UNIVERSITY OF CYPRUS

DEPARTMENT OF EDUCATION

FACTORS PROMOTING QUALITY OF EDUCATION AT CLASSROOM AND SCHOOL LEVEL: A EUROPEAN EFFECTIVENESS STUDY IN MATHEMATICS AND SCIENCE

DOCTOR OF PHILOSOPHY DISSERTATION

ANASTASIA PANAYIOTOU

2015

ANASTASIA PANAYIOTOU

University of Cyprus

DEPARTMENT OF EDUCATION

FACTORS PROMOTING QUALITY OF EDUCATION AT CLASSROOM AND SCHOOL LEVEL: A EUROPEAN EFFECTIVENESS STUDY IN MATHEMATICS AND SCIENCE

ANASTASIA PANAYIOTOU

A Dissertation Submitted to the University of Cyprus in Partial Fullfilment of the Requirements for the Degree of Doctor of Philosophy

November 2015

ANASTASIA PANAYIOTOU

ANASTASIA PANAYIOTOU

© Anastasia Panayiotou, 2015

VALIDATION PAGE

Doctoral Candidate: Anastasia Panayiotou

Doctoral Thesis Title: Factors promoting quality of education at classroom and school level: a European effectiveness study in mathematics and science

The present Doctoral Dissertation was submitted in partial fullfilment of the requirements for the Degree of Doctor of Philosophy at the Department of Education and was approved on the 20th of November 2015 by the members of the Examination Committee.

Examination Committee:

Research Supervisor: Leonidas Kyriakides, Professor
Department of Education, University of Cyprus

……….……………………

Committee Members: Maria Eliophotou Menon, Associate Professor
Department of Education, University of Cyprus (Chair)

……………………………

Charalambos Charalambous, Assistant Professor
Department of Education, University of Cyprus

Pamela Sammons, Professor
Department of Education, University of Oxford

Jan-Eric Gustafsson, Professor
Department of Education and Special Education,
University of Gothenburg

ANASTASIA PANAYIOTOU

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Anastasia Panayiotou

………………………..

ANASTASIA PANAYIOTOU

STATEMENT OF JOINT WORK

The present doctoral dissertation presents the results of an international study with the participation of six European countries (i.e., Belgium/Flanders, Cyprus, Germany, Greece, Ireland, and Slovenia). Therefore the contribution of international partners is acknowledged.

School selection, test and questionnaire administration as well as data entering in Belgium, Germany, Greece, Ireland, and Slovenia were conducted by members of the corresponding national teams. School selection, test and questionnaire administration as well as data entering in Cyprus were conducted by the author (A. Panayiotou). Data processing (i.e., cleaning and recoding data from Cyprus and Greece as well as merging all data to create an international database) was performed by the author.

All statistical analyses presented in this doctoral dissertation were performed by the author.

ANASTASIA PANAYIOTOU

Abstract

National studies conducted in the past few decades were able to provide a significant insight as to the functioning of education and led to the identification of several factors explaining variance in student outcomes. These studies have also led to the development of the theoretical base of educational effectiveness research and to the establishment of a mutual basis for discussion among researchers. However, the need of expanding country restricted knowledge and sharing research results has led researchers to an attempt of not only conducting international conferences where knowledge can be more widely spread but also to the recognition of the contribution of international studies. International longitudinal studies may contribute to the field of educational effectiveness by providing empirical support to existing theoretical models, thus assisting the establishment of a solid theoretical framework, demonstrating factors that are associated with student progress irrespective of the context of each country.

Taking in mind the contribution of international longitudinal research, this study collected data from six European countries (i.e., Belgium/Flanders, Cyprus, Germany, Greece, Ireland, and Slovenia) aiming to provide support to the assumptions of the dynamic model of educational effectiveness in regard to factors operating at two different levels; the classroom and school level. Specifically, this study aimed to examine the generic and differential effects of the classroom and school level factors of the dynamic model in six different countries and two different subjects (i.e., mathematics and science) and also test whether the relationship of some factors with achievement is not linear but curvilinear.

Thus, in each participating country a sample of at least 50 primary schools was drawn and tests in mathematics and science were administered to all grade 4 students (n=10742) at the beginning and at the end of school year 2010-2011. For the construction of the tests, permission was obtained from IEA to use the released items of TIMSS 2007.

ANASTASIA PANAYIOTOU

Students were also asked to complete a questionnaire measuring teacher behavior in classroom (i.e., quality of teaching). Similarly, questionnaires were administered to the teacher sample (n=2923) to collect information concerning the different aspects of school policy as they are described in the dynamic model. Since both, students and teachers were asked to report on factors belonging to a higher level, one-way analysis of variance was used and showed that the student and teacher data can be generalized at the classroom and school level, respectively.

Structural Equation Modeling also provided support to the validity of student and teacher responses to these questionnaires. Separate multilevel modelling analyses for each subject were conducted both across- and within country, to identify the impact of the teacher and school factors on student achievement. The results of these analyses revealed that most of the teacher and school factors of the dynamic model explain student achievement gains in mathematics and science and provided support to the generic nature of these factors in six different countries and two different outcomes. Multivariate analyses also supported the generic nature of the teacher and school factors in mathematics and science. It was also found that teachers and schools that are effective in one subject are also effective in the other. Curvilinear relations between some classroom level factors and student achievement were also identified. Implications of findings for theory and practice are drawn.

ANASTASIA PANAYIOTOU

Περίληψη

Οι έρευνες που πραγματοποιήθηκαν τα τελευταία χρόνια στο χώρο της εκπαιδευτικής αποτελεσματικότητας ήταν σε θέση να προσφέρουν σημαντικές πληροφορίες ως προς τον τρόπο λειτουργίας της εκπαίδευσης και οδήγησαν στον εντοπισμό διάφορων παραγόντων που μπορούν να εξηγήσουν τις διαφορές που παρατηρούνται στα μαθησιακά αποτελέσματα. Οι έρευνες αυτές οδήγησαν επίσης στην ανάπτυξη του θεωρητικού πλαισίου της έρευνας για την εκπαιδευτική αποτελεσματικότητα και στη δημιουργία μια κοινής βάσης για συζήτηση μεταξύ των ερευνητών. Ωστόσο, η ανάγκη για διαχύση των αποτελεσμάτων των εθνικών ερευνών και η επιδίωξη για ανταλλαγή γνώσεων μεταξύ ερευνητών που προέρχονται από διαφορετικές χώρες, έχει οδηγήσει τους ερευνητές σε μια προσπάθεια όχι μόνο διεξαγωγής διεθνών συνεδρίων, όπου η γνώση μπορεί να διαδοθεί ευρύτερα, αλλά και στην αναγνώριση της συμβολής των διεθνών ερευνών. Διεθνείς διαχρονικές έρευνες (longitudinal) μπορούν να συμβάλουν στον τομέα της εκπαιδευτικής αποτελεσματικότητας παρέχοντας εμπειρική υποστήριξη σε υπάρχουσες θεωρίες και βοηθώντας έτσι τη δημιουργία ενός στέρεου θεωρητικού πλαισίου, εντοπίζοντας παράγοντες που σχετίζονται με την πρόοδο των μαθητών, ανεξάρτητα από το συγκείμενο της κάθε χώρας.

Λαμβάνοντας υπόψη τη συμβολή των διεθνών διαχρονικών ερευνών, η έρευνα αυτή συνέλεξε δεδομένα από έξι ευρωπαϊκές χώρες (Βέλγιο, Κύπρο, Γερμανία, Ελλάδα, Ιρλανδία και Σλοβενία) με στόχο την παροχή υποστήριξης στις βασικές υποθέσεις του δυναμικού μοντέλου εκπαιδευτικής αποτελεσματικότητας σε σχέση με τους παράγοντες που λειτουργούν σε δύο διαφορετικά επίπεδα: το επίπεδο της τάξης και του σχολείου. Συγκεκριμένα, στόχος της παρούσας έρευνας ήταν η εξέταση της επίδρασης των παραγόντων που περιλαμβάνονται στο επίπεδο της τάξης και του σχολείου στο δυναμικό μοντέλο, στην πρόοδο των μαθητών σε έξι διαφορετικές χώρες και δύο διαφορετικά γνωστικά αντικείμενα (μαθηματικά και επιστήμη). Στόχος επίσης της παρούσας έρευνας

ANASTASIA PANAYIOTOU

ήταν να εξετάσει κατά πόσο η σχέση κάποιων παραγόντων με τα μαθησιακά αποτελέσματα δεν είναι γραμμική αλλά καμπυλόγραμμη.

