learning	-0.14	0.03	–4.04	959	.000
Meta-cognitive ability	3.62	0.34	10.56	959	.000
Average initial status (intercept)	253.03	1.36	186.01	23	.000
Leadership of the administration	1.04	0.25	4.06	23	.001
School environment	1.58	0.34	4.66	23	.000
Private versus public school	–7.08	2.87	-2.47	23	.022
Slope of change					
Mean change	14.15	0.55	25.89	22	.000
Expectations	0.25	0.09	2.61	22	.016
Private versus public school	–2.71	1.35	-2.01	22	.057
Rate of grade repeating	0.12	0.03	3.51	22	.002
Level 2 and 3 change interactions					
Sociocultural context (intercept)	–0.10	0.03	–2.76	962	.006
Meta-cognitive ability (intercept)	–0.73	0.19	–3.91	962	.000

Random effects	Components of Variance	*df*	χ^2	*p*-value
Level 2 intercept	395.75	953	3258.54	
Level 2 slope	70.85	956	1643.10	.000
Level 3 intercept	22.05	23	65.31	.000
Level 3 slope	2.85	23	43.63	.004

Table 6
Three-Level Model: Mathematics

Fixed effects	Coefficient	*SE*	*t*	*df*	*p*-value
Student variables (Level 2)					
Sociocultural context	0.16	0.05	3.33	958	.000
Memorization-based learning	–0.04	0.02	–1.90	958	.059
Meta-cognitive ability	3.24	0.28	11.64	958	.000
Mean initial status (intercept)	253.03	1.36	186.01	23	.000
School environment	1.02	0.27	3.72	23	.000
Expectations	0.69	0.13	5.52	23	.000
Slope of change					
Mean change	14.30	0.52	27.57	24	.000
School size	–1.90	0.50	–3.80	24	.001
Level 2 and 3 change interactions					
Meta-cognitive ability (intercept)	–0.94	0.13	–7.47	962	.000

Random effects	Components of Variance	*df*	χ^2	*p*-value
Level 2 intercept	304.61	906	3,855.03	.000
Level 2 slope	46.66	934	1,693.42	.000
Level 3 intercept	38.19	24	93.25	.000
Level 3 slope	4.08	25	64.27	.000

Table 7
Three-Level Model: Social Science

Fixed effects	Coefficient	*SE*	*t*	*df*	*p*-value
Student variables (Level 2)					
Sex	-5.49	1.11	–4.94	959	.000
Sociocultural context	0.22	0.04	5.50	959	.000
Strategic understanding	0.23	0.05	4.24	959	.000
Meta-cognitive ability	2.78	0.30	9.28	959	.000
Mean initial status (intercept)	250.84	1.92	130.33	23	.000
School environment	0.98	0.43	2.28	23	.030
Mean sociocultural context	0.38	0.11	3.38	23	.003
Slope of change					
Mean change	19.99	1.06	18.83	22	.000
Expectations	0.39	0.14	2.69	22	.016
Levels 2 and 3 change interactions					
Memorization-based learning (intercept)	–0.05	0.03	–1.90	25	.067

Random effects	Components of Variance	*df*	χ^2	*p*-value
Level 2 intercept	351.59	978	2815.25	.000
Level 2 slope	6272	953	1496.21	.000
Level 3 intercept	64.43	25	109.46	.000
Level 3 slope	21.31	26	126.59	.000

found between achievement level and learning based on memorization. Along those lines, Language turned out to be the most sensitive to the students' focus when trying to reach an understanding. Last, sex appeared in only two of the three models – Language and Social Science – and with a distinctive importance in each case. Being a female student was found to imply a higher probability of achieving academic success in Language, while in Social Science, that was true for the males.

With respect to the school-level variables, the school environment was part of all three models. Meanwhile, the other level 2 factors appear in only some models. Achievement in Language was found to be associated with efficient leadership on the part of the administration, and with being a private school. Achievement in Math was found to be associated with the school and its faculty having high expectations of their students, and of the school itself. Finally, in Social Science, the mean sociocultural status of the school appears to be a predictor, although that coefficient was very low.

