Co-funded by the Erasmus+ Programme of the European Union

Erasmus+

This project has been funded with support from the European Commission. This publication [The Impact of the European School Improvement Study on Quality and Equity in Education] reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

The Impact of the European School Improvement Study on Quality and Equity in Education

Introduction

Given that children in socially disadvantaged areas are more likely to have low basic skills, the study reported under the Erasmus+ KA2 program "*Promoting quality and equity: a dynamic approach to school improvement" (PROMQE)* sought the extent to which the dynamic approach to school improvement (DASI, Creemers and Kyriakides 2012) could be applied in these schools and could promote both quality and equity in education. This was done by investigating the extent to which schools and teachers were able to reduce the gaps in schooling outcomes among students with differences in their background characteristics, such as SES, gender and ethnicity. The idea behind this is that education can contribute to social justice and democracy by closing the gap in learning outcomes between students with regard to their background, particularly when their abilities and the socio-cultural status of their family (equity) are taken into consideration. As a consequence, this project aimed to develop further the DASI and evaluate its impact on both promoting student achievement gains (quality) and reducing the impact that SES can have on student achievement in mathematics (equity). Therefore, the project was designed mainly to benefit children in socially disadvantaged areas whose schooling had not equipped them with basic skills in mathematics. It was expected that this would be achieved by supporting their teachers and the school management teams to take actions to promote quality and equity. Special attention was also paid to students coming from low socioeconomic backgrounds, who might receive fewer learning opportunities at home and face difficulties meeting the aims of the school curriculum (Strand 2010). Other students who were at risk of dropping out of school might also benefit, since the project was expected to increase teachers' and schools' awareness of how to differentiate instruction and meet the needs of this group of students (Tomlinson 1999). Moreover, parents of these groups of students were expected to benefit indirectly since the proposed theoretical framework, which emerged from the dynamic model, is concerned with the development of partnership policy and parental involvement in order to promote quality and equity. DASI also encourages differentiation both at classroom level and in terms of the functioning of school policy to help school management teams to respond to the special needs of this group of parents who face difficulties in being actively involved in their children's schools (Devine 2013). It is also important to note that DASI was implemented for the first time in socially disadvantaged areas (see Creemers and Kyriakides 2015). Hence, we could discover whether school stakeholders and the advisory and research team (A&R Team) of each country participating in this project faced special challenges in implementing this approach in this group of schools. The project results may help us further develop DASI and identify factors important for promoting quality in this group of schools and/or factors important for promoting equity. Finally, this project aimed to establish stronger links between theory and practice by attempting to develop and evaluate a school improvement strategy which took into account elements of research and improvement. Major elements of this combination were an emphasis on the evidence stemming from theory and research, a need to collect multiple data about the achievement of students and school processes, on one hand, and an emphasis on the context of individual schools and the development and implementation of programs by the schools themselves, on the other. This project can help identify the extent to which each of these elements of DASI is essential for promoting quality and equity in education.

With regard to the countries participating in this project (i.e., Cyprus, England, Greece and Ireland), a recent European project (ESF Research Programme: Establishing a Knowledge Base for Quality in Education: Testing a Dynamic Theory for Education [08-ECRP-012]) provided empirical support for this theoretical model. Cyprus, Greece and Ireland were amongst the countries participating in this project and generated data supporting the validity of the dynamic model (see Kyriakides, Creemers, Panayiotou et al. 2014; Panayiotou, Kyriakides, and Creemers 2016; Panayiotou et al. 2014; Vanlaar et al. 2016). This implies that the framework of the proposed school improvement approach was found to be valid and relevant to the three countries participating in this project. Thus, the project drew on the experiences of these three country teams in developing valid instruments and collecting data on the functioning of school factors included in the framework of DASI; it also gave us the opportunity to develop instruments and generate a methodology for evaluating school policy and the school learning environment in an additional European country (i.e., England). Moreover, DASI was used in a project aiming to support schools in developing strategies and action plans to face and reduce bullying (DAPHNE Research Programme: Designing Evidence-Based Strategies and Actions to Face Bullying by Considering Socio-ethnic Diversities in School Populations and Evaluating their Effects [JLS/2007/DAP-1/226]). Three of the participating countries (i.e., Cyprus, Greece and England) made use of this approach in a substantial number of primary schools (n=115) and the project revealed that their schools improved their school learning environment and, through that, reduced bullying (Kyriakides, Creemers, Muijs, et al. 2014). Thus, the project reported here built on the experiences of these three country teams in using the DASI and sought the extent to which the dynamic approach can be used equally effectively in another country (i.e., Ireland). Finally, all country teams (i.e., Cyprus, Greece, England, and Ireland) have rich experience in school improvement projects and have developed different approaches to working with schools in disadvantaged areas.

Research Aims

The project aimed to measure the impact of the dynamic approach to promoting quality and equity in the participating schools. For this reason, a value-added approach was used by collecting data both at the beginning and at the end of the intervention (see the methods section). Moreover, schools were randomly allocated to two groups and two different types of intervention (the dynamic and the holistic approach) were implemented. By following this design, we were in a position to conduct summative evaluation of the dynamic approach and to seek its impact on:

a) Student achievement gains in mathematics;

b) Reducing the gaps in learning outcomes in mathematics among students with differences in their background characteristics, such as SES, gender, and ethnicity;

c) Improving the school learning environment (SLE);

d) Improving the school policy on teaching.

Specifically, the data were analysed by using multilevel modelling techniques and the impact of each type of intervention on each of these four indicators was examined. Both cross- and within-country analyses were conducted. In this way, we were able to examine whether DASI could be used equally effectively in the participating countries. Moreover, we used multilevel structural equation modelling techniques (as explained in the next section) to find out whether the intervention had an impact on student learning outcomes as a result of improving the school factors (i.e., school policy on teaching and SLE). In this way, not only the impact of DASI, but also its main assumptions, were examined. Thus, the methods of this experimental study are described and the main results concerning the impact of this intervention on promoting quality and equity are presented.

Methods

Participants

At the beginning of school year 2015-2016, a sample of 72 primary schools in four European countries (i.e., Cyprus, England, Greece and Ireland) in socially disadvantaged areas was selected. In each country, participating schools were randomly split into two groups: the experimental (N=36) and the control group (N=36). A pre-measure of achievement in mathematics of all grade 4, 5 and 6 students (n=5560) in the participating schools, and of the functioning of school factors, was also conducted. Table 1 provides some descriptive data on the background factors of students participating in this study and their prior achievement by country. One can see that the t-test did not reveal any statistically significant difference at .05 level between the two groups in relation to two of the background characteristics of their students (i.e., SES, age), and their prior achievement in mathematics. With regard to the other two student background factors (i.e., gender and ethnicity), the chi-square test did not reveal any statistically significant difference at .05 level.

Table 1. Descriptive data about the background factors of the students

	Experimental (N=2899)		Control (N=2661)		T-test		
Student background factors	**Mean**	**S.D.**	**Mean**	**S.D.**	**T**	**Df**	**p**
Prior Achievement	0.76	1.34	0.73	1.30	0.85	5558	0.40
Post Achievement	1.13	1.30	0.91	1.25	6.42	5558	0.0001
SES	0.68	0.65	0.67	0.66	0.57	5558	0.559
Age in days	3765	367	3781	363	-1.63	5558	0.103

Both groups of schools were asked to develop improvement strategies and action plans to promote quality and equity. The next part of this section provides a brief description of the intervention that took place in these schools in order to promote quality and equity. At the end of the school year, student achievement in mathematics and the functioning of school factors were measured.

The treatment offered to the experimental group

At the first stage of the intervention, training and provision of guidelines were offered to the participating primary schools (n=36). In detail, an external seminar for the head teachers of these schools was organised by each country research team at the beginning of the school year 2015-2016 to reach a consensus on the general purpose of the intervention and to inform them about the main phases of the project, the role of the A&R Team, and the role of the school stakeholders. A handbook (see Intellectual Output 3 of this project) was also produced which presented the theoretical framework and provided suggestions for schools on how to build school evaluation mechanisms that aim to improve educational practices at school and classroom level. The handbook also included the rationale of the project and clarified the role of the A&R Team. It was made clear that the A&R Team should provide support for school stakeholders in order to assist them in carefully setting up their own strategies and action plans for promoting both quality and equity. Thus the aim of the handbook (produced by the A&R Team) was mainly to help schools develop and implement their strategies and action plans by providing concrete and specific guidelines for the teachers (the practitioners) and the school management team (principal and deputy heads). Specifically, the A&R Team provided the aims, content, target groups and, most importantly, the activities and actions that schools could carry out in order to promote equity. The handbook also provided clear suggestions on how to build school evaluation mechanisms, including the collection of relevant data, and the use of this information to promote quality and equity, both at the school and classroom levels. The handbook can be accessed from the web page of the project (www.ucy.ac.cy/promqe).