Συνεπώς, σε κάθε συμμετέχουσα χώρα επιλέγηκε δείγμα τουλάχιστον 50 σχολείων πρωτοβάθμιας εκπαίδευσης και σε όλους τους μαθητές Δ' τάξης χορηγήθηκε δοκίμιο για τη μέτρηση της επίδοσής τους στα μαθηματικά και την επιστήμη (n = 10742), στην αρχή και στο τέλος του σχολικού έτους 2010-2011. Για την κατασκευή των δοκιμίων, λήφθηκε άδεια από το διεθνή οργανισμό IEA για να χρησιμοποιηθούν θέματα από τη διεθνή έρευνα TIMSS 2007. Οι μαθητές κλήθηκαν επίσης να συμπληρώσουν ένα ερωτηματολόγιο για τη μέτρηση της διδακτικής συμπεριφοράς του εκπαιδευτικού τους στην τάξη (δηλαδή, την ποιότητα διδασκαλίας). Παρομοίως, ερωτηματολόγια χορηγήθηκαν και στο δείγμα των εκπαιδευτικών (n = 2923) για τη συλλογή πληροφοριών σχετικά με τις διάφορες πτυχές της πολιτικής του σχολείου, όπως αυτές περιγράφονται στο δυναμικό μοντέλο. Εφόσον τόσο οι μαθητές, όσο και οι εκπαιδευτικοί κλήθηκαν να δηλώσουν τις απόψεις τους σχετικά με παράγοντες που ανήκουν σε υψηλότερο επίπεδο, εφαρμόστηκε ανάλυση διακύμανσης μιας εξαρτημένης μεταβλητής και έδειξε ότι τα δεδομένα που συλλέχθηκαν από τους μαθητές και τους εκπαιδευτικούς μπορούν να γενικευθούν στο επίπεδο της τάξης και του σχολείου, αντίστοιχα.

Παράλληλα, η εγκυρότητα των δεδομένων που συλλέχθηκαν από τα ερωτηματολόγια των μαθητών και των εκπαιδευτικών του δείγματος ελέγχθηκε και επιβεβαιώθηκε με τη χρήση των Μοντέλων Δομικών Εξισώσεων (Structural Equation Modeling). Στη συνέχεια, για τη διαπίστωση της επίδρασης των παραγόντων σε επίπεδο τάξης και σχολείου στην επίδοση των μαθητών σε κάθε ένα από τα δύο γνωστικά αντικείμενα που εξετάστηκαν στην παρούσα έρευνα, διεξάχθηκαν ξεχωριστές πολυεπίπεδες αναλύσεις ανά γνωστικό αντικείμενο, αρχικά με τα δεδομένα όλων των χωρών (across-country analysis) και στη συνέχεια με τα δεδομένα της κάθε χώρας ξεχωριστά (within-country analysis). Τα αποτελέσματα των αναλύσεων αυτών κατέδειξαν

ANASTASIA PANAYIOTOU

ότι η πλειονότητα των παραγόντων σε επίπεδο τάξης και σχολείου είναι σε θέση να εξηγήσουν την πρόοδο των μαθητών στα μαθηματικά και στην επιστήμη σε έξι διαφορετικές χώρες. Ταυτόχρονα, πολυμεταβλητή ανάλυση (multivariate analysis) που εφαρμόστηκε υποστήριξε τη γενική φύση (generic nature) των παραγόντων σε επίπεδο τάξης και σχολείου, ενώ κατέδειξε επίσης ότι οι εκπαιδευτικοί και τα σχολεία που είναι αποτελεσματικά στο ένα γνωστικό αντικείμενο είναι επίσης αποτελεσματικά και στο άλλο. Τέλος, διαφάνηκε καμπυλόγραμμη σχέση μεταξύ ορισμένων παραγόντων του επιπέδου της τάξης και της επίδοσης των μαθητών στα μαθηματικά και την επιστήμη.

ANASTASIA PANAYIOTOU

Acknowledgments

Coming to the end of a four year journey I would like to thank all those who provided me with support, knowledge and inspiration and contributed to making it all possible.

Firstly, I would like to express my sincere gratitude to my advisor Prof. Leonidas Kyriakides for the continuous support he provided me with during my PhD study. His guidance helped me in all the time of research and writing of this thesis and his belief in my abilities helped me evolve not only as a researcher but also as a person. I could not have imagined having a better advisor and mentor.

Besides my advisor, I would also like to thank the members of my committee: Dr. Maria Eliophotou Menon, Dr. Charalambos Charalambous, Prof. Pamela Sammons and Prof. Jan-Eric Gustafsson, for their time and insightful comments, and also for their constructive feedback helping me to widen my research from various perspectives.

Also, I would like to thank my family and especially my parents who provided me with their full support during the entire time of my studies. I would also like to thank my sister Andrie who provided me with guidance and was always willing to answer my questions. Thanks are also due to the newest addition in our family, my godson George who provided me with much needed breaks during the writing of my thesis!

I am also grateful to my best friend throughout the years, Evangelia who showed me that people can be so different and yet so much alike and provided me with unconditional support even though she never really understood what I did! Last but not least, I would like to thank my partner Philippos who always reminded me that success and growth only come when you manage to turn your greatest weaknesses into your greatest strengths...

ANASTASIA PANAYIOTOU

To all those who believed in me…

“The price of victory is high but so are the rewards”

Paul Bryant

ANASTASIA PANAYIOTOU

TABLE OF CONTENTS

ANASTASIA PANAYIOTOU

ANASTASIA PANAYIOTOU

ANASTASIA PANAYIOTOU

LIST OF TABLES

ANASTASIA PANAYIOTOU

ANASTASIA PANAYIOTOU

LIST OF FIGURES

ANASTASIA PANAYIOTOU

CHAPTER 1

INTRODUCTION TO THE STUDY

This chapter presents an overall view of the study. Specifically, an overview of the main advances and limitations in Educational Effectiveness Research (EER) is presented, the research purpose of the study is stated and specific research aims and questions are set. Then the practical and theoretical contribution of the study is outlined. Finally, a brief presentation of the study's outline is included in order to facilitate further reading.

Introduction

Educational Effectiveness Research (EER) concentrates on identifying the reasons for which differences in student outcomes exist, in terms of explaining why some schools and teachers are more effective than others. EER has shown great improvement in the last three decades both on methodological (Goldstein, 2003; Creemers, Kyriakides & Sammons, 2010) as well as on theoretical aspects (Levin & Lezotte, 1990; Scheerens & Bosker, 1997), by taking into account the criticism on the research practices used in the field. Along with the methodological advances in EER, such as the development of multilevel mathematical models which could provide a more accurate perspective on the different effects of all the levels of education (Teddlie & Reynolds, 2000), the need for international studies aiming to search for factors that can contribute in improving the different outcomes of schooling (i.e., cognitive, affective, psychomotor etc) has been outlined and the contribution of international effectiveness studies has been recognized (Sammons, 2006). In this context, the importance of international comparative studies has led to the development of the International Congress for School Effectiveness and Improvement (ICSEI), which aims to expand country restricted knowledge and provide a basis for the exchange of experience and expertise among researchers, contributing to the

ANASTASIA PANAYIOTOU

better understanding of what constitutes educational effectiveness, so that suggestions can be provided to policy makers and practitioners in different countries on methods to increase school effectiveness; and consequently student outcomes (Creemers, et al., 1998; Townsend, 2007).

Based on the assumption that the educational effectiveness knowledge base can be used for the improvement of education, the need for international studies searching for methods that can increase national standards has extensively been discussed by researchers across countries (e.g., Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002; Creemers, 2006; Sammons, 2006) since most of the school effectiveness studies are conducted in one single country, in opposition to other scientific fields, such as psychology and the pure and applied sciences (Reynolds, 2000).

The Contribution of International Studies

The identification of factors that have an effect on student outcomes in different countries is essential, since we cannot refer to one method that works in each country irrespective of the context and the characteristics of different student populations. Effective policies in one country may not have an effect on student achievement in a country with a different context and thus, across-country studies are required to investigate generic factors that function regardless of the country's context (Kyriakides, 2006a). However, the results of comparative studies that took place in one country and investigated the factors associated with student outcomes were expected to provide information to policy-makers, researchers and other stakeholders on the functioning of the educational systems in general (Mullis et al., 2000), possibly creating the misconception that the suggestions for raising national standards can be applied in an equally effective way in other countries. International studies can be the answer to this tendency as they may provide evidence for

ANASTASIA PANAYIOTOU

the functioning of effectiveness factors in different contexts and reveal generic factors that can be a basis for policy development in different countries.

Additionally, another issue that has raised concern among researchers is the one of ensuring enough statistical power which will allow the estimation of the true factor effects (Basagaña & Spiegelman, 2010; Maxwell, Kelley & Rausch, 2008; Moerbeek, 2008; Schochet, 2008). In opposition to national studies and meta-analyses, only international studies may provide enough statistical power to identify the full range of classroom and school effects, by increasing the variation at the school and classroom level (Kyriakides, 2006b). The effects of the teacher and school factors are expected to be much smaller within a country, while in international samples the variation is expected to be much bigger giving us the opportunity to identify the true size of the educational effects upon student outcomes. This perspective has also been supported by a secondary analysis of the PISA 2000 study by Kyriakides and Demetriou, (2005), which showed that the variation at the student level is smaller (less than 50%) than most national effectiveness studies in which the variation at student level is greater than 60% (e.g., Sammons, 1999; De Jong, Westerhof, & Kruiter, 2004; Kyriakides, 2005).