On the other hand, these analyses have allowed us to identify the variables that may be responsible for the differences in learning increases between schools over the course of ESO. On the individual level of the student, once again, meta-cognitive abilities —in Language and Math— and memorization learning—in Social Science—are variables which show a certain predictive capacity, albeit low and an inverse relationship.

The level of progress among the students at each school (average change) is related to the expectations for their achievement in Language and Math. Being a private school and, to a lesser extent, having a greater rate of grade repetition, increases the probability of having a sizeable increase in Language achievement. In Math, it was the size of the school that was found to be related to the average change.

Finally, in Table 8, the reduced proportions of unexplained variance are presented that were obtained by applying the aforementioned models. These proportions were obtained separately for both the intercept—or average, initial

Table 8
Percentage Reduction in Variance Statistics

	Intercept		Slope	
	Between individuals	Between schools	Between individuals	Between schools
Language	20.39	81.36	4.91	74.96
Math	18.11	60.33	8.96	10.33
Social Science	20.85	50.89	0.46	18.54

status —and for the slope of change. The data demonstrate that the differences in how much increase in academic achievement was produced in the students, between schools, is much higher in the subject of Language in comparison to Math and Social Science.

Discussion

What may we conclude about the objectives and the hypotheses of this study? First of all, it confirmed a finding of the majority of the studies reviewed in the Introduction that a greater proportion of variance between schools was related to the variables associated with the students themselves than to the school variables. Nevertheless, the models used in this study have detected a greater percentage of difference between schools than usual.

We will now discuss the school variables that were shown to have a predictive capacity, the most clear of which was the school environment. This result confirms the findings of recent studies that included that measurement (Haark et al, 2005; OECD, 2004). It is a fact of great importance to the field of education, in that it affirms the importance of paying attention to subjects of coexistence. The need to teach in school how to positively relate to and interact with others is justified by the transference of positive, social learning abilities it would bring into society to unfold. These results show that an environment of respect and peace in schools actually results in higher levels of academic achievement.

The sociocultural level of the schools only appears to be a predictor of students' level of learning in Social Science, however, and with only a low coefficient. However, in Language it seems that public versus private status does appear to be a predictor of achievement, which is a characteristic that is widely related to sociocultural status. In private schools, students tend to demonstrate a higher level of achievement in Language when they begin their ESO education than in public schools, and in this study, that relationship was very strong. Those results, in light of the direction of the relationship – higher achievement in schools with a high average sociocultural status and in private schools – confirm the results of the majority of previous studies of achievement. Where our data differ is on the extent of this influence. In this study, the influence was found to be less than normal, even when compared to other data about the Spanish education system. Also, the between-schools variance was lower than in other countries. Perhaps this is due to the small number of schools in the sample and the elevated presence of private schools.

However, although the influence of the schools' sociocultural levels was slight, we would like to reiterate the need for the politics of education to steer itself toward favoring a distribution of students that reduces sociocultural disparity between schools. For the education of students in both public and private school systems to be equal, and to avoid excessively homogenous concentrations of students in the same school, should be priorities of all educational administrations. A commitment should be made on the part of schools to share the responsibility to compensate for social inequalities, and that without falling short of excellence.

The results surrounding the "expectations" variable were quite interesting. It seems that the very fact of having separated this factor from the general, school environment – remember that in the PISA studies, this was included as one more aspect comprising the wider construct of school environment – has allowed us to more clearly appreciate its importance. The expectations that the faculty have of their school and of its students seem to be connected to the initial level of achievement at the start of ESO in Math, and to growth over time in Language and Social Science. This result supports the idea that faculty politics should include ways of augmenting teachers' confidence in their students' capacity to learn. In the case of schools with a low socioeconomic context, this implies completing evaluations of students' growth that highlight the importance of those schools that contribute largely to their students' lives, even if they may receive lower test scores than other schools. A differential treatment of schools that face greater difficulties than others could also help raise the faculty's expectations of their school at large.