In the next step of the intervention, the A&R Team provided support for the schools to help them establish school evaluation mechanisms. In addition, the A&R Team analysed the data that emerged from the pre-measure (i.e. from the teacher questionnaire) and supplied each school with feedback indicating its priorities for improvement. Specifically, confirmatory factor analysis (CFA), using EQS (Byrne 1994), was conducted for each school factor of the dynamic model to test whether the data fitted a hypothesised measurement model, that is, the assumptions of the dynamic model regarding the measurement dimensions for each school factor. Two sets of CFA were conducted: across countries (i.e. using the full data-set) and within countries (i.e. separate analysis for each country). The results of the cross-country CFA confirmed the construct validity of the questionnaire. Although the scaled chi-square was statistically significant, the values of RMSEA were smaller than 0.05 and the values of CFI were greater than 0.95, thus meeting the criteria for an acceptable level of fit. Moreover, the standardised factor loadings were all positive and moderately high, ranging from 0.63 to 0.84, with most of them higher than 0.72. Thus the analysis of data helped in generating scores for each of the school factors of the dynamic model. Descriptive statistical analysis was then conducted, and each one of the school factors was ordered based on the value of its mean score. In this way, those items (=factors) for which each school performance was less good could be identified. Moreover, the standard deviations for those items which were relatively high (implying that there was less consensus among teachers about the quality of policy in their school) were highlighted.

Finally, Kendall’s W non-parametric test (Kendall and Babington 1939) was applied to rank all the school factors based on their functioning. Kendall’s W test was used to determine whether there was consensus among the teachers’ perceptions with regard to the functioning of the factors. By using this test, it was found that teachers in each school agreed among themselves as to how the factors could be ordered. Then by using the Wilcoxon signed-rank test (Wilcoxon 1945) which factors seemed to perform less well than all the others were identified and each school was able to develop an action plan to improve their functioning.

Subsequently, each country research team visited the schools in the experimental group and participated in staff meetings to announce the results of the teacher questionnaire. In this way, school stakeholders had the chance to discuss the findings of school evaluation and decide whether their action plans would address one or a combination of priorities concerning the factors included in the dynamic model. It was strongly recommended that decisions about their priorities for improvement should not be taken only by the teachers and the school management team: students and parents should also be actively involved in the decision-making process. For this reason, schools were encouraged to establish a committee with representatives of parents, students and teachers to discuss the results and gradually reach a consensus about the priorities of the school and how to address them. The final decision was announced to the whole school community and feedback was provided which helped schools to produce a clear definition of their area requiring improvement.

At the third stage of the intervention, school stakeholders (in cooperation with the A&R Team) developed their strategies and action plans to address specific aspects of the domains on which they were focusing. Readers may see the action plans developed by schools on the web page of the project (i.e., www.ucy.ac.cy/promqe). Table 2 presents the improvement areas addressed by each experimental school according to country. Although one can see that similar areas were addressed by the schools of each country (i.e. in Cyprus 8 out of the 12 schools and in Greece 6 out of the 8 schools were concerned with improving their policy on quality of teaching, whereas in Ireland 5 out of the 8 schools chose to improve their policy on the provision of sufficient learning resources for students and teachers, and in England 6 out the 8 schools chose to improve their policy on student behaviour outside the classroom), one should bear in mind that DASI is based on the assumption that each school should develop strategies and action plans addressing its own needs and after a decision taken by the school stakeholders. DASI is based on the value assumption that authentic change comes primarily from within the organisation and does not expect the centre (e.g., the Ministry of Education) to ask all schools to develop strategies and action plans addressing the same school factor(s).

Table 2. School factors chosen for improvement by each experimental school by country during the school year 2015-2016

School Code	CYPRUS
1	A. Provision of sufficient learning resources for students and teachers B. Quality of teaching C. Quantity of teaching
2	A. Partnership policy
3	A. Quality of teaching B. Student behaviour outside the classroom
4	A. Quantity of teaching B. Quality of teaching
5	A. Partnership policy B. Provision of sufficient learning resources for students and teachers
6	A. Quantity of teaching B. Quality of teaching C. Partnership policy
7	A. Provision of learning opportunities
8	A. Quality of teaching B. Partnership policy
9	A. Partnership policy
10	A. Quality of teaching
11	A. Quality of teaching
12	A. Student behaviour outside the classroom B. Quality of teaching
	GREECE
1	A. Student behaviour outside the classroom B. Collaboration and interaction between teachers C. Partnership policy D. Provision of sufficient learning resources for students and teachers
2	A. Quality of teaching
3	A. Quality of teaching B. Quantity of teaching
4	A. Partnership policy B. Quantity of teaching C. Quality of teaching
5	A. Quality of teaching B. Quantity of teaching
6	A. Quality of teaching B. Quantity of teaching
7	A. Quality of teaching B. Quantity of teaching
8	A. Partnership policy

	IRELAND
1	A. Provision of sufficient learning resources for students and teachers B. Collaboration and interaction between teachers C. Quality of teaching
2	A. Provision of sufficient learning resources to students and teachers B. Collaboration and interaction between teachers
3	A. Quantity of teaching B. Quality of teaching
4	A. Quality of teaching
5	A. Student behaviour outside the classroom B. Provision of sufficient learning resources for students and teachers C. Quantity of teaching
6	A. Provision of sufficient learning resources for students and teachers B. Quantity of teaching C. Quality of teaching
7	A. Student behaviour outside the classroom B. Quantity of teaching C. Quality of teaching
8	A. Student behaviour outside the classroom B. Quantity of teaching C. Provision of sufficient learning resources for students and teachers
	ENGLAND
1	A. Student behaviour outside the classroom B. Provision of learning opportunities
2	A. Student behaviour outside the classroom B. Quantity of teaching
3	A. Quantity of teaching B. Student behaviour outside the classroom
4	A. Student behaviour outside the classroom B. Collaboration and interaction between teachers
5	A. Student behaviour outside the classroom
6	A. Student behaviour outside the classroom
7	A. Quality of teaching B. Quantity of teaching
8	A. Provision of learning opportunities

In developing their action plans, it was explicitly stated by the A&R Team, that the action plan should not only *refer to the activities* that should be taken, but should also indicate *who was supposed to do each activity*, what the *time-schedule* was and what *resources* were needed (see action plan specified in Intellectual Output 3). At this point, the schools were also reminded to make use of the suggestions and additional reading sources provided in the handbook produced by the A&R Team, in order to specify the activities involved in their improvement strategies and action plans. School stakeholders then had to divide the work on developing their action plans by appointing different groups or committees to be responsible for specific areas. At all stages, and especially in developing the action plans, members of the A&R Team provided support for the school stakeholders.

Beyond designing action plans, school stakeholders were further asked to make decisions regarding the monitoring of the implementation of their strategies and action plans. For example, some schools decided that a logbook should be kept by the coordinator of the improvement effort and also by those stakeholders who were responsible for implementing specific aspects of their action plans. School stakeholders were also asked to share their experiences/views with the management team and other stakeholders. If a problem arose in implementing aspects of the action plans, school stakeholders (in cooperation with the A&R Team) had to make improvements to their action plans and/or provide support for those stakeholders not in a position to implement particular tasks of the action plans.

In the next stage, the intervention was implemented. The implementation of DASI lasted for approximately eight months and the A&R Team provided support for the school stakeholders by helping them overcome difficulties and problems that emerged during the implementation of their action plans. Specifically, the country teams visited the schools in the experimental group once every six weeks to provide feedback and support in the implementation and/or in redesigning the action plans. A network within and across countries of schools addressing the same factors was also developed in order to share experiences during the implementation of their school improvement strategies. Moreover, the A&R Team helped school stakeholders to use the evaluation data in order to modify their strategies and action plans, according to the circumstances and specific needs of different groups of the school population. The appropriate modification of action plans was found to reduce the chance of a school discovering too late that no progress had been made during the school year due to the poor implementation of its action plans.