Finally, EER will benefit from the results of international studies as they will contribute to the validation of existing theoretical models in the field and assist the establishment of a solid theoretical framework which will indicate the factors that can be used across countries for the increase of student gains from education (Creemers, 2006). International studies will provide us with a more holistic view of what works in education and will help us draw implications for practices that are proven to be effective, taking in mind the complex nature of education.

The international comparative studies that take place such as the Third International Mathematics and Science Study (TIMSS) are designed to provide policy makers, practitioners and researchers with information about educational achievement in different

ANASTASIA PANAYIOTOU

learning contexts, aiming also to sustain high standards and prevent bias (Beaton et al., 1996; Mullis, Martin, Gonzalez, & Chrostowski, 2004). However, in spite of the fact that they provide the opportunity to compare different countries on an equal basis, the methodology used seems to have some important limitations, with the most important being that the data collected in these international studies are cross-sectional, which means that they assess student achievement in only one point in time. This, not being a value-added approach, does not take in mind student progress over time and consequently the factors that have an effect on student achievement gains, also because data on the school and classroom procedures that are related to student achievement are not usually collected in these types of international studies (Reynolds, 2006). Therefore, they may provide policy makers with information regarding national standards in a large number of countries, but they have very few to offer as far as raising national standards through the provision of evidence-based information on ways to increase student gains, looking at student progress. These international studies could only be considered as longitudinal at country level and not at school, classroom and student level, thus not allowing the investigation of student progress over time at either of these lower levels (Gustafsson, 2007).

Cross-sectional studies may also lead to misleading notions regarding cause and effect relations of some effectiveness factors with student achievement. In particular, the mere association of some factors with student achievement obtained at one time point may be falsely perceived as a causal relation. In these cases however, the direction of causality may not be determined (e.g., in cases where reverse causality exists). For instance, a cross-sectional study may demonstrate a negative association between student achievement and school policy for communication with parents. This could be falsely be interpreted by concluding that schools establishing more actions for communication with parents lead to lower student achievement. However, this negative association could be explained by the

ANASTASIA PANAYIOTOU

fact that schools where student initial achievement is lower need to take further actions regarding different aspects of their policies (Gustafsson, 2013).

For this reason, international longitudinal studies are needed as they can facilitate the demonstration of causal relationships between variables and identify the factors that contribute to educational effectiveness (Creemers & Kyriakides, 2006). Specifically, longitudinal research is needed to allow the identification of the relationship between student performance and teacher - school effectiveness (quality of teaching and school policies) and to provide practitioners and policy-makers with feedback in order to improve the quality of education offered to students (Teddlie & Reynolds, 2000). The ultimate goal of EER is to locate those factors that are related to student learning/progress and can be adjusted through changes in policy and practices (Martin, 1996) in an attempt to model educational effectiveness and create a theory that can be used across countries to enhance education.

Thus, taking in mind that the major contribution of research on educational effectiveness lies on the development of theoretical models aiming to interpret the differences in student outcomes, as well as on providing suggestions for using multilevel models which offer a broader spectrum of data analysis (Kyriakides & Charalambous, 2005; Cools, De Fraine, Van den Noortgate & Onghena, 2009), international longitudinal studies may constitute a significant step towards establishing a solid theoretical framework that presents generic factors affecting student outcomes in the different levels of education, by providing empirical support to the assumptions of the existing theoretical models in the field.

One of the most recent theoretical models which deals extensively with the factors operating at the different levels of education and searches for the relationship of factors operating both across and within levels, is the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008), which was developed based on Creemers' Comprehensive

ANASTASIA PANAYIOTOU

model (Creemers, 1994a). The dynamic model places emphasis on the factors operating at the classroom and school level and was thus chosen to provide a basis for this study.

Research Purpose and Aims

This research aims to provide answers to current debates on the improvement of learning outcomes and on specific issues such as whether factors that operate at the classroom and school level can explain variation in student achievement gains in six European countries with rather different educational contexts. In this context, this research aims to further develop a theoretical framework that can be used to provide answers to questions concerned with the factors that are considered to have an effect on student achievement. Specifically, the extent to which the dynamic model of educational effectiveness can be used as a starting point for establishing a theory-driven and evidence-based approach to educational effectiveness will be investigated.

Although this framework is more complex than previous models of effectiveness such as Creemers' Comprehensive Model (Creemers, 1994a) or Scheerens' (1992) and Stringfield and Slavin's (1992) models, it is based upon research evidence. Empirical support to the validity of the model has also been provided through three national longitudinal studies testing the effects of classroom and school level factors upon achievement (Creemers & Kyriakides, 2010a; Kyriakides & Creemers, 2008, 2009) and through a quantitative synthesis of studies on school effectiveness conducted during the last two decades (Creemers & Kyriakides, 2008). It should however be acknowledged that these studies were national in character and thereby their findings cannot be used for supporting the generic nature of the factors. An international study may provide further support to the dynamic model and test the generalisability of the findings of these studies by searching for the impact of the classroom and school factors on student achievement in mathematics and science in six different European countries.

ANASTASIA PANAYIOTOU

Thus the main purpose of this study is the further development and testing of the validity of the dynamic model of educational effectiveness at the classroom and school level, in order to improve the effectiveness of education based on scientific, validated model(s) and thus, test the extent that the dynamic model can be used as a starting point for establishing the theoretical framework of EER.

More specifically this study aims to:

1. Examine the generic nature of the classroom and school factors included in the dynamic model of educational effectiveness by investigating the extent to which these factors have an effect on student achievement a) in two different outcomes (i.e., mathematics and science) and b) in six different countries.
2. Test the assumption of whether the classroom and school factors of the dynamic model have differential effects on student achievement in the two different outcomes and in six different countries.
3. Search for curvilinear relations of the factors included in the dynamic model at the classroom and school level with student achievement in mathematics and science.

Based on the results of this study suggestions for policy development in each of the participating countries will arise. For this study to be conducted, international instruments for measuring teacher behaviour in the classroom (classroom level factors) and the school level factors of the dynamic model will be developed and validated which will also provide a basis for measuring effectiveness factors in further international studies.

Study Summary

In each participating country (Belgium/Flanders, Cyprus, Germany, Greece, Ireland, and Slovenia), stratified sampling procedure (Cohen, Manion, & Morrison, 2000) was used to select a sample of at least 50 primary schools. Data on student achievement in

ANASTASIA PANAYIOTOU

mathematics and science were collected at the beginning and the end of grade 4 by using external forms of assessment which were designed to assess knowledge and skills in these two subjects. Collecting data in more than one phase helps to draw credible conclusions about causal relations between factors and outcomes. For the construction of the tests which were used to measure student achievement, permission was obtained from IEA to use the released items of TIMSS 2007. The properties of each item and its relation with the curricula of grades 3 and 4 in each country were taken into account for developing the two types of test. Data in each participating country was collected following the specific regulations of each country in relation to gaining permission for collection of data from parents, teachers, and schools. All the participants involved (schools, parents, children) were also informed that confidentiality would be ensured and kept throughout the procedure. Code numbers were also assigned to students, teachers, and schools to ensure confidentiality and provide the possibility to analyse the longitudinal data by linking the relevant data sets in the different phases of the study.

Additionally, all classroom and school level factors of the dynamic model were measured. Concerning the classroom factors, the student questionnaires which have been developed and tested in a national study conducted in Cyprus (Creemers & Kyriakides, 2008) were adapted in order to fit to the context of each country. These questionnaires were administered to the students of the sample in order to collect data on their teacher's instructional behaviour. Similarly, the teacher questionnaire which has again been previously used in the national study mentioned above was adapted in order to measure the school factors. This questionnaire was administered to all the teachers of the sample to gather information on their school's policy on teaching and the SLE and the school evaluation methods. For the construction of these instruments, the procedure of translation and back translation was used.

ANASTASIA PANAYIOTOU

To test the construct validity of the instruments which were used to measure the classroom and school factors, Structural Equation Modelling (SEM) techniques (Byrne, 1994) were used. More specifically, the scaled chi-square, Bentler's (1990) Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA) (Brown & Mels, 1990) were examined. Then, separate multilevel modelling techniques were used (Goldstein, 2003; Snijders & Bosker, 1999), to identify the extent to which each classroom- and school-level factor is associated with achievement in each outcome (mathematics and science) both linearly and non-linearly. Additionally, multivariate analysis was conducted to test the consistency between the two outcomes. Multilevel analyses helps identify the importance of each factor regarding student outcomes and reveal both generic and differential factors operating at different levels.