Finally, in the models, two school-related variables were identified as not appearing to be predictors for students' academic success at the beginning of their secondary education; rather, they were predictors of progress over the four-year period of ESO. The first refers to the size of the secondary school. In smaller schools, the level of achievement of the students at the beginning of their secondary education was higher in Math. In the other subject areas, it does not seem to have an influence. This result was already asserted in other research (Raywid, 1997; Marchesi & Martín, 2002). This last study demonstrated an interactive relationship between sociocultural context and school size. In low income schools, small school size was associated with higher achievement. We are aware of the difficulties involved in planning for smaller schools in ESO. It becomes more difficult to organize the staff and faculty. That factor, as well as others related to scheduling, usually involves reduced flexibility and fewer options. However, both of these difficulties would be quite possible to resolve if it were accepted that schools faced with the most difficulties should receive more support from the administration and the state. If this alternative method of organization were to be adopted, although it would indeed imply greater spending, according to the results that the present research, it would pay off in terms of achievement.

The second variable associated with progress is the percentage of students who have to repeat grades. The schools with more students repeating grades have a higher level of achievement in Language. How should that data be

interpreted? One possible explanation for this result is that in middle and high schools that are strict in their criteria for students to be promoted a grade, students with poor results during the first year of data collection may have had to repeat a grade, in which case they would no longer be included in the sample.

What about individual variables, those associated with the students? The students' families' sociocultural status appears in all three models. In all cases, the students who belong to families with a high sociocultural status also have high levels of academic achievement. However, surprisingly, that effect is the opposite when its influence is analyzed not in terms of the student's initial level, but in terms of progress over the course of the educational period. For example, in Language achievement, progress was lower in students that came from families with a high sociocultural status. This result may be a consequence of the fact that, although these students have the highest levels of achievement, school teaches them less than it does those students who come from lower sociocultural contexts. Language is one of the aspects on which a student's family context has the most influence, which is consistent with this data.

The variable of sex appears to be linked to two of the academic subjects, and in different ways. Where Language is concerned, the female students obtained better results, while on the contrary, in social science, being male predicted a higher probability of having good academic results. This data confirms a tendency about Language found in other studies on education (OECD, 2001; INECSE, 2001, 2003; Van Damme, De Fraine, Van Landeghem, Opnedakker, 2002). Nevertheless, the data does not coincide with the results of other research on Math achievement; the majority of research has emphasized male students' performance in Math (OECD, 2004, INCE, 1997). However, it is important to remember that in the Basque Country and Castilla and Leon samples of other studies, no differences were found between the sexes in math, which indeed confirms our results. Last, in the case of Social Science, our data did not coincide with that of the INECSE study (2001, 2003) either, where male students obtained significantly lower scores than their female counterparts. One possible explanation for these results could be the type of test used, which emphasized procedural content, not just the concepts, which could perhaps begin to approach difficult questions and forge a closer understanding of what are traditionally considered "areas of science."

Without a doubt the most interesting results of this study have to do with learning strategies and meta-cognitive abilities. As for the first, two of the factors studied in the questionnaire have turned out to be especially strong predictors. The "strategic understanding" factor refers to students' ability to derive the main ideas from information presented and discriminate essential content from accessory content, and also to plan a series of tasks to be done. In two of the three subjects – Language and Social Science – the students that scored highest on that factor obtained higher academic results. It seems, then, that students know themselves well when it comes to how they approach learning, since their responses about their study habits are related to their results. This occurs both when they learn through significant strategies and when they primarily use memorization to learn. The "memorization learning" factor is a predictor of learning in the case of all three subjects; in this case, it logically predicts a low level of academic achievement. This result coincides with the findings of Marchesi & Martínez Arias (2006) in a PISA study that used a Spanish sample.

However, meta-cognitive abilities appear in all three cases and with a high capacity to predict. This may be due to at least two reasons. First, the test used to evaluate these competencies in students may have turned out to be the instrument best fit for the objectives sought. Second, there may truly be abilities associated with students' meta-cognition that are strongly related to significant learning. It is also possible that both of those reasons influence the results. In any case, this data is of great interest, in that it supports the relevance of abilities associated with *learning to learn* to academic success.