Support offered to the schools of the control group

In order to evaluate the impact of DASI, the A&R Team gave feedback to a control group of primary schools (n=36) regarding the results that emerged from a pre-measure concerned with student achievement in mathematics and the functioning of their school (but without mentioning what their improvement priorities were) as revealed by teacher responses to the questionnaire measuring school factors. These schools were also offered support to develop their own strategies and actions to promote quality and equity, but without using the DASI. Support was, therefore, provided by the A&R Team and the same amount of time and effort was allocated to each treatment group. In addition, accounts of the effort that schools put into the project were collected but we did not identify any statistically significant difference between the two groups. By following this approach, we were able to provide equal support for each group and to control for the Hawthorne effect in two ways: all experimental groups put the same amount of effort into their specific intervention and schools in each group were not aware of the other interventions thus avoiding compensatory rivalry or resentful demoralisation on the part of either of the two groups (Shadish, Cook, and Campbell 2002). This group of schools was therefore treated as a control group and data were collected in order to compare the impact of DASI on promoting quality and equity. The next part of this section refers to the methods used to measure the main variables of this study.

Measures

Student achievement in mathematics at the beginning and at the end of the intervention: For each year group of students, criterion-reference tests in mathematics were constructed in order to measure their knowledge and skills in mathematics in relation to the objectives of the national curriculum in these four European countries (see Intellectual Output 2). Specifically, each country team analysed its own mathematics curriculum and developed a specification table covering the basic skills in mathematics expected to be taught to students in grades 3 to 6. The core-team compared the tables and developed a common one that addressed aspects of numeracy covered by all countries. In addition, each country team collected instruments measuring the constructs mentioned in the common specification table, and by comparing the items used in these instruments, we managed to develop our own battery of written tests. Each country team had to translate the items and develop its own version. To test the construct validity of this battery of tests, a validation study was conducted. Specifically, the written tests were administered to all grade 3, 4, 5 and 6 students of 20 primary schools (i.e., 5 per participating country) at the end of school year 2014-2015. At this point, it is important to mention that the test developed to measure the achievement of grade 4 students at the beginning of the school year was administered to grade 3 students who were at that time at the end of their school year. Data were then analysed by using the Extended Logistic Model of Rasch (Andrich 1988), which revealed that each scale had satisfactory psychometric properties. Based on the results of the validation study, we also made some minor amendments to the written tests and to the correction schemes developed by our team. The final version of the battery of written tests was used to collect data at the beginning and at the end of the intervention.

The written tests administered during the main study were also subject to control for reliability and validity. It is important to note that none of the respondents achieved a full score, and none scored zero. Moreover, less than 3 per cent of the students achieved over 80 per cent of the maximum score, and less than 9 per cent of the students achieved over 70 per cent of the maximum score. Based on the range of the results, the ceiling and floor effects in the attainment data were not observed.

Equating of tests. The test administered to Grade 6 students when they were at the end of the school year was obviously more difficult than that administered to Grade 4 students when they were at the beginning of the school year. Prior to making comparison of test scores meaningful, the scores had to be made comparable. Equating was done using Item Response Theory (IRT) modelling (Hambleton and Swaminathan 1985). The method of equating followed the same procedure as that used in PISA studies. However, in PISA, equating is horizontal (equating the different versions of tests), whereas in this study the equating was vertical. Specifically, the scores were transformed into the same scale on the basis of characteristics of IRT models with students' latent level of ability (y) and difficulty level of an item (b) being identical when certain preconditions were fulfilled (Bond and Fox 2001). The latent ability level for each student could be determined in every version as long as there were so-called 'anchoring items' connecting the versions. For the purposes of this study, we used enough common items (i.e., approximately 15 per cent of anchoring items across the tests) with representative content to be measured (Kolen and Brennan 1995). Estimation was made using Rasch's the Extended Logistic Model (Andrich 1988), which revealed that each scale had satisfactory psychometric properties. Thus, for each assessment period, achievement in mathematics was estimated by calculating the Rasch person estimates.

Student background factors: Information was collected on four student background factors: gender (0=boys, 1=girls), ethnicity, language spoken at home (0=other language, 1=language of instruction at school) and SES. Five SES variables were available: father's and mother's education level, the social status of father's job, the social status of mother's job and the main elements of the home learning environment. Specifically, it was possible to classify parents' occupation into three groups: occupations followed by the working class (63 per cent), occupations followed by the middle class (28 per cent), and occupations followed by the upper-middle class (9 per cent). The student questionnaire (see this questionnaire after the Mathematics test in the Intellectual Output 2) was also concerned with the main elements of the home learning environment (i.e., learning materials available at home and learning opportunities offered at home). With regard to the home learning materials, students were asked whether they had newspapers at home, a car and/or a second car, a lawnmower, a room of their own, musical instruments, computer, access to the Internet, and their own desk where they could study and do their homework. They were also asked to indicate how many books they had at home. Furthermore, in the final part of the questionnaire, students had to provide information about whether their parents/guardians knew their classmates' and friends' parents/guardians and also, by completing a Likert scale, how often specific activities took place involving themselves and their family (i.e. seeing a play, visiting a museum or art gallery, attending a popular music concert, and going to a public reading with their family).

Rasch's Extended Logistic Model (Andrich 1988) was used to analyse the ordinal data that emerged from the questionnaire. Thus a scale which referred to the student SES score was created and analysed for reliability, fit with the model, meaning and validity. Analysis of the data revealed that the scale had relatively satisfactory psychometric properties. Specifically, the indices of cases (i.e. students) and item separation were higher than 0.87, indicating that the separability of the scale was satisfactory (Wright 1985). Furthermore, the infit mean squares and the outfit mean squares of each scale were near one and the values of the infit t-scores and the outfit t-scores were approximately zero. The analysis revealed that there was a good fit with the model (Keeves and Alagumalai 1999). Thus a score for the SES of each student was calculated using the relevant Rasch person estimate in each scale.

Using a teacher questionnaire to measure school factors. The explanatory variables which refer to the school level factors of the dynamic model were measured by asking the teachers in the school sample to complete a questionnaire. The questionnaire (see Intellectual Output 1) was designed in such a way that information about the five dimensions of the school-level factors of the dynamic model could be collected, and a Likert scale was used to collect data on teachers' perceptions of the school-level factors. Since it is expected that teachers within a school will view the policy of their school and the evaluation mechanisms of their school similarly, but differently from teachers in other schools, a generalisability study was initially conducted. For each participating country, it was found that in all the questionnaire items, the object of measurement was the school. Reliability was then computed for each of the dimensions of the school factors by calculating multilevel λ (Snijders and Bosker 1999) and the Cronbach alpha for data aggregated at the school level. The value of the Cronbach alpha represented consistency across items, whereas multilevel λ represented consistency across groups of teachers. The results are presented in Table 3. We can observe that for all factors and their dimensions the reliability coefficients were high (around .80).

Using the Mplus (Muthén and Muthén 1999) software, the intra-class correlations (ICC) of the scales were computed. The ICC, which indicate what amount of variance in the teacher questionnaire is located at the between level, are also illustrated in Table 3. We can observe that the percentages of variance at the between level (school level) were between 29 and 38. These percentages are rather high compared to the results of other instruments that measure perceptions of people or objects in clustered or interdependent situations (den Brok, Brekelmans, Levy, and Wubbels 2002).

Table 3. Cronbach Alpha (Reliability), Multilevel λ (Consistency), and Intra-Class Correlations (ICC) of scales emerging from teacher questionnaire concerned with each school factor at the school level

School Factors	**Cronbach Alpha (Reliability)**	**Multilevel λ (Consistency)**	**Intra-Class Correlations (ICC)**
School policy for teaching			
Quantity of teaching	0.83	0.82	0.35
Provision of learning opportunities	0.82	0.84	0.33
Quality of teaching	0.81	0.82	0.36
Policy on the school learning environment (SLE)			
Student behaviour outside the classroom	0.87	0.88	0.29
Collaboration and interaction between teachers	0.84	0.83	0.31
Partnership policy	0.83	0.86	0.38
Provision of resources	0.85	0.87	0.35
School evaluation			
Evaluation of school policy for teaching	0.82	0.86	0.35
Evaluation of the SLE	0.84	0.83	0.33

To test the validity of the questionnaire, separate SEM analyses were conducted for each of the three overarching factors: a) *school policy on teaching*, b) *policy on the school learning environment*, and c) *school evaluation*. The first overarching factor was *school policy on teaching* and consisted of factors measuring the quantity of teaching, provision of learning opportunities and quality of teaching (see Table 3). The second overarching factor was *policy on the school learning environment* and comprised factors measuring student behaviour outside the classroom, teacher collaboration, partnership policy and provision of sufficient learning resources. Finally, the third overarching factor was *policy on evaluation* and was composed of factors measuring evaluation of the school policy on teaching and evaluation of the learning environment. Additionally, for each of the three overarching factors another model was tested in order to compare its fit with the data from the three proposed theoretical models (i.e., model 1). In these alternative models (model 2) all the items that were used for the SEM analyses of each of the three overarching factors were considered to belong to a single factor. These models were an attempt to test whether there was a social desirability factor involved in the questionnaire items. If the one-factor models (models 2) had been found to fit with the data, this would have raised doubts about whether we could have scores for each factor separately. The fit indices of each model per school factor are shown in Table 4, where it can be seen that model 1 was found to best fit the data and that in each case (i.e., overarching factor) the fit indices of this model were satisfactory.