At the end of the study, feedback was provided to each school participating in the study but only general trends were mentioned and not the performance of individuals. This way, schools were encouraged to develop policies and actions for improving their effectiveness.

Contribution to the Theory

This study aims to investigate the extent to which the dynamic model of educational effectiveness (Creemers & Kyriakides, 2008) can be used as a starting point for establishing an evidence-based and theory-driven approach for designing reform policies to improve education in different contexts. Hitherto, national studies testing the effects of school and classroom level factors upon achievement of both cognitive and affective outcomes and two quantitative syntheses of studies on school and classroom effectiveness conducted during the last two decades (Creemers & Kyriakides, 2008) have provided empirical support to the validity of the dynamic model at national level. These studies reveal that basic elements of the model are associated with effectiveness in at least

ANASTASIA PANAYIOTOU

one country and can constitute a basis for an evidence-based and theory-driven approach to improvement of education (Kyriakides & Creemers, 2008). Therefore, one of the main contributions of an international study is the fact that it will be in a position to provide further support to the dynamic model and also investigate some further issues concerned with the effective functioning of education.

The first issue is concerned with the extent to which some factors of the dynamic model are related to learning outcomes irrespective of the context, whereas others have differential effects and are therefore more relevant for policy making in specific socio-cultural contexts. Findings on this issue will reveal the conditions under which a factor is associated with outcomes in some countries. Thus, the study will provide the first step so as to help policy makers understand the complexity of educational effectiveness and avoid practices adopted without any detailed knowledge of the possible contextual factors that might explain how factors that work in one country may be ineffective in another country (Reynolds, 2006). Finding differential effects may lead to further studies searching for individual characteristics of each country that explain these differences in the functioning of the factors.

The second issue regards the theoretical assumption of the dynamic model that the relation of some factors with achievement is non-linear (Creemers & Kyriakides, 2008). In order to test this assumption, a wide variation in the functioning of factors is needed and this variation may be provided by collecting data from schools and classrooms in different countries. Thus, findings of this study may reveal non-linear relations of some factors with achievement and therefore be in a position to help policy-makers and practitioners use these results for improvement purposes. Specifically, by finding non-linear relations of some factors with achievement, optimal points of their functioning could be identified and thereby different priorities for improvement may emerge in each school and classroom.

ANASTASIA PANAYIOTOU

Finally, this study will assist the establishment of internationally valid instruments for measuring teacher behaviour in the classroom (classroom level factors) and the school level factors of the dynamic model of educational effectiveness. Taking in mind the methodological lessons drawn from previous attempts for international research (e.g., the ISERP project which will be further elaborated on in the next chapter), providing researchers with instruments that can be used in across countries comparative studies for the measurement of effectiveness factors, is considered to be a crucial step in the success of international effectiveness studies, especially due to the complexities in measuring classroom processes (Teddlie, Creemers, Kyriakides, Muijs & Yu, 2006). Therefore, the contribution of this study expands in terms of not only providing answers to questions concerning the functioning of effectiveness factors in different contexts and thus broadening the knowledge on educational effectiveness, but also by contributing to the methodological progress made in EER in the last three decades.

Significance of the Study

One of the key questions modern educational effectiveness research faces is the search for factors that act at the different levels of education (system, school, class and student) and contribute to the improvement of learning outcomes. However, many studies have focused solely on the investigation of the individual factors which lie at one level of education. Instead, this study aims to search for the effects of factors operating at two different levels of education and provide answers on how these are likely to affect learning outcomes. By acknowledging how changes in the functioning of the factors at classroom and school level may affect student outcomes, we may be in a position to make practical suggestions on methods to improve student learning. Even though, only a small amount of variance in student outcomes (approximately 12-18%) can be explained by the classroom and school factors when controlling for student background (Creemers, 1994b) is

ANASTASIA PANAYIOTOU

important to identify which of these factors have a greater effect so as to inform policy makers accordingly and form improvement action plans.

Therefore, in spite of the scientific contributions of this study, its significance also lies on the practical contributions as its results will give us the possibility to inform national and European policy makers about effective practices at school and classroom level, contributing to the improvement of educational quality in terms of higher average achievement. Although national studies address some specific aspects of the functioning of effectiveness factors at different levels, the use of a common framework in international studies enables the provision of suggestions to policy makers on how to make use of research findings by transferring effective practices in other countries, taking into consideration their own educational context. Specifically, by having across-country data, specific improvement priorities for each country can be identified and suggestions can be made for factors that are found to be associated with student outcomes. For example, in case a factor is not found to perform well in a country but it is not associated with learning, it might not be a priority for improvement. On the contrary, a factor that is found to have a large impact on learning and its functioning is not satisfactory could constitute a basis for setting improvement priorities and creating improvement action plans.

Furthermore, this study gives the opportunity to investigate and explain differences in the added value of education, for which the collection of data from different countries is a prerequisite. Collection of data about quality of teaching, policies on teaching and on the school learning environment, and evaluation of policies and actions taken for improvement from different countries will establish variation in the functioning of factors of the dynamic model and allow for the measurement of their impact on quality in education. Based on this information, generic and differential factors will be identified and recommendations will be made for improving education in each participating country by taking into account and respecting each country's specific educational needs and improvement priorities.

ANASTASIA PANAYIOTOU

Thesis Structure

The complete thesis consists of five chapters. The first chapter provides an introduction to the research background, the research problem addressed as well as the research questions this study aims to answer, leading to a justification of the significance of the study; both in terms of its practical as well as on its theoretical contribution in the field of EER. The second chapter aims to provide a critical review of the literature elaborating on the theoretical framework upon which the study is based and on the fundamental issues regarding the purpose of the study. Specifically, it provides a brief overview of previous studies in the field of EER which focus on the effects of school and classroom factors on student achievement and examines the contribution of international studies in the field of EER through a review of the current international studies. Then, Chapter 3 describes the research methodology followed in the study. In this chapter, a detailed description of the processes of sampling and data collection is provided with particular reference to the data collection instruments and the statistical techniques used. Finally, the main limitations of the study are acknowledged and discussed. Chapter 4 presents the analysis of the data collected during the study. The analysis is made in order to provide answers to the main research questions, presented in Chapter 1. Finally, in the last chapter of the study, Chapter 5, the main results that occur from the analyses are discussed, with reference to each research question and to the overall aims of the study. Implications for theory, policy and practice are also drawn. Finally, Chapter 5 ends with suggestions for further research.

ANASTASIA PANAYIOTOU

CHAPTER 2

LITERATURE REVIEW

Introduction

The literature review, presented in this chapter, aims to provide a description of the theoretical framework that is related to the study's purpose. It aims to provide a broader view of the relevant concepts described in the study and demonstrate links with previous work conducted in the field of EER. Through a critical literature review, it creates a framework for the examination of the research problem and questions stated in the previous chapter. Therefore, this chapter concentrates on providing a review of the available literature within the field of EER, providing insight to the advancements made in the past few decades which lead to the theoretical and methodological developments as well as to the need for the internalization of EER.

Specifically, in the first section, a historical overview of EER is presented where the different phases of EER are discussed to demonstrate the growth that the field has met through the years. Then, the methodological as well as the theoretical advances in EER are elaborated on and the main models of educational effectiveness are described recognizing the leap from the earlier, simpler models to the most recent multilevel models which capture the complexity of education. Moving on, the next section presents and describes in detail the theoretical framework proposed and used in this study as well as the attempts that were made to validate its main assumptions. Recognizing both the contribution as well as the limitations of the studies aiming to provide support to the theoretical framework of this study, the final section of this chapter presents the main reasons for conducting international studies. The main conclusions drawn from previous international studies, as well as the main aims of the present study are also summarized.

ANASTASIA PANAYIOTOU

Educational Effectiveness Research: a Historical Overview

Educational Effectiveness Research (EER) aims to address the question of what works in education and why, thus identifying the reasons for which differences in student outcomes exist and provide an understanding of existing practices (Creemers, 1993). Specifically, EER aims at identifying factors which operate at the different levels of education, such as the classroom, school and system which have a direct or indirect effect on student outcomes (i.e., cognitive, affective, psychomotor), taking also into consideration student background characteristics (Raudenbush & Bryk, 1986).

Since the 1980s, EER has enjoyed rapid expansion in many countries (Teddlie & Reynolds, 2000). Methodological and technological advances have improved the power of estimation of teacher and school differences in relation to student achievement (Goldstein, 2003), while theoretical advances have also been observed since more specific definitions have been provided and an attempt to identify clearer relations between different concepts has been made (e.g., Scheerens, 1992; Teddlie & Reynolds, 2000 etc.). Additionally, in relation to the methodological advances that were made, the criticism placed on EER has led to the improvement of the measurement methods of student outcomes as well as to more general improvements in regard to the research design followed (Creemers, 2006).