It has always been clear that the primary contribution of school should be to teach students to learn on their own, but at this time, in a society of knowing, autonomous learning is the only guarantee we have to complete the necessary objective of learning throughout life. International, educational organizations have insisted for a long time on the need to include among those basic competencies the leaning to learn competency (DeSeCo, 2005 European Commission, 2004). The Organic Law of Education (LOE) has introduced it into the minimum required curriculum, from nursery school all the way through the end of high school (Martín & Moreno, 2007). In addition to that theoretical support, the data found in this study show empirical evidence that corroborate the importance of schools paying this aspect of learning the attention it deserves.

The models demonstrate that having meta-cognitive abilities is a predictor of students' progress over the course of their Compulsory Secondary Education in Language and Social Science, but the present study's data shows that relationship to be inverse. In other words, the students with the highest abilities at the start of ESO exhibited less progress than their classmates. This result seems surprising at first. However, it may have an explanation similar to that offered earlier about the effect of sociocultural status on achievement. These are those students that scored the highest on academic tests when they began their secondary educational stage (keep in mind that this means the presence of the meta-cognitive variable in the first level model, explained above). They certainly progress over the course of their secondary educational stage, but coming from a higher baseline level of achievement, their progress is not as great as the rest of their classmates. One might assert that these students could have learned more than they did.

Alternately, perhaps this result could be interpreted in the same way the PISA results were, as indicating that Spanish students do not achieve a high level of excellence in academic achievement even though they do reach mean levels of achievement. Along those lines, note Marchesi & Martínez Arias's data (2006) about differential progress in Math by students that participated in the PISA study between 2000 and 2003. An improvement was found only in students with low achievement. The students with higher levels did not improve their scores.

The last result we would like to comment upon in these concluding remarks is about the students' level of progress during the secondary education stage. As Figures 1, 2 and 3 show, in all three academic subjects evaluated, clear progress was produced in all the schools that participated in the study. On the other hand, that average progression could be masking certain differences between schools. Some schools may not follow that pattern of progress, even though it may appear as a mean tendency. However, as the figures in the present study show, all schools' data behaves this way, in terms of mean tendency. In all of the middle and high schools analyzed, important progress was produced during the four years of ESO. Nevertheless, what may seem obvious is not so; the image often transmitted of this stage of compulsory, secondary education is one of deterioration in teaching that causes some students to learn very little. The results of this study cast doubt on those assertions. It is true that our research sample was made up of students that have completed the secondary education stage in four years as expected, but that is true of the majority of the student population. According to the Spanish Ministry of Education and Science ([Ministerio de Educación y Ciencia; MEC], 2007) data, 78.4% of students pass on to the next grade in their 1st and 2nd years of ESO, while only 72% pass on to the next grade in their 3rd and 4th years. For that reason, we wish to highlight these results and emphasize the importance of conducting rigorous evaluations in educational centers through longitudinal studies wherein data from cohorts of students may be analyzed.

However, in spite of the interesting contributions implicit in these results, the small number of participating schools and the voluntary nature of their evaluations oblige us to interpret their results prudently and to continue by pursuing studies of this type in representative samples.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (AERA, APA, NCME) (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association, American Psychological Association, National Council on Measurement in Education.

Ayala, C., Martínez Arias, R., & Yuste, C. (2004). *CEAM. Cuestionario de estrategias de aprendizaje y motivación.* Madrid: EOS.

Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzales, E.J., Smith, T.A., & Kelly, D.L. (1996). *Science achievement in the middle school years. IEA's Third International Mathematics and Science Study (TIMSS),* Chestnut Hill, MA: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

Braun, H. (2005). *Using student progress to evaluate teachers: A primer on value-added models.* Princeton, NJ: Educational Testing Service.

Braun, H., Jenkins, F., & Grigg, W. (2006). *Comparing private schools and public schools using hierarchical linear modeling* (NCES 2006-461). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office.