Table 4. Fit indices of the models that emerged from the SEM analyses of the teacher questionnaire used to measure each overarching school factor

Models	**X^2**	**Df**	**X^2/df**	**p**	**CFI**	**RMSEA**	**Range RMSEA**
School policy on teaching							
Model 1	140	16	8.75	0.001	0.992	0.051	0.045 – 0.058
Model 2 (one factor model)	493	20	24.7	0.001	0.941	0.093	0.085 – 0.099
Policy on the school learning environment							
Model 1	679	96	7.1	0.001	0.967	0.052	0.045 – 0.063
Model 2 (one factor model)	3888	135	28.8	0.001	0.738	0.099	0.096 – 0.107
School evaluation							
Model 1	544	57	9.54	0.001	0.969	0.056	0.048 – 0.060
Model 2 (one factor model)	1545	65	23.8	0.001	0.895	0.093	0.089 – 0.096

Having established the validity and reliability of the relevant measures, it was decided to generate factor scores by taking into account teacher responses to the relevant questionnaire items. For each school, separate analysis of the teacher responses to the questionnaire items was conducted, and those factors which had the lowest mean rank values were identified. The results were reported to each school (readers can find the format of the report given to schools in Appendix), and stakeholders in the experimental group were encouraged to develop their strategies and action plans in order to improve the functioning of those factors for which lower mean rank values were estimated.

A similar approach was used in analysing teachers' responses to the questionnaire at the end of the intervention. The reports sent to the schools at the end of the intervention made suggestions regarding the improvement areas that each school could consider in developing its own strategies and action plans during the next school year (i.e., 2016-2017). It was also possible to classify the experimental schools into three groups: a) those that had managed to improve with respect to all the factors addressed during the intervention and therefore new factors had to be addressed in the following year, b) schools which had managed to improve some of the factors but needed to show further improvement regarding some other factors addressed during the intervention and c) schools which had to work on improving further all the factors addressed during the intervention. Therefore three different types of report were sent to the schools in the experimental group. At this point, it is important to note that all experimental schools in Cyprus, Greece and Ireland (28 out of 36) asked from the research team to work together with them on using DASI to promote quality and equity for another school year. Although this was not part of the original research design, the research team continued to support these schools for an extra school year as requested. Since data from all schools participating in this project have been collected, the sustainability of DASI and the effects of offering DASI to schools for a period of more than one school year will be examined. However, results from this follow-up phase of the study are not reported here since at the time of writing, this phase had not been completed.

Findings

Since the experimental study reported here attempted to support schools in developing strategies and action plans in order to improve the functioning of specific school factors and, through that, promote quality and equity, this section is split into four parts, in which the main assumptions of this intervention are tested. The first part refers to the impact of the intervention upon the functioning school factors since these factors were directly addressed through the action plans developed by the experimental groups. The second part investigates the extent to which the intervention had an impact on student achievement gains, and the results of a multilevel regression analysis are presented. The impact of this intervention on promoting equity is examined in the third part of this section. We finally use multilevel SEM techniques to investigate the extent to which the impact of this intervention on student achievement gains resulted from improving the functioning of school factors, and in this way, we test the main assumption of DASI upon which this study is based.

A) The impact of intervention on improving school factors

Table 5 presents the means and standard deviations of the three school-level overarching factor scores before the implementation of the intervention and at the end of the intervention in the experimental and control schools across the four countries participating in this study. Although a repeated measures MANOVA of treatment (following DASI/not following the proposed approach) according to time {before (i.e. pre) /end (i.e. post)} could have been carried out with the three factor scores (i.e., policy on teaching, school learning environment, and school evaluation) as dependent variables, we decided to compare the school factor scores of these two groups by using non-parametric statistical tests due to our small sample size at the school level (i.e., 36 schools in each group). Siegel and Castellan (1988) argue that when the sample size is small, non-parametric tests are preferable to parametric tests, even when interval data have been collected. In addition, Dixon (1954) claims that when compared with the t-test, the Kolmogorov-Smirnov test has high-power efficiency (about 96 per cent) for small samples. Thus, the Kolmogorov-Smirnov Two-Sample Test was initially employed to identify any statistically significant difference between the two groups in terms of the functioning of the school factors before the intervention.

Table 5. Means and standard deviations of the functioning of each overarching school factor in the experimental and control schools and values of the Kolmogorov-Smirnov Two-Sample Test

	Experimental School		**Control School**			
Overarching school factor	**Mean**	**S.D.**	**Mean**	**S.D.**	**K-S Z**	**p**
A) Before the intervention						
School policy on teaching						
Quantity of teaching	3.03	0.64	3.05	0.67	0.823	0.507
Provision of learning opportunities	3.08	0.75	3.02	0.70	0.774	0.587
Quality of teaching	2.98	0.78	3.03	0.74	0.799	0.547
School learning environment						
Student behaviour	3.02	0.68	3.06	0.73	0.840	0.481
Teacher collaboration	3.01	0.75	3.07	0.76	0.832	0.493
Partnership policy	3.10	0.69	3.05	0.71	0.766	0.601
Provision of resources	3.01	0.78	2.99	0.72	0.656	0.782
School evaluation						
Evaluation of teaching policy	2.81	0.84	2.77	0.83	0.914	0.874
Evaluation of the school learning environment	2.78	0.85	2.74	0.79	0.963	0.312
B) At the end of intervention						
School policy on teaching						
Quantity of teaching	3.48	0.63	3.04	0.68	1.47	0.03
Provision of learning opportunities	3.80	0.64	3.10	0.66	1.99	0.01
Quality of teaching	3.40	0.78	2.98	0.68	1.46	0.03
School learning environment						
Student behaviour	3.45	0.77	3.09	0.61	1.39	0.04
Teacher collaboration	3.69	0.85	3.10	0.64	1.76	0.02
Partnership policy	3.55	0.84	3.06	0.63	1.62	0.03
Provision of resources	3.65	0.77	3.08	0.69	1.71	0.02
School evaluation						
Evaluation of teaching policy	3.20	0.74	2.79	0.74	1.41	0.03
Evaluation of the school learning environment	3.19	0.75	2.83	0.81	1.28	0.06

No statistically significant difference was identified at the .05 level. This implies that the two groups were performing equally well in relation to the functioning of the three overarching school factors. However, at the end of the intervention, the Kolmogorov-Smirnov Two-Sample Test revealed statistically significant differences at the .05 level between these two groups of schools in relation to each school factor (see Table 5). Moreover, the Wilcoxon Test was used to identify whether there was any statistically significant progress in the performance of each group of schools in relation to the three overarching school factors. Only the schools in the experimental group managed to improve the functioning of their school factors at a statistically significant level.

B) The impact of the intervention on improving student achievement gains in mathematics

Multilevel analysis of mathematical achievement was conducted in order to evaluate the impact of DASI on promoting quality. Since the number of countries involved in this project was small, it was decided to model the country effects by adding relevant dummy variables to the empty model (treating Cyprus as the reference group) and not to consider the country as an extra level of data. Thus the first step in the analysis was to determine which levels should be included to reflect the hierarchical structure of the data. Empty models with all possible combinations of the levels of analysis (i.e., student, class and school) were established and the likelihood statistics of each model were compared (Snijders and Bosker 1999). An empty model consisting of student, class and school levels represented the best solution. It was found that 60.2 per cent of the total variance was situated at the student level, 23.5 per cent, at the classroom level and 16.3 per cent, at the school level. In subsequent steps, explanatory variables at different levels were added, starting at the student level. Explanatory variables, apart from grouping variables, were centred as Z-scores with a mean of 0 and a standard deviation of 1. Grouping variables were entered as dummies with one of the groups as the baseline (e.g., boys = 0). The models presented in Table 6 were estimated without the variables that had no statistically significant effect at level 0.05. In model 1 the context variables at each level (i.e., prior achievement, gender, SES, age and ethnicity) were added to the empty model. The following observations arose from the figures in the third column of Table 6. Firstly, model 1 explained 36.9 per cent of the total variance, and most of the explained variance was at the student level. The likelihood statistic revealed a statistically significant change between the empty model and model 1 ($p < 0.001$). Secondly, the effects of all contextual factors, other than ethnicity and age, were significant. Prior knowledge in mathematics was found to have the strongest effect in predicting student achievement at the end of the school year.