Originally, the attention given to EER was a result of the early sociological and psychological studies of Coleman et al. (1966) and Jencks et al. (1972), respectively, which concluded that education had a very small contribution on student outcomes especially in case that student background characteristics were taken into consideration. These results were also reinforced by the failure of large-scale programmes applied in pre-primary schools the USA, such as "Headstart" and "Follow Through", which aimed at reducing the initial differences between students and address equity issues. These disappointing results led to reactions, both among practitioners as well as among researchers, especially since they opposed to the notion that schools had few to offer as far

ANASTASIA PANAYIOTOU

as student outcomes (Stringfield & Teddlie, 2011). These studies and the reactions their results caused were thus a catalyst for a line of early studies in the field of EER such as the smaller studies undertaken by Weber (1971) and Reynolds (1976) and the larger studies of Brookover, Beady, Flood, and Schweitzer (1979) in the USA, Rutter, Maughan, Mortimore and Ouston (1979) in the United Kingdom and the later study of Mortimore, Sammons, Stoll, Lewis, and Ecob (1988). These studies in the first phase of EER were able to demonstrate that differences in school effectiveness exist even when controlling for student background characteristics, assuming that these differences could be attributed to differences in the quality of education offered by schools (Goldstein & Woodhouse, 2000). In spite of the methodological weaknesses of these studies their optimistic results which showed that effective teachers and schools play an important role in student achievement, gave thrust to further research in the field of educational effectiveness which would then lead to questions in regard to the origin of those differences (Creemers & Scheerens, 1994).

Therefore, since at this early stage of EER, variation in student outcomes was identified, in the second phase of EER researchers aimed at explaining the reasons for which these differences exist and identify factors that can explain variation in outcomes among students (Levine & Lezotte, 1990; Sammons, Hillman & Mortimore, 1995; Scheerens & Bosker, 1997). Thus, a series of process-product studies have taken place and led to the identification of a list of factors that link specific teaching behaviors and characteristics to student outcomes (Doyle, 1986; Brophy & Good, 1986; Reynolds et al, 1996; Borich, 1992; Galton, 1987; Evertson et al., 1980). One of the first studies that were conducted and has led to the identification of five factors which were considered to be correlated with each other and linked to better student outcomes was the study undertaken by Edmonds (1979). Edmonds' "five-factor model" which included the factors of: a) strong educational leadership, b) high expectations of student achievement, c) emphasis on basic skills, d) safe and orderly climate and e) frequent evaluation of student progress was however heavily criticized as far as its methodology (e.g., Ralph & Fennessey, 1983). The methodological criticism of the studies conducted during the first and second phase of EER had gradually led the focus of researchers to not only the possible identification of separate factors which could explain variation in student outcomes, but also to the demonstration of causal relations between factors and achievement. This turn in focus was based on the framework developed by Scheerens and Creemers (1989a), which called attention to the possible contribution of the different levels of education to student outcomes.

Thus, in the third phase of EER, of which the beginning can be placed around the early 90s, researchers moved from searching effectiveness factors to answering the question of why specific factors are correlated with achievement (Scheerens & Bosker, 1997). In this context, three basic approaches have been used so as to identify the reasons for which certain factors or characteristics contribute to educational effectiveness. The first approach lies on the economic aspects of education and focuses on the relationship between schooling inputs and educational outputs controlling for the influence of several background characteristics (Monk, 1992; Hanushek, 1994; 1997). This implies that this approach places emphasis on the educational costs and attempts to identify their linkage with student outcomes (Creemers & Van der Werf, 2000) assuming that increased inputs can lead to improved outcomes (Creemers & Kyriakides, 2008). Based on this assumption several 'education production' models were developed, such as the models by Elberts and Stone (1988) as well as Brown and Saks (1986). However, the relationship between inputs and outputs in education is more complex than assumed (Creemers & Kyriakides, 2008). Thus, the education production studies that were conducted were not in a position to reveal the school inputs that can contribute to maximizing student gains from education (Monk, 1992) and in spite of the extensive research regarding educational processes and the economic aspects of education, no clear policy was drawn on the matter (Hanushek, 1986).

The second approach focuses on the sociological perspective of EER. This approach refers to factors relating to students' background characteristics, gender, as well as other social and cultural factors which can possibly affect student outcomes (Creemers & Kyriakides, 2008). Based on this approach, the possibility of adjusting for these background and social differences through education is examined. Therefore, apart from quality in education another aspect that gradually started to gain attention was the equity dimension (Creemers & Scheerens, 1994) which led to several studies searching for the differential effectiveness of schools in regard to different student populations (e.g. Campbell et al., 2004; Strand, 2010) and the effect of contextual factors on student outcomes (Opdenakker & Van Damme, 2006).

Finally, the third approach lies on the psychological aspects of EER and focuses on student background factors associated with motivation and learning aptitude, as well as with the learning process itself. Therefore, this approach called for more attention on the two main actors involved in the teaching and learning process (i.e., students and teachers), and led to a list of teacher behaviors in the classroom which were found to be consistently, positively related to student achievement (Creemers, Kyriakides & Sammons, 2010). Through the emphasis given to the teaching process and the combination of the results of studies conducted in the third phase of EER, two models emerged as described by Rosenshine (1983) and Brophy and Good (1986). The first model, called the "Direct Instruction Model of teaching" or "Structured approach", emerged from a series of experimental and correlational studies (e.g., Good & Grouws, 1979; Fitzpatrick, 1982) and is based on their results which demonstrated that students taught with structured curricula and under teacher supervision have a better achievement in comparison to students taught with more individualized or discovery learning approaches. However, a somewhat different model, the "Active teaching" model, was also developed which supported the

ANASTASIA PANAYIOTOU

increased involvement of children in the learning process and the focus of teacher on elements of academic interactions, such as asking questions and giving feedback.

Subsequent to the search for factors that are associated to student achievement and to the identification of their relationships which led in the past three decades to the development of the theoretical grounds of EER, in the fourth phase of EER, which is still present, the need for expanding country restricted knowledge has been stressed (Reynolds, 2000). Together with the advancements in the methodological approaches behind the studies in the field of EER and the realization that educational effectiveness cannot be considered as a static situation, but as a rather changing condition that adapts to different contexts, eras and student populations (Creemers, Kyriakides & Sammons, 2010), it has been gradually acknowledged that collaboration among researchers in various countries can contribute to the further progression of knowledge in the field of EER (Reynolds et al., 2014). National studies have also raised methodological concerns both regarding the small sample size which does not allow for safe conclusions to be drawn and also because of the mostly broad concepts examined; at least as far as the school factors are concerned (Reynolds et al., 1994). In addition, the methodological advancements of the last few decades which have led to the development of multilevel modeling techniques have led researchers to broaden their views on the functioning of education and realize that uni-level and uni-dimensional viewing of educational effectiveness can be seen as rather restricting and apart from reality.

Methodological Advances in EER

As mentioned in the previous section, the early studies conducted, especially in the first two phases of EER, have been understandably criticized not only on conceptual, but also on methodological grounds relating to the methodological design used and the sampling and data analysis procedures followed. For example, the two studies of Coleman

ANASTASIA PANAYIOTOU

et al. (1966) and Jencks et al. (1972), the results of which have led to great reactions and had an important influence on research in the field of EER, were based on traditional multiple regression techniques. In addition, even though they had a very large number of schools as part of their sample and were able to collect a variety of information in regard to school and student level variables, their research design had some significant weaknesses, since these studies were cross-sectional and were thus, not able to observe school effects and changes over time (Goldstein, 1997). Similarly, in spite of the fact that the work of Rutter, Maughan, Mortimore and Ouston (1979) had a longitudinal design and was able to demonstrate that differences in school effectiveness exist, the sample of the study was very small since the number of participating schools was only 12. The measurement methods of studies such as the one of Edmonds (1979) have also been criticized since the collection of data through self reporting and/or unstandardised instruments is not thought to provide an accurate measurement method of factors (Coe & Fitz-Gibbon, 1998).

Moreover, as stressed by Luyten, Visscher and Witziers (2005), concerns have been expressed by several researchers (e.g., Coe & Fitz-Gibbon, 1998; Goldstein, 1997; Goldstein & Woodhouse, 2000; Reynolds, Hopkins, & Stoll, 1993; Scheerens & Bosker, 1997) as to the extensive use of quantitative methods which are not in a position to allow for a more in depth investigation of certain aspects; and to the preference of cross-sectional studies which are not in a position to provide information on more complex relations, such as indirect, reciprocal, and curvilinear relationships as well as differential effects. Cross-sectional approaches can only provide a "snapshot' of a school at one point in time" (Reynolds et al.1993, p. 51) which opposes to the aim of observing the contribution of schooling over the course of time.