Coleman, J. S., Hoffer, T. B., & Kilgore, S. (1982). *High school achievement: Public, Catholic, and other private schools compared.* New York: Basic Books.

Corten, R., & Dronkers, J. (2006). School achievement of pupils from the lower strata in public, private government-dependent and private government-independent schools: A cross-national test of the Coleman-Hoffer thesis. *Educational Research and Evaluation, 12,* 179-208.

De Fraine, B., Van Damme, J., & Onghena, P. (2002) *The effect of school and classes upon language achievement.* Unpublished manuscript, K.U. Leuven, Secondary and Higher Education Research Center, Leuven, Belgium.

DeSeCo (2005): *The definition and selection of key competencies*. Executive Summary. OECD.

EURYDICE (2002). *Key competencies. A developing concept in general compulsory education.* Retrieved April 25, 2006, at: http://www.eurydice.org/Documents/survey5/en/FrameSet.htm

European Commision (2004). *Competencias clave para un aprendizaje a lo largo de la vida. Un marcode referencia europeo*. Working program «Educación y Formación 2010». Grupo de trabajo B. «Competencias clave». E.C. General Direction for Education and Culture. Retrieved May, 1, 2006, at: http://www.educastur.princast.es/info/calidad/indicadores/doc/comision_europea.pdf

Freiberg, J.H. (1999). *School Climate*. Londres: Falmer Press.

Goldstein, H. (2003). *Multilevel statistical models.* London: Arnold.

Goldstein, J., & Behuniak, P. (2005). Practical assessment, research, and evaluation, *Practical Assessment Research & Evaluation, 11,* 1-17.

González Nieto, L. (2002). El aprendizaje de la lengua. In A. Marchesi & E. Martín (Eds.), *Evaluación de la Educación Secundaria. Fotografía de una etapa polémica* (pp. 179-196). Madrid: Editorial SM.

Haahr, J.H., Nielsen, T.K., Hansen, M.E., & Jakobsen, S.T. (2005). *explaining student performance evidence from the international PISA, TIMSS and PIRLS surveys.* Danish Technological Institute. Retrieved February 18, 2006, at http://ec.europa.eu/education/doc/reports/doc/basicskill.pdf.

Hambleton, R.K., & Swaminathan, H. (1987). *Item response theory: Principles and applications.* Boston, MA: Kluwer.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

INCE (1997). *Resultados de matemáticas. Tercer estudio internacional de matemáticas y ciencias (TIMMS).* Madrid: MEC-INCE

INECSE (2001). *Diagnóstico del sistema educativo. La Evaluación de la educación secundaria 1997.* Madrid: MEC-INECSE

INECSE (2003). *Evaluación de la educación secundaria obligatoria 2000.* Madrid: MEC-INECSE.

INECSE (2004). *Aprender para el mundo del mañana. Resumen de resultados-PISA 2003.* Madrid: MEC-INECSE.

INECSE (2005). *Resultados en España del Estudio Pisa 2000.* Madrid: MEC-INECSE.

Kolen, M.J., & Brennan, R.L. (2004). *Test equating, scaling, and linking: Methods and practices.* New York: Springer.

Lissitz, R.W. (Ed.). (2005). *Longitudinal and value-added modeling of student performance.* Maple Grove, MN: JAM Press.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* New Jersey: Lawrence Erlbaum Associates.

Martín, E. & Moreno, A. (2007). *Competencia de aprender a aprender.* Madrid: Alianza Editorial.

Marchesi, A. & Martín, E. (Eds.) (2002). *Evaluación de la Educación Secundaria. Fotografía de una etapa polémica*. Madrid: Editorial SM.

Marchesi, A. y Martínez Arias, R. (2006). *Escuelas de éxito en España. Sugerencias e interrogantes a partir del Informe PISA, 2003.* Madrid: Fundación Santillana

Marchesi, A., Martínez Arias, R. &Martín, E. (2004). Estudio longitudinal sobre la influencia del nivel sociocultural en el aprendizaje de los alumnos en la Educación Secundaria Obligatoria. *Infancia y Aprendizaje, 27*(3), 307-323.