Table 6. Parameter estimates and standard errors for the analysis of mathematics achievement (students within classes, within schools)

Factors	**Model 0**	**Model 1**	**Model 2**
Fixed part			
Intercept	0.77 (.04)*	0.73 (.07)*	0.69 (.05)*
Student level			
Prior achievement		0.57 (.01)*	0.57 (.01)*
Gender (0=boy, 1=girl)		-0.07 (.02)*	-0.07 (.03)*
SES		0.18 (.02)*	0.18 (.02)*
Age		0.05 (.04)	
Ethnicity (0=other, 1=immigrant background)		-0.06 (.07)	
Class level			
Average prior achievement		0.08 (.04)	
Percentage of girls		0.03 (.04)	
Average SES		0.05 (.03)	
Average age		0.04 (.05)	
Percentage of students with immigrant background		-0.03 (.04)	
School level			
<u>Context</u>			
Average prior achievement		0.28 (.09) *	0.21 (.09) *
Percentage of girls		0.03 (.05)	
Average SES		0.04 (.07)	
Average age		0.03 (.03)	
Percentage of students with immigrant background		-0.02 (.04)	
<u>*Countries*</u>			
Greece		-0.11 (.11)	-0.10 (.11)
England		0.29 (.11) *	0.28 (.12) *
Ireland		-0.13 (.09)	-0.12 (.08)
DASI (0=control, 1=experimental)			0.19 (.05) *
Variance components			
School	16.3 %	10.5 %	6.1 %
Class	23.5 %	17.3 %	13.2 %
Student	60.2 %	35.3 %	35.1 %
Explained		36.9 %	45.6 %
Significant test			
X^2	17936	13794	13768
Reduction		4142	26
Degrees of freedom		5**	1
p-value		.000	.000

* Statistically significant effect at .05 level

** The models presented in this table were estimated without the variables that did not have a statistically

Moreover, prior knowledge was the only contextual variable which had a significant effect on student achievement when aggregated at the school level. Thirdly, England was found to have better results in mathematics, whereas no statistically significant difference between the other three countries was found. However, we should not draw any conclusions about country effects, especially since a purposive sampling procedure was used and only schools in socially disadvantaged areas were selected. In model 2 the impact of DASI was tested by adding to model 1 a dummy variable. By considering the control group as a reference group, it was found that the group which made use of DASI managed to achieve better results than the control group. But although the effect of DASI on achievement was found to be statistically significant at 0.05 level, its effect size was small (i.e., d=0.31). However, by comparing it with other interventions in education, it could be considered as very satisfactory, given that most interventions in education have much smaller effects on student learning outcomes (see Creemers and Kyriakides 2015).

This finding seems to provide support for the use of DASI in socially disadvantaged schools to promote quality in education. To test this argument further, separate within-country analyses were also conducted. In each country, it was found that the parameter estimate of the dummy variable DASI was statistically significant, which implies that schools which made use of DASI managed to achieve better learning outcomes in mathematics than schools in the control group (after controlling for the effects of all contextual factors). Table 7 summarises the main results concerning the impact of DASI that emerged from each within-country analysis. The fixed effects obtained from multilevel analysis can readily be converted into standardised effects or 'Cohen's d' by dividing them by the standard deviations in the 'treatment group' which made use of the DASI to promote student learning outcomes in mathematics. Thus Table 7 presents the effect sizes of using DASI according to country and we can observe some differences in the reported effect sizes. The impact of DASI in some countries (i.e., Greece and Ireland) was bigger, whereas in England its impact was smaller. These differences could be attributed to differences between countries in terms of the support that schools in socially disadvantaged areas received from the system level to promote quality. Differences in the extent to which the school factors of the dynamic model are addressed by national policies and the relevant school evaluation mechanisms should also be considered. For example, in Cyprus, the Ministry of Education refers explicitly to the school factors of the dynamic model and expects headteachers to develop action plans in order to improve the functioning of school factors, and this might have had an effect on schools in the control group.

Table 7. Effect of using the DASI approach on student achievement gains in mathematics

Country	**Effect**	**Pooled *SD***	**Cohen's *d***
Cyprus	0.24	0.73	0.33
Greece	0.28	0.67	0.42
England	0.16	0.71	0.23
Ireland	0.32	0.84	0.38
Across countries	0.19	0.62	0.31

C) The impact of intervention on equity

To investigate the impact of DASI on promoting equity, two separate multilevel analyses were conducted for each group of schools to identify the direct impact of SES on prior and final student achievement in mathematics. In this way, it was possible to discover whether the direct effect of SES on achievement became smaller in the experimental and/or in the control group. The final multilevel models that emerged from each analysis (after controlling for all contextual factors) are presented in Table 8.

The results of the two separate multilevel analyses of student achievement at the beginning of the intervention are presented in the second and fourth columns of the table. Each analysis revealed that two background factors (i.e., SES and gender) were associated with the achievement of each group of students at the beginning of the intervention. The results of the two analyses were exactly the same in terms of the background factors found to be associated with achievement at the beginning of the intervention of the two groups of students and were almost the same in terms of the percentage of explained variance of student achievement for each group of students (i.e. 19.6 per cent and 17.8 per cent).

Table 8. Parameter estimates and standard errors for the analysis of the impact of SES on student achievement in mathematics (students within classrooms within schools)

Factors	**Experimental Group**		**Control Group**	
	Pre-measure	**Post-measure**	**Pre-measure**	**Post-measure**
Fixed part				
Intercept	0.70 (.12) *	0.73 (.07) *	0.67 (.11) *	0.74 (.07) *
Student level				
Prior achievement	NA**	0.55 (.01) *	NA**	0.59 (.02) *
Gender (0=boy, 1=girl)	-0.08 (.03) *	-0.07 (.03) *	-0.09 (.04) *	-0.08 (.03) *
SES	0.30 (.01) *	0.11 (.02) *	0.28 (.01) *	0.23 (.02) *
Age	0.02 (.04)	0.03 (.04)	0.02 (.04)	0.04 (.04)
Ethnicity (0=other, 1=immigrant background)	-0.02 (.03)	-0.01 (.03)	-0.02 (.04)	-0.03 (.04)
Class level				
Average prior achievement	NA**	0.09 (.04) *	NA**	0.08 (.03) *
Percentage of girls	0.05 (.04)	0.04 (.04)	0.05 (.05)	0.05 (.04)
Average SES	0.03 (.03)	0.02 (.03)	0.03 (.03)	0.04 (.03)
Average age	0.04 (.05)	0.03 (.05)	0.04 (.05)	0.04 (.04)
Percentage of students with immigrant background	-0.03 (.04)	-0.03 (.04)	-0.04 (.04)	-0.06 (.04)
School level				
Average prior achievement	NA**	0.24 (.08) *	NA**	0.18 (.08) *
Percentage of girls	0.02 (.03)	0.02 (.03)	0.03 (.03)	0.04 (.03)
Average SES	0.02 (.03)	0.02 (.03)	0.02 (.03)	0.03 (.03)
Average age	0.03 (.02)	0.03 (.02)	0.03 (.02)	0.02 (.02)
Percentage of students with immigrant background	-0.02 (.04)	-0.02 (.03)	-0.02 (.04)	-0.01 (.03)
<u>Countries</u>				
Greece	-0.13 (.12)	-0.14 (.10)	-0.14 (.11)	-0.11 (.09)
England	-0.05 (.11)	0.29 (.10) *	-0.04 (.10)	0.28 (.10) *
Ireland	0.21 (.13)	-0.14 (.08)	0.15 (.12)	-0.15 (.09)
Variance components				
School	13.2 %	10.2 %	14.8 %	10.9 %
Class	24.7 %	17.1 %	25.6 %	17.5 %
Student	42.5 %	35.0 %	41.8 %	34.1 %
Explained	19.6 %	37.7 %	17.8 %	37.5 %
Significant test				
X^2	11878	13892	10098	12144
Reduction	231.9	571.9	182.4	504.2
Degrees of freedom***	2	6	2	6
p-value	.001	.001	.001	.001

* Statistically significant effect at .05 level

** There was no measure of prior achievement that could be used in analysing student achievement at the beginning of the intervention.

*** The models presented in this table were estimated without the variables that did not have a statistically significant effect at 0.05 level.

In order to estimate the relative importance of the SES on student achievement at the beginning of the intervention for each group of students, the fixed effect obtained from each multilevel analysis was converted to standardised effects or 'Cohen's d' by following the approach proposed by Elliot and Sammons (2004). By using this approach, it was found that the effect size was equally high for each group of students (i.e., the experimental group d = 0.41 and the control group d = 0.42). The results that emerged from analysing the impact of student background factors on achievement at the end of the intervention for each group of students are presented in the third and fifth columns of Table 8. For each group of students, achievement at the end of the intervention was found to be associated with all student background factors apart from ethnicity (i.e., prior achievement, SES and gender). Moreover, the findings of both analyses show that prior achievement was the only background factor associated with achievement at the end of the intervention when aggregated either at the level of class or school. Thus the two separate analyses of achievement at the end of the intervention reveal exactly the same results in terms of the background factors found to be associated with achievement and the percentage of explained variance of achievement at the end of the intervention. Finally, with regard to the direct effect of SES on student achievement at the end of the intervention, using the approach described above, it was found that, for students in the control group, the effect of SES was greater (d = 0.39) than for those in the experimental group (d = 0.28). It should, however, be acknowledged that this finding reveals that the direct effect of SES was substantially reduced in the schools in the experimental group. Further analysis was therefore conducted to measure the total effect of SES in each group of schools by taking into account the fact that SES has an indirect effect on final achievement through its impact on prior achievement. Table 8 shows that prior achievement has almost the same effect on final achievement these two groups {i.e., experimental group (d=0.39) and control group (d=0.34)}. Earlier it was pointed out that the effect of SES on prior achievement was almost the same in the experimental and control groups. As a consequence, differences in the total effect of SES between these two groups were also found {i.e., experimental group (d=0.42) and control group (d=0.52)} and revealed that schools which made use of DASI not only managed to reduce the direct effect of SES on final achievement in mathematics, but also had smaller total effects of SES on achievement by the end of the intervention. However, by comparing the effect of SES on student achievement at the beginning of the intervention with the total effect of SES on achievement at the end of the intervention, one can see that almost neither a reduction nor an increase was observed in the schools of the experimental group. Nevertheless, one should not underestimate the impact of DASI on promoting equity since the effect of SES in the control schools increased substantially, and this finding is in line with the results of longitudinal studies conducted in different countries, which reveal that the total effect of SES gradually increases over time (Hansen, Rosén, and Gustafsson 2011; Sammons 2008).

D) Searching for direct and indirect effects of DASI on promoting student learning outcomes

In the final part of this section, the main theoretical assumption of DASI is tested by searching for both the direct and indirect effects of this intervention on student achievement in mathematics. Specifically, DASI is based on the assumption that improving the functioning of school factors can aid schools in becoming more effective. Thus the following three conceptual models were used to test the effect of the intervention on student achievement at the end the intervention: *(1) the direct effect model*, *(2) the indirect effect model*, and *(3) the direct and indirect effect model.* In the first model, we assumed that the intervention would have only direct effects upon final achievement and upon improvement in the functioning of school factors. It was also assumed that only student background factors (i.e., gender, SES, ethnicity) and prior achievement would have direct effects on students' final achievement in mathematics (see figure 1). In the second model (see figure 2), we did not expect any direct effect of the intervention on final achievement. Only the existence of an indirect effect of DASI on final achievement was anticipated since DASI was expected to have an effect on school factors which were thought likely to affect final student achievement together with the background factors (including the prior achievement). Finally, the third model is based on the assumption that the intervention would have both direct and indirect effects on student achievement in mathematics at the end of the intervention (see Figure 3). For the purposes of this analysis, MPlus (Muthén and Muthén 2001) was used to test the three conceptual models and identify which of the three models fitted our data most accurately.

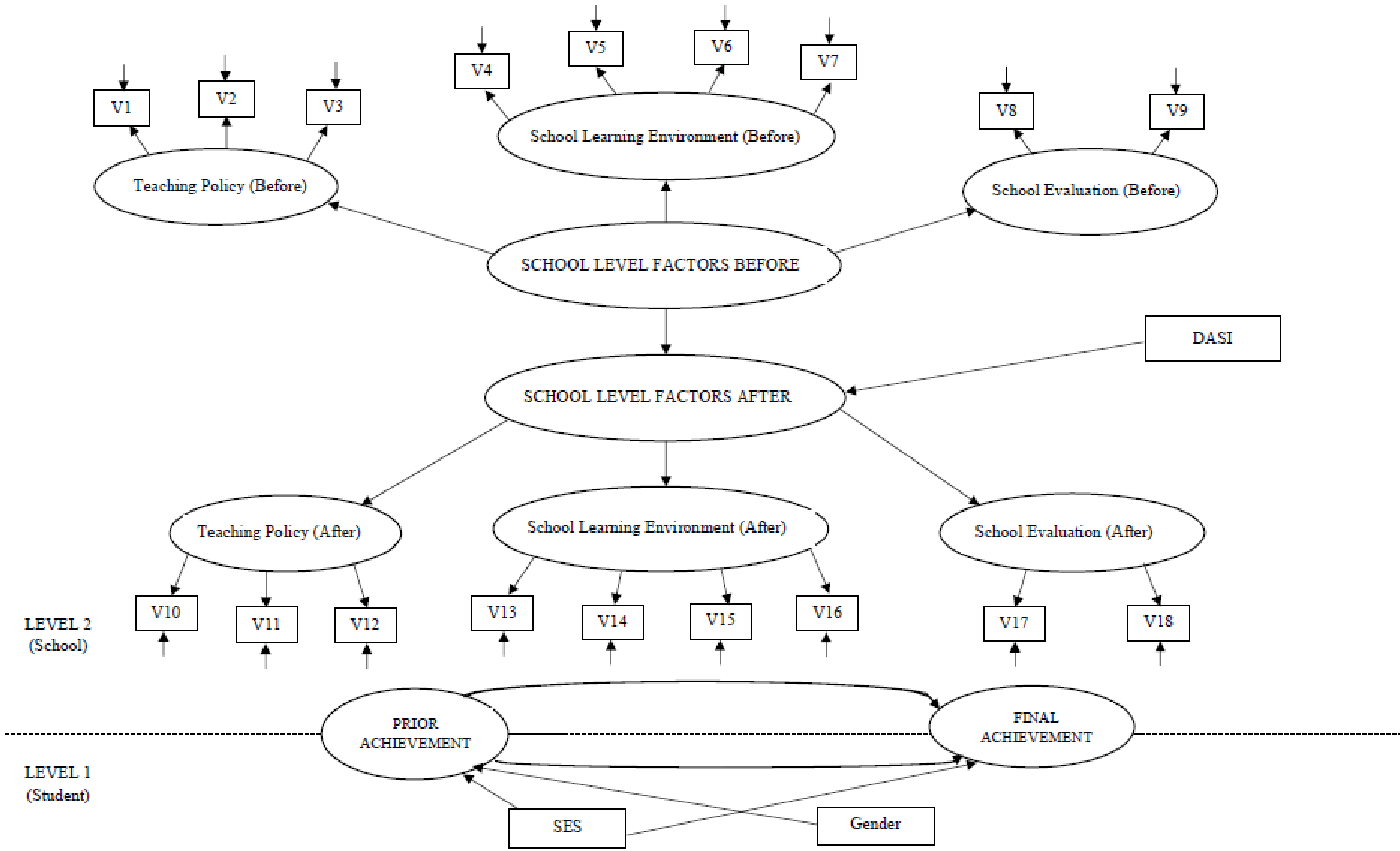

Figure 1. The theoretical multilevel model supporting that DASI had only direct effect on student achievement gains in mathematics

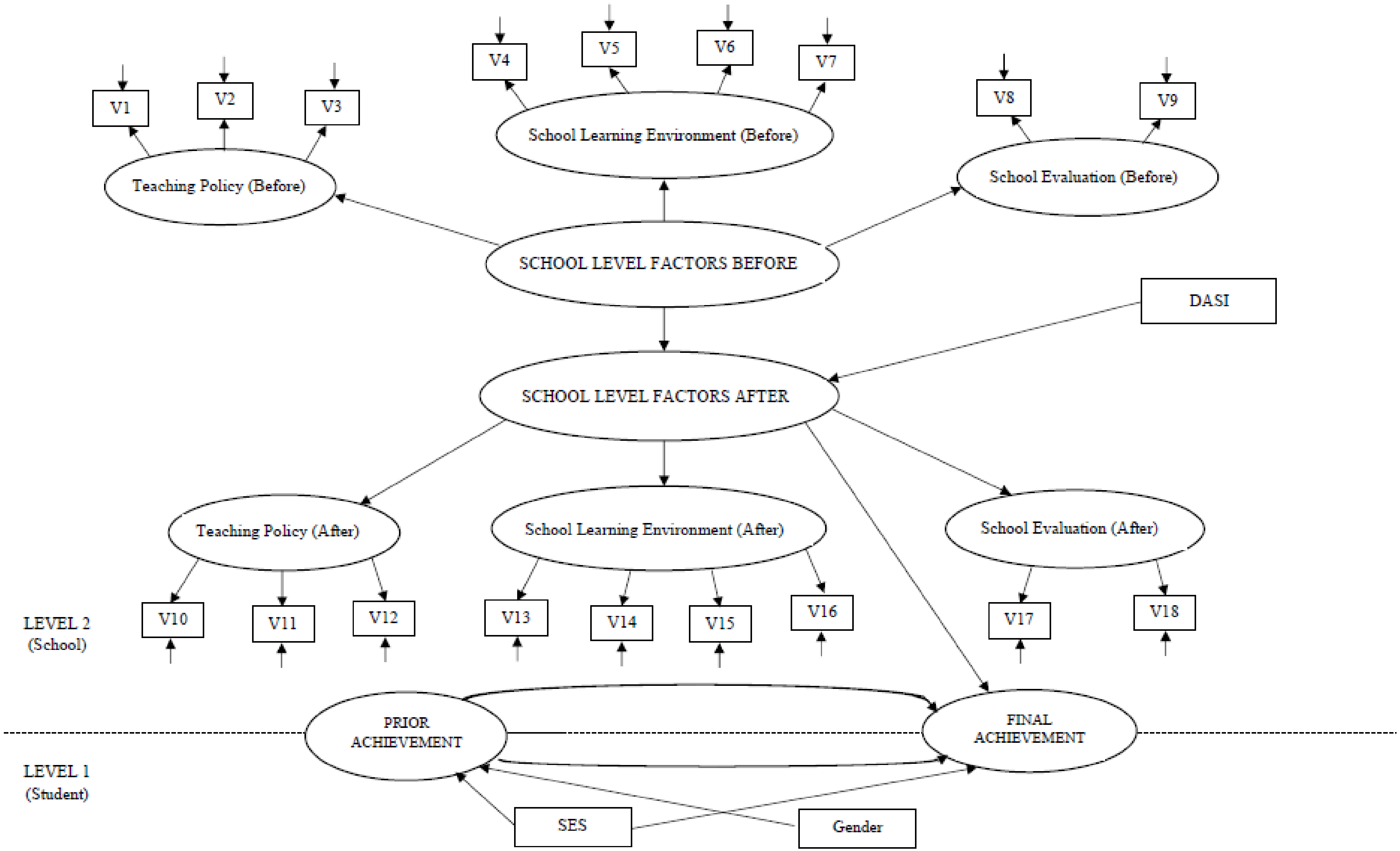

Figure 2. The multilevel model supporting that DASI had an indirect effect on student achievement gains in mathematics

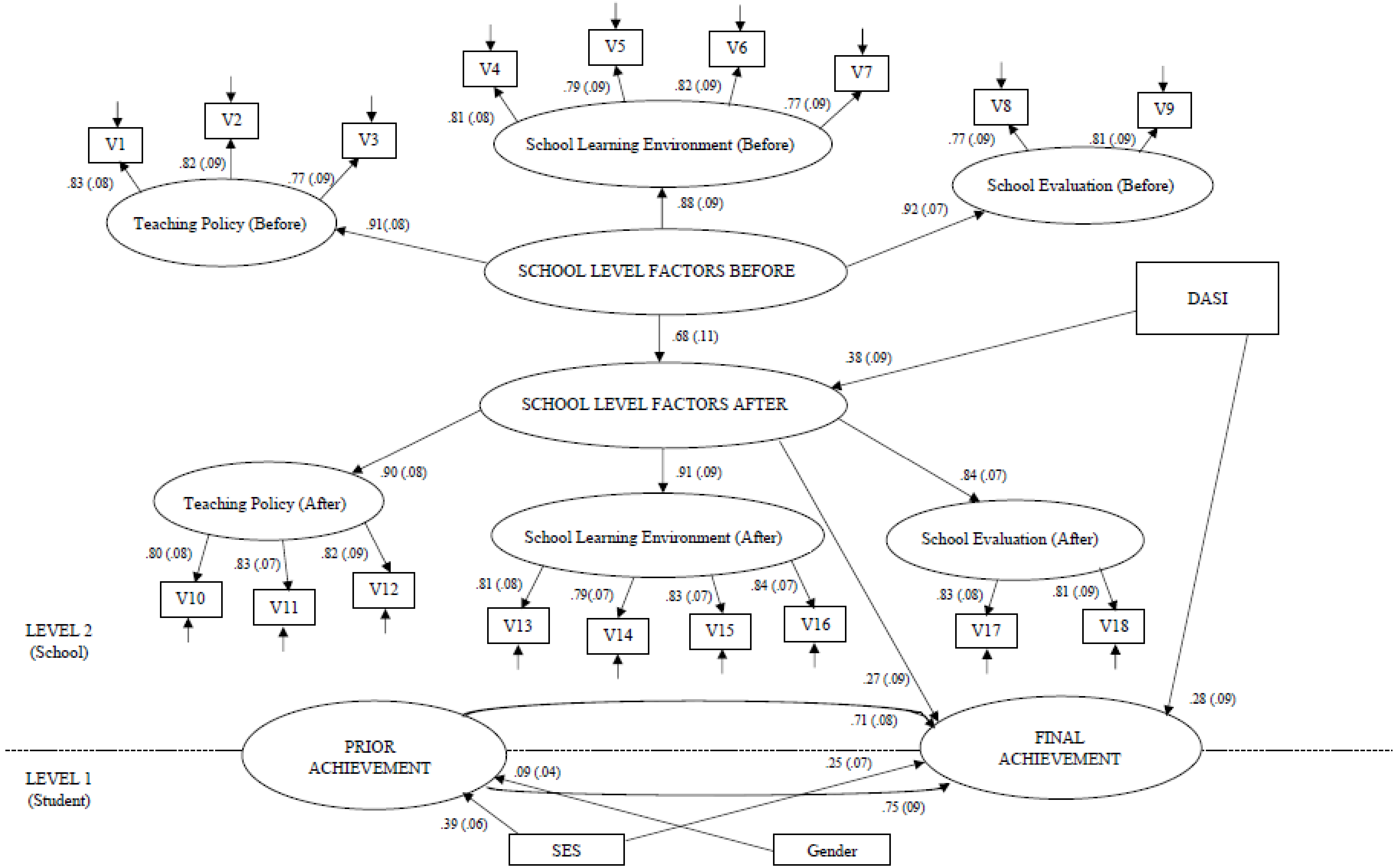

Figure 3. The multilevel model illustrating the direct and indirect effects of DASI on student achievement gains in mathematics

V1: Quantity of teaching (Before)
V2: Provision of learning opportunities (Before)
V3: Quality of teaching (Before)
V4: Student behaviour (Before)
V5: Teacher collaboration (Before)
V6: Partnership policy (Before)
V7: Provision of recourses (Before)
V8: Evaluation of teaching policy (Before)
V9: Evaluation of school learning environment (Before)
V10: Quantity of teaching (After)
V11: Provision of learning opportunities (After)
V12: Quality of teaching (After)
V13: Student behaviour (After)
V14: Teacher collaboration (After)
V15: Partnership policy (After)
V16: Provision of recourses (After)
V17: Evaluation of teaching policy (After)
V18: Evaluation of school learning environment (After)

Model fit statistics for each of the three models are reported in Table 9. We can observe that model 3 fits the data best. Specifically, the p value for the chi-square test of model 3 was found to be higher than 0.05. Moreover, both the CFI and the Tucker-Lewis Index (TLI) were higher than 0.95. As far as the value of the RMSEA is concerned, it was lower than 0.06. These results reveal that model 3 fitted the data well (see Hu and Bentler 1999). Figure 3 illustrates the model with the best fit. The estimated standardised parameters are also presented (standard errors are put in parentheses). All parameter estimates are statistically significant at the .01 level. At the lowest level, only two background variables (i.e., SES, and gender) were found to be associated with prior achievement in mathematics, but only SES was found to have an effect on final achievement. When the three background variables were aggregated at the school level, none of them was found to have an effect either on prior or on final achievement. On the other hand, prior achievement was found to be a good predictor of final achievement, both at the student and the school level. This finding provides support for the predictive validity of the battery of tests used to measure student achievement in mathematics. At the school level, the figure also reveals that the use of the proposed whole-school approach had a positive direct impact on student achievement in mathematics. An indirect effect upon student achievement in mathematics can also be identified due to school use of the approach which improved the functioning of school factors. It is important to take into consideration the fact that the measures of the three school-level factors were found to belong to a single latent variable, implying that the school-level factors were related to each other. Thus the results of the SEM analysis provided support for the main assumption upon which this intervention was based. This is due to the fact that the intervention was found to have both a direct effect on student achievement gains in mathematics and an indirect effect through improvement in the functioning of school factors. It should, however, be acknowledged that the direct effect of DASI was found to be much stronger than its indirect effect, and this result seems not to be in line with the assumption that DASI has mainly indirect effects on student achievement.

Table 9. Summary of fit results for the three alternative models concerned with the effect of intervention upon students' final achievement in mathematics

Alternative models	X^2	Df	p	CFI	TLI	RMSEA	SRMR (B)	SRMR (W)
1) Direct effect model	1835	227	0.001	0.92	0.93	0.14	0.189	0.015
2) Indirect effect model	1722	227	0.001	0.90	0.91	0.12	0.174	0.011
3) The direct and indirect effect model	249	226	0.141	0.99	0.99	0.03	0.115	0.006

Note: CFI = Comparative Fit Index; TLI = Tucker-Lewis Fit Index; RMSEA = Root Mean Square Error of Approximation; SRMR(W) = Square root mean error for the student level; SRMR(B) = Square root mean error for the school level.

A Synopsis of the Main Findings of the Intervention

The experimental study attempted to measure the effect of a dynamic approach to school improvement on improving student achievement in mathematics (quality) and reducing the effect of SES on mathematics achievement (equity). With regard to the main findings of this experimental study, cross-countries analysis revealed that students of schools implementing DASI managed to achieve better results in mathematics than schools in the control group. This finding revealed that DASI had an impact on the quality dimension and its actual effect size was found to be higher than the impact reported in most studies investigating the impact of interventions in education on student learning outcomes. Within-country analyses were also conducted and revealed positive effects of DASI in each country. Small differences in the effect of DASI were observed in the four participating countries. With regard to the effect of DASI on the equity dimension, we show that the impact of SES on student achievement was similar at the beginning of the intervention in both the experimental and the control schools. However, at the end of the intervention, the impact of SES on student achievement in mathematics was smaller in schools implementing DASI. In addition, the effect of SES on achievement gradually increased in the schools in the control group, whereas it remained the same in the experimental group. This means that the initial gap in achievement based on SES increased in the schools in the control group but no further increase was observed in the experimental schools. It can therefore be argued that DASI had an impact not only on the quality, but also on the equity dimension of school effectiveness. In the final part of this report, it is also demonstrated that DASI had an impact on student learning outcomes through improvement in the school factors of the dynamic model. In this way, empirical support for the main assumption of DASI was provided. Implications of findings for policy, research and practice are discussed in the Intellectual Output 8 of this project.

References

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, *1*(4), 363–378.

Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Byrne, M.B. (1994). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, Applications and Programming.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Creemers, B.P.M., & Kyriakides, L. (2012). *Improving quality in education: Dynamic approaches to school improvement.* London and New York: Routledge.

Creemers, B.P.M., & Kyriakides, L. (2015). Developing, testing and using theoretical models of educational effectiveness for promoting quality in education. *School Effectiveness and School Improvement*, *26*(1), 102-119.

den Brok, P., Brekelmans, M., Levy, J., & Wubbels, T. (2002). Diagnosing and improving the quality of teachers' interpersonal behavior. *International Journal of Educational Management, 16*(4), 176–184.

Devine, D (2013). Valuing children differently? Migrant children in education. *Children & Society, 27*(4), 282-294.

Dixon, W.J. (1954). Power under normality of several non-parametric tests. *Annals of Mathematical Statistics, 25*(3), 610-614.

Elliot, K., & Sammons, P. (2004). Exploring the use of effect sizes to evaluate the impact of different influences on child outcomes: Possibilities and limitations. In I. Schagen & K. Elliott (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 6-24). Slough: National Foundation for Educational Research.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer.

Hansen, K.Y., Rosén, M., & Gustafsson, J.-E. (2011). Changes in the Multi-Level Effects of Socio-Economic Status on Reading Achievement in Sweden in 1991 and 2001. *Scandinavian Journal of Educational Research, 55*(2), 197-211.

Hu, L., & Bentler, P.M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Keeves, J.P., & Alagumalai, S. (1999). New Approaches to Measurement. In G.N. Masters & J.P. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment* (pp. 23-42). Oxford: Pergamon.

Kendall, M.G., & Babington, S.B. (1939). The Problem of *m* Rankings. *The Annals of Mathematical Statistics, 10*(3), 275–287.

Kolen, M.J., & Brennan, R.L. (1995). *Test equating: Methods and practices.* New York: Springer.

Kyriakides, L., Creemers, B.P.M., Muijs, D., Rekers-Mombarg, L., Papastylianou, D., Van Petegem, P., & Pearson, D. (2014). Using the dynamic model of educational effectiveness to design strategies and actions to face bullying. *School Effectiveness and School Improvement, 25*(1), 83-104.

Kyriakides, L., Creemers, B.P.M., Panayiotou, A., Vanlaar, G., Pfeifer, M., Gašper, C., & McMahon, L. (2014). Using student ratings to measure quality of teaching in six European countries. *European Journal of Teacher Education, 37*(2), 125-143.

Muthén, L.K., & Muthén, B.O. (1999). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Muthén, L.K., & Muthén, B.O. (2001). *Mplus: statistical analysis with latent variables: user's guide* (2nd Ed.). Los Angeles, CA: Muthén & Muthén.

Panayiotou, A., Kyriakides L., & Creemers, B.P.M. (2016). Testing the validity of the dynamic model at school level: a European study. *School leadership and management*, *36*(1), 1-20.

Panayiotou, A., Kyriakides, L., Creemers, B.P.M., McMahon, L., Vanlaar, G., Pfeifer, M., Rekalidou, G., & Bren, M. (2014). Teacher Behavior and Student Outcomes: Results of a European Study. *Educational Assessment, Evaluation and Accountability*, *26,* 73–93.

Sammons, P. (2008). Zero tolerance of failure and New Labour approaches to school improvement in England. *Oxford Review of Education, 34*(6), 651-664.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Siegel, S., & Castellan, N.J. (1988). *Non-Parametric Statistics: For the Behavioural Sciences* (2nd Ed). New York: McGraw-Hill.

Snijders, T., & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling.* London: Sage.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement, 21*(3), 289–314.

Tomlinson, C. (1999). *The differentiated classroom: Responding to the needs of all learners*. Alexandria, VA: Association for Supervision and Curriculum Development.

Vanlaar, G., Kyriakides, L., Panayiotou, A., Vandecandelaere, M., McMahon, L., De Fraine, B., & Van Damme, J. (2016). Do the teacher and school factors of the dynamic model affect high- and low-achieving student groups to the same extent? A cross-country study. *Research Papers in Education, 31*(2), 183-211.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1(6), 80–83.*

Wright, B.D. (1985). Additivity in psychological measurement. In E. E. Roskam (Ed), *Measurement and personality assessment,* (pp. 101-112). Amsterdam: Elsevier Science Publishers BV.

Appendix

An Example of the Report Given to Each School of the Experimental Group of the Study

……………………………………………………………

ERASMUS+ Key Action 2 - Cooperation for innovation and the exchange of good practices, Field: School Education

Promoting quality and equity: A dynamic approach to school improvement – PROMQE

Outcomes of the Teacher Questionnaire

School 1

During the third phase of our research programme in which the initial evaluation of your school policy has been conducted, a questionnaire was administered to teachers examining the policy developed by your school in relation to the factors of the Dynamic Model of Educational Effectiveness operating at the school level. These factors refer to the a) school policy for teaching, and b) school policy for creating a learning environment at school.

We would like to thank you for completing the questionnaire (the questionnaire had been completed by 10 teachers) and inform you that we have managed to identify the improvement areas of your school based on your responses. Data analysis was based on the non-parametric statistical criterion Kendall's W.

According to the opinions of your school teachers, the three factors with the lower average ranking in relation to the other school factors are the following:

A. Provision of sufficient learning resources for students and teachers

B. Quality of teaching

C. Quantity of teaching

Therefore, your school may choose *one* of the above areas on which to focus its improvement strategies and actions or *combine some of these areas*. In the handbook, which will be given to you during our next meeting, you will find several suggestions for actions that need to be taken in order to improve each one of these factors. Of course, with the support of our research team, you can develop your own suggestions and actions depending on the needs and context of your school. Thus the next step is to select the area or areas you wish to improve and develop your action plan.

Thank you once again for your cooperation.