Therefore, in the past couple of decades, researchers in the field of EER have been extensively discussing the importance of using value-added approaches in order to measure the contribution of schooling on student achievement, as well as the weaknesses of the

ANASTASIA PANAYIOTOU

cross-sectional design (Luyten, Tymms & Jones, 2009). This criticism placed on the early studies described above, as well as the criticism on the general methodological approaches followed in the field of School Effectiveness Research (SER) and – more broadly – EER, has progressively led to advancements in the research design, the sampling methods and the data analysis procedures, giving room for more accurate and reliable estimations and for more in depth investigations.

It has previously been argued that researchers have come to realizing the multilevel composition of education, where students are nested in classrooms, classrooms are nested in schools and schools are nested in educational systems (Creemers & Kyriakides, 2008; Hill & Rowe, 1996). Thus, it has been acknowledged that a significant element in the methodological development in regard to research on the classroom and school effects on student outcomes, lies on the development of models and longitudinal studies that take into consideration the complexity and hierarchical structure of most educational systems and on the use of methods for analyzing such multilevel data (Burstein, 1980; Reynolds et al., 2014).

However, despite the fact that the nested nature of education has gained recognition even from the 1980s, one of the major problems researchers had to address was the lack of statistical methods of data analysis which would allow them to deal effectively with nested data (Hill & Rowe, 1996). During the last two decades, an important progress was made in the development of statistical packages such as HLM (Raudenbush & Bryk, 1991, 2002), Mlwin (Rabash et al., 2000), Mplus (Muthén & Muthén, 2008) and Stata (Rabe-Hesketh & Skrondal, 2008), which allow the analysis of hierarchical data and provide a basis for large-scale studies with the collection of a variety of rich, longitudinal data from large sample populations. In fact, the use of multilevel approaches has led to the demonstration of not only the fact that schools and teachers can explain a significant amount of variance in student outcomes, providing further and more methodologically advanced support to the

ANASTASIA PANAYIOTOU

results of earlier studies, but was also able to support the assumption that teacher factors have a larger effect than school factors (Reynolds et al., 2014).

Additionally, the use of Structural Equation Modeling (SEM) for the analysis of data which has gained recognition especially among the social sciences in the last decade, allows the identification of both indirect and reciprocal effects (Reynolds et al., 2014) and increases the credibility of results (Bowen & Guo, 2011). SEM can also be used for the measurement of the construct validity of the instruments used in the studies of EER for the measurement of effectiveness factors. Establishing the construct validity of the instruments used can be the answer to the criticism that was placed in the earlier studies of EER as far as the quality of data the instruments used were able to offer (e.g., Coe & Fitz-Gibbon, 1998). Furthermore, recent developments in the software used for multilevel analysis have allowed researchers to not only be in a position to identify relations among factors that operate at the same level, but also at the different levels of education, by using multilevel SEM models (Creemers, Kyriakides & Sammons, 2010).

The opportunity provided to researchers with the revolutionary advances in the methodological aspects of EER to understand and – as to a certain degree – decode the complex nature of educational effectiveness, has promoted not only research into more in depth aspects of education and the search for more complex relations among factors, but also theory development. Specifically, the application of more prominent methodological approaches has allowed the shift from the more simplistic earlier studies which led to the identification of separate, simply correlated factors, such as Edmonds' (1979) 'five factor model', to the more recent multilevel models of educational effectiveness.

The following section provides a description of some of the most influential theoretical models in the field of EER which led to the development of the current multilevel models that present both direct and indirect relations among different factors and student outcomes.

ANASTASIA PANAYIOTOU

Theoretical Advances in EER: Modeling Educational Effectiveness

The previous sections of this chapter provided a historical overview of the developments in the field of EER which led from the original pessimistic results of Coleman et al. (1966) and Jencks et al. (1972), to showing that education can in fact explain variance in student outcomes; and also provided a summary of the advancements made in regard to the methodological aspects of EER during the past three decades.

However, apart from the methodological limitations, one of the main reasons for which EER has been heavily criticized in the past years, lies on the absence of an efficient theoretical framework upon which research efforts could be based (Reezigt, Guldemond & Creemers, 1999; D'Haenens, Van Damme & Onghena, 2010). This lack of theoretical models has led researchers to the random selection of factors and to the mere establishment of statistical relations among them instead of the development and further testing of theories that could explain those relationships (Creemers, 1997; Creemers & Kyriakides, 2006; Kyriakides, 2012). The problem also seems to be magnified by the infrequent use of existing models, since most studies in the field of EER can be considered as atheoretical (Bosker & Scheerens, 1994; Creemers, 2002b; Scheerens, 1993; Scheerens & Bosker, 1997; Scheerens, 2014). Theoretical models are needed so as to assist researchers by summing up the relationships identified among factors through previous research in a simple and easily understandable manner. This way, acquired knowledge can be identified and acknowledged, and researchers may avoid conducting already existing research (Creemers & Kyriakides, 2012).

In this context, the criticism made to the early studies of EER which led to developments as far as the methodology used, has inevitably led to progress also on the theoretical foundations of the field. Specifically, more accurate definitions were provided to the concepts used and cleared relations between them were identified (e.g., Mortimore et al., 1988; Scheerens, 1992; Levin & Lezotte, 1990; Teddlie & Reynolds, 2000). Below,

ANASTASIA PANAYIOTOU

some of the most influential models in the field of EER which led to the recognition of the complexity and hierarchical nature of education are presented and elaborated on.

Educational Effectiveness Models

One of the most influential models of educational effectiveness that can be considered as a starting point to modeling educational effectiveness is Carroll's model (1963) which has also provided a basis for the application of large-scale educational innovations such as the Beginning Teacher Evaluation Study of Denham and Leiberman in 1980 (Anderson, 1984). The model originates from the work on foreign language learning that showed that persons with low aptitude needed longer time for learning a specific criterion in comparison to persons with high aptitude and consists of five basic classes of variables that would account for variations in school achievement (Carroll, 1989).

The five classes of variables based on Carroll's (1963, 1989) description are:

- *Aptitude:* variables that determine the amount of time a student needs to learn a given task under optimal conditions of instruction and student motivation;
- *Opportunity to learn*: the amount of time allowed for learning;
- *Perseverance:* the amount of time a student is willing to spend on learning the task or unit of instruction;
- *Quality of instruction:* when the quality of instruction is sub-optimal, the time needed for learning is increased;
- *Ability to understand instruction*, e.g., language comprehension, the learners' ability to figure out independently what the learning task is and how to go about learning it.

Generally, as indicated in Figure 2.1, the Carroll model assumes that achievement can be seen as the ratio between the time spend by students on a learning activity and the time

ANASTASIA PANAYIOTOU

actually needed by them considering time, quantity and quality as important elements for learning.

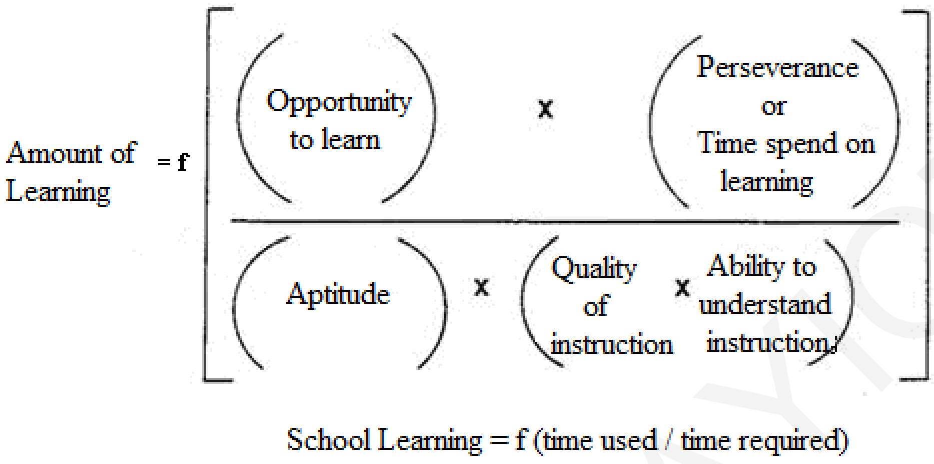


Figure 2.1. The Carroll model (1963)

In spite of the fact that the Carroll model is considered one of the leading models in the field upon which many of the later models and theories have been based and it has been provided with empirical support, it has received considerable criticism on the basis of not presenting predominant constructs such as 'quality of instruction' in more detail. In fact, 25 years after the development of the model, Carroll himself recognized the vagueness of the term "high-quality instruction" stating that it only mentions "that learners must be clearly told what they are to learn, that they must be put into adequate contact with learning materials, and that steps in learning must be carefully planned and ordered" (Carroll, 1989, p. 26).

However, the relation between time, perseverance, aptitude and quality of instruction has been further analyzed by Bloom (1976), who was based on the main elements of Carroll's model to develop the theory of "mastery learning". The rationale behind Bloom's theory led to the development of some of the later, multilevel models of EER such as the models of Creemers (1994), Scheerens (1993) and Slavin (1996).

Another model that was based on Carrolls' model and on the main factors that comprised it was the Walberg model (1984). Walberg's model (1984) was also part of a series of educational productivity models, aiming to estimate the size of effects of educational inputs on outputs which has expanded Carroll's model by adding a supplementary category of environmental variables and most importantly, by demonstrating relations among the different effectiveness factors. Specifically, Walberg's model consists of nine factors that can be grouped into three broader categories, which are considered to contribute to the acquisition of higher affective, behavioral and cognitive outcomes. These three categories and nine factors are:

A) *Student aptitude*

1. Ability or prior achievement; as indicated by standardized tests
2. Development; as measured by chronological age or stage of maturation
3. Motivation or self-concept; showing students' perseverance or willingness to engage with learning tasks

B) *Instruction*

4. Quality of instruction
5. Amount of instruction

C) *Environmental factors*

6. The home
7. The classroom social group
8. The peer group outside the house, and
9. The use of out-of-school time (Walberg 1984).

Figure 2.2 provides an illustration of the above factors as well as the relationships between them.

ANASTASIA PANAYIOTOU

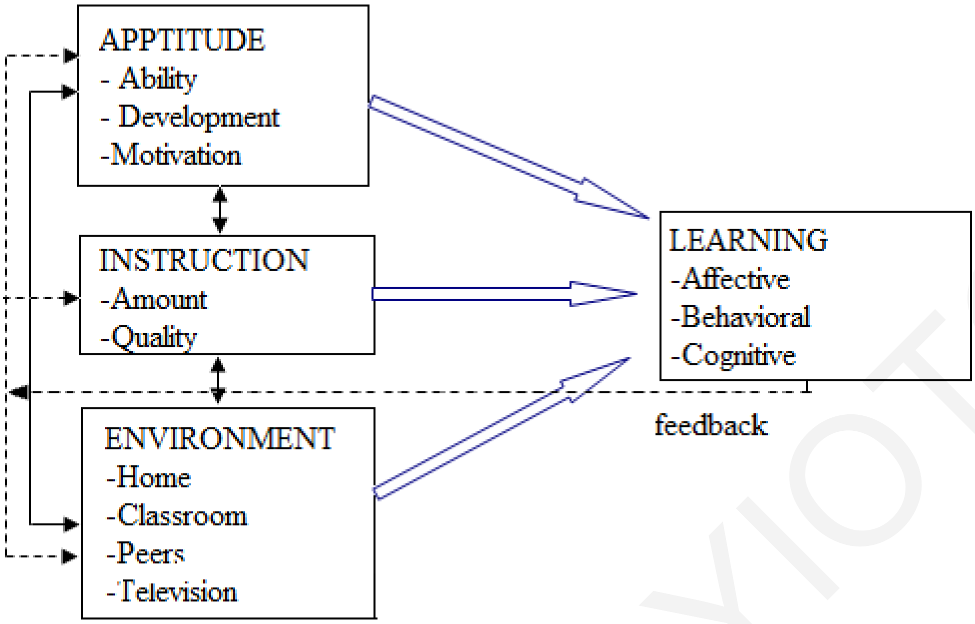


Figure 2.2. Walberg's model (1984)

Even though a large number of theories proposed have not been sufficiently supported through empirical studies in order to provide linkage between research and practice (Walberg, 1986), the validity of Walberg's model has been examined through a variety of studies and has demonstrated complex, indirect relations among the proposed factors (Reynolds & Walberg, 1992). These results reveal the need for establishing a theoretical framework in EER which takes into account the possible relations among factors operating both at the same and also at the different levels of education (Creemers & Kyriakides, 2008).

Taking the above into consideration, a model of educational effectiveness that attempts to identify and investigate the relations among factors within and across levels, is the Comprehensive model developed by Creemers in 1994 on the basis of Carroll's (1963) model (Kyriakides, 2005). Specifically, Creemers (1994) extended Carroll's model by further elaborating on the factor of quality of instruction which is seen as one of the central elements of his model. Thus, Creemers (1994) places emphasis on the factors associated with effective teaching which can be considered as the core of his model, dealing with both the teaching and learning processes. While earlier studies were focused on school effectiveness, investigating factors that operate at the school level, such as educational leadership and an orderly and safe school climate (Edmonds, 1979), in later research it has been acknowledged that classroom factors had greater effects on achievement (Postlethwaite & Ross, 1992). Specifically, factors such as time for learning (the time offered to students to engage in learning) and opportunity to learn (the amount of learning) were given focal attention and included in effectiveness models such as Creemers' Comprehensive model (van der Werf, Creemers, de Jong & Klaver, 2000).

However Creemers (1994) distinguishes between time and opportunities offered to students for learning, and time and opportunities actually used by students, supporting that quality of instruction can ultimately play a substantial role in the use that students make of time and opportunity (Kyriakides, Campbell & Gagatsis, 2000). His model also shows that the factors included in the different levels of education (i.e., school, classroom and student) are associated with student outcomes by indicating both direct and indirect relations. Specifically, it is assumed that the factors operating at the student level, such as the time and learning opportunities used by the students as well as motivation, aptitudes and social background are directly related with achievement, whilst factors operating at higher levels have indirect effects through their influence on time and opportunities used by students. Even though Creemers (1994) has taken into account the criticism placed on Carroll's (1963) model in regard to the specificity of quality of instruction and has provided a more elaborate description, it has been discussed that the motivation factor included in Creemers' Comprehensive model (1994) at the student level lacks equivalent specificity (de Jong et al., 2004).

Additionally, neither Carroll nor Creemers assume that motivation can be influenced by teaching behavior or school factors and neither also refers to any other student personal characteristics for which previous research has shown a relationship with

ANASTASIA PANAYIOTOU

student achievement (Kyriakides, 2005). However, an aspect that differentiates Creemers' model from Carroll's is that Creemers has considered students' social background as a variable that has a possible effect on student achievement. This way it is acknowledged that even though SES is a factor that cannot be altered it has prevailing effects on student outcomes (Creemers & Kyriakides, 2008). SES effects on student outcomes have been consistently investigated through a series of studies in the field of EER and different aspects of SES are still being examined in more recent studies (Acosta & Hsu, 2014; Lindo, 2014; Kieffer, 2012; Strand, 2014).

Moreover, Creemers' Comprehensive model (1994) is based on four main principles which enlist it as one of the most improved and influential models of its time of development. These principles as described by Creemers and Kyriakides (2008) are: *a) consistency, b) cohesion, c) constancy* and *d) control*. *Consistency* is based on the assumption that individual factors cannot contribute effectively to student learning when applied apart from and without reference to one another. On the other hand learning is better promoted when the factors operating at the different levels support and complement each other. *Cohesion* refers to the similar application of the school policies by all school members and to the stability of the policies throughout the year. Sustaining cohesion then leads to *Constancy* which refers to the maintenance of high quality instruction throughout a student's academic course and prerequisites consistency and cohesion. Finally, *Control* places emphasis on the evaluation processes of both student outcomes and also teaching behavior.

The validity of Creemers' model (1994) has been examined through several national studies (i.e., de Jong et al., 2004; Driessen & Sleegers, 2000; Kyriakides et al., 2000; Kyriakides, 2005; Kyriakides & Tsangaridou, 2004; Reezigt et al., 1999) and a secondary analysis of TIMSS 1999 (Kyriakides, 2006). These studies were in a position to demonstrate the effects of the main factors included in the model on student outcomes and

ANASTASIA PANAYIOTOU

reveal that the Comprehensive model can provide a solid basis for the development of the theoretical grounds of EER.

Taking in mind the basic assumptions of Creemers' model which, unlike the early models of EER, aim to call attention to the fact that the influences on student achievement are multilevel, and also the results of the validation studies that have provided support to the main factors included in the Comprehensive model, the Dynamic Model of Educational Effectiveness (Creemers & Kyriakides, 2008) was developed. An overview of the dynamic model of educational effectiveness, which provided the theoretical basis of the current study, is presented in the following section.

The Dynamic Model of Educational Effectiveness: moving a step forward

One of the most recent theoretical models which takes into account the criticism placed on the earlier models in the field of EER and which incorporates the findings of studies conducted in regard to the factors that have an influence on student outcomes, is the Dynamic Model of Educational Effectiveness (Creemers & Kyriakides, 2008), which was developed based on Creemers' Comprehensive model (Creemers, 1994a). This model takes into account the new goals of education, which means that apart from its reference to the cognitive outcomes of schooling, it also refers to other outcomes, such as affective, psychomotor and new learning (i.e., metacognition). Additionally, the dynamic model is multilevel in nature, since various effectiveness studies supported the basic assumption for the multilevel character of education (Teddlie & Reynolds, 2000) and also supported the fact that multilevel models may enhance the explanatory power of research in the area of educational effectiveness (Scheerens, 1993; Stringfield & Slavin, 1992). Thus, the dynamic model refers to effectiveness factors that operate at four different levels - at the level of the student, classroom, school and system - focusing, however, on the classroom and school level. The dynamic model emphasizes the factors operating at the classroom and school

ANASTASIA PANAYIOTOU

level, recognizing on one hand the direct effects of teachers' instructional behavior on student outcomes and on the other hand, the mainly indirect effects of the school factors, through their effect on the classroom level factors.

The dynamic model was developed taking in mind the weaknesses of the previous educational effectiveness models and the main outcomes of educational effectiveness studies conducted to test the validity of earlier effectiveness models – and especially Creemers' (1994) model (Kyriakides, 2008). One of the main weaknesses of the educational effectiveness studies is the fact that they do not refer to the measurement of each effectiveness factor, implying that the factors can be measured using only one dimension (i.e., frequency). Unlike this implication, the dynamic theory of education considers effectiveness factors as multidimensional constructs (Kyriakides & Creemers, 2008) and proposes the following five measurement dimensions which are assumed to provide more information concerning not only the quantitative aspects of the factors, but also the qualitative: a) frequency, b) stage, c) focus, d) quality and e) differentiation (Creemers & Kyriakides, 2008). The five measurement dimensions will be further elaborated on in the next section of this chapter.

Another weakness of EER is the exclusive attention given on the cognitive outcomes of schooling and especially on mathematics and language, which fails to examine students' progress across the full range of the curriculum and to relate with the new goals of education, as the development of metacognitive skills (Campbell, Kyriakides, Muijs, & Robinson, 2003). However, a series of studies which were not restricted to the sole examination of cognitive outcomes and/or on the fore mentioned two subjects (e.g., Creemers & Kyriakides, 2010a; Knuver & Brandsma, 1993; Kyriakides, 2005, 2006c; Kyriakides & Tsangaridou, 2008; Opdenakker & Van Damme, 2000), demonstrated that it is possible to reliably measure student outcomes at a broader extend.

ANASTASIA PANAYIOTOU

In addition, many of the concepts used in the previous models (e.g., Levine & Lezotte, 1990; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Edmonds, 1979) are not defined in detail and the relations between these concepts are not demonstrated. On the contrary, the dynamic model gives emphasis on providing a clear description of the quality of teaching through eight factors included at the classroom level, and the factors operating at the school level, and assumes that there are relations between factors operating both at the same and different levels. Such relations were also demonstrated through earlier models such as Walberg's (1984) who indicated that aptitude, instruction and the psychological environment influence one another and are also influenced by feedback on the amount of learning that occurs. As mentioned earlier, studies testing Walberg's model demonstrated more complex relationships (Reynolds & Walberg, 1990) which were taken into account in the development of the dynamic model, showing that factors that have an effect on teacher and school effectiveness can be grouped. However, apart from the grouping of factors, more complex assumptions are made for the functioning of the effectiveness factors included in the dynamic model, such as that the relationship of some factors with student outcomes may not be linear, but curvilinear (Campbell et al., 2004; Kyriakides, 2007; Scheerens & Bosker, 1997).

The search for non-linear relations may only be accomplished through longitudinal international studies as it may be assumed that only studies that are conducted using an international sample can provide enough variance and statistical power allowing for the demonstration of more complex relations. Even though significant progress has been made in the measurement of the factors at the different levels of education, which has led to a more precise understanding of the functioning of the factors associated with effectiveness, further international research is needed to examine whether the effects on student outcomes are causal, linear, or non-linear (Creemers et al., 1998). Therefore, as will be further described in Chapter 3, this study will be based on the assumptions of the dynamic model

ANASTASIA PANAYIOTOU

of educational effectiveness and examine the relationship of the teacher and school factors included in the model with student achievement, attempting not only to identify generic factors, by also to search for curvilinear relations and differential effects; providing additional support to the main assumptions of the model.

Finally, another criticism made in the earlier theories of EER, regarded their practical use and the possibility of using their basic principles for policy development in education (Creemers & Kyriakides, 2012). It can therefore be claimed that the dynamic model was designed in such way that can be used not exclusively for research and theory purposes, but also for promoting improvement in education (Carlo, 2012; Creemers & Kyriakides, 2015; Savage, 2012; Sucharita, 2013). The practical use of the model for improvement purposes, both at the classroom and school level, has already been demonstrated through several experimental studies which gradually provided support to the notion that the dynamic model can be used as a basis for policy development in educational contexts (Antoniou, 2013; Antoniou & Kyriakides, 2011, 2013; Antoniou, Kyriakides & Creemers, 2011; Creemers & Kyriakides, 2010b; Kyriakides, Archambault & Janosz, 2013; Kyriakides, Creemers & Antoniou, 2009; Kyriakides et al., 2014).

The significance of the classroom level through Teacher Effectiveness Research

Both researchers and theoreticians have continuously dealt during the past decades with the investigation of the effects of the teacher and school factors upon student outcomes (Teddlie & Liu, 2008). This focus and the long history of research into effective teaching has shown that the classroom level has a more immediate and direct effect on student achievement than the school level (e.g., Caldwell & Spinks, 1993; Creemers, 1994a; Muijs & Reynolds, 2000; Scheerens, 1992; Yair, 1997), which was shown to have mostly indirect influences on student performance (Creemers & Kyriakides, 2008). Additionally, the meta-analysis conducted by Marzano and colleagues showed that the

ANASTASIA PANAYIOTOU

effects were stronger in primary schools and weaker in the upper levels of education (i.e., middle and high schools) (Haystead & Marzano, 2009). Therefore since teacher behavior in the classroom was found to be in a position to explain a significant amount of variance at the classroom level (Muijs & Reynolds, 2011), apart from the extensive research and discussion around teacher factors, they have also been incorporated in the theoretical models developed in the past two decades (e.g. Creemers, 1994a; Creemers & Kyriakides, 2008).

Research during the past 35 years has led to the demonstration of a number of teacher factors that are positively related to student outcomes (e.g., Brophy & Good, 1986; Creemers, 1994a; Doyle, 1986; Galton, 1987; Muijs & Reynolds, 2000). Such factors include management of the classroom, expectations of student performance, teacher objectives, structuring of lessons, questioning skills, and immediate exercise after presentation, as well as evaluation, feedback, and corrective instruction (Creemers, 1994a). Management of the classroom is linked with opportunity to learn and time on task which have been consistently found to positively influence learning (Brophy & Good, 1986; Muijs & Reynolds, 2003), whilst dealing with student misbehavior is related to the time students are actually engaged with learning tasks. Along with dealing with student misbehavior, research in the field of teacher effectiveness has indicated that the creation of a well structured and orderly climate in which interactions among students are encouraged and learning occurs effortlessly through the maximization of student collaboration and elimination of excessive competition among students, can contribute to maximizing student gains and attainment (Muijs & Reynolds, 2000).

It is important to note that these factors do not stem from only one theory of learning such as the direct or active teaching approach (Rosenshine, 1983; Brophy & Good, 1986), but reflect a wider spectrum of teacher behaviors that incorporates characteristics of both approaches. Moreover, in spite of the fact that isolated studies were

ANASTASIA PANAYIOTOU

in a position to demonstrate the effect of the classroom level on student learning, meta-analyses conducted which have provided a synthesis of findings in the field of TER, have also identified specific teacher factors that have an influence on student outcomes. Specifically, the thorough meta-analysis of Scheerens and Bosker (1997) provided evidence showing that factors such as reinforcement of content and feedback have the strongest impact on student outcomes, followed by the factors of cooperation, differentiation, adaptive instruction and time on task (Muijs et al., 2014).

Likewise, Seidel and Shavelson (2007) conducted a meta-analysis expanding however their search to not only cognitive outcomes but also motivational-affective and learning processes. The results of their meta-analysis showed that domain-specific learning activities had the strongest effects on all three outcomes, while social experiences and time for learning were also found to have the highest effect sizes for affective outcomes and learning processes. Nevertheless, the largest synthesis of studies which is comprised of over 800 different meta-analyses is the one carried out by John Hattie as described in his book: "Visible Learning" (Hattie, 2009). The results of this dominant meta-analysis reinforced previous findings in the field of TER not only regarding the factors that have an effect on student outcomes, but also in regard to the overall size of effects of the teacher level factors on student learning. Specifically, Hattie's synthesis verified previous findings in the field of TER, showing that factors such as classroom behavior, teacher clarity, teacher–student relationships, cooperative learning and classroom management can be held accountable for student learning; and also added to previous research by pointing out that assessment, problem-solving skills and meta-cognitive strategies can also be seen as important elements of student learning (Muijs et al., 2014). Taking into account the results of research in the field of TER, as well as the results of the dominant meta-analyses conducted in the field, the dynamic model which aims to provide a complete framework upon which educational improvement efforts can be based, gives special emphasis to the

ANASTASIA PANAYIOTOU

classroom level factors and