Ministerio de Educación y Ciencia (2007). Las cifras de la educación en España. Estadísticas e Indicadores. Retrieved: July 25, 2007 at: http://www.mec.es/mecd/jsp/plantilla.jsp?id=3131&area=estadisticas&contenido=/estadisticas/educativas/cee/2006A/cee_2006A.html

Mislevy, R. J., & Bock, D. R. (1990). *BILOG. Ítem analysis and test scoring with binary logistic models.* Chicago: SSI-Scientific Software International.

Moreno, A. (2002). La evaluación de las habilidades metacognitivas. In A. Marchesi & E. Martín (Eds.), *Evaluación de la Educación Secundaria. Fotografía de una etapa polémica* (pp. 119-136). Madrid: Editorial SM.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years.* Wells, UK: Open Books.

Mullis, I.V.S., Martin, M.O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T.A. (IEA) (1998). *Mathematics and science achievement in the final years of secondary school: IEA's Third International Mathematics and Science Report.* Chestnut Hill, MA: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

OECD (2001). *Knowledge and skills for life-first results from PISA 2000.* Paris: OECD.

OECD (200). *Learning for tomorrow's World. First results from PISA 2003.* Paris: OECD

Opdenakker, M. C., Van Damme, J., De Fraile, B., Van Landeghem, G., & Onghena, P. (2002). The effect of school and classes mathematics achievement. *School Effectiveness and School Improvement, 13*, 339-427.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., Bryk, A., Cheong, Y., & Congdon, R. (2004). *HLM6: Hierarchical linear and non-linear modeling.* Homewood, IL: Scientific Software International.

Raywid, M. (1997). Small schools: A reform that works. *Educational leadership*, *55*, 34-39.

Rivière, V. (2002). El aprendizaje de las matemáticas. In A. Marchesi & E. Martín (Eds.), *Evaluación de la Educación Secundaria. Fotografía de una etapa polémica* (pp. 153-177). Madrid: Editorial SM.

Roca, E. (2002). El aprendizaje de las ciencias sociales. In A. Marchesi & E. Martín (Eds.), *Evaluación de la Educación Secundaria. Fotografía de una etapa polémica* (pp. 197-219). Madrid: Editorial SM.

Sellström, E., & Bremberg, S. (2006). Is there a "school effect" on pupil outcomes? A review of multilevel studies. *Journal of Epidemiological Community Health*, *60*,149-155.

Stufflebeam, D. L., & Shinkfield, A. J. (1985). Systematic evaluation: A self-instructional guide to theory and practice. Boston: Kluwer-Nijhoff Publishing [Spanish translation: *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Paidós- MEC, 1987].

Teddlie, Ch., Reynolds, D., & Sammons, P. (2000). Context issues within school effectiveness research. In D. Reynolds & Ch. Teddlie (Eds.), *The international handbook of school effectiveness research* (pp. 134-159). London: Falmer Press.

Tiana (2002) El context sociocultural en la evaluación de los centros educativos. In A. Marchesi & E. Martín (Eds.), *Evaluación de la Educación Secundaria. Fotografía de una etapa polémica* (pp. 61-76). Madrid: Editorial SM.

Van Damme, J., De Fraine, B.; Van Landeghem, G.; Opnedakker, M. (2002). A new study on Educational Effectiveness in Secondary Schools in Flanders: An introduction. *School Effectiveness and School Improvement, 13*(4), 383-397.

Wilson, D.T., Wood, R., Kandola-Downs, P., & Gibbons, R. (1991). *TESTFACT: Test scoring, item statistics, and item factor analysis.* Chicago: SSI-Scientific Software.

Zvoch, K., & Stevens, J. (2006). Successive student cohorts and longitudinal growth models: An investigation of elementary school mathematics performance. *Education Policy Analysis Archives, 14*(2). Retrieved February 16, 2007 at: http://epaa.asu.edu/epaa/v14n2/.

Received July 30, 2007
Revision received May 8, 2008
Accepted June 2, 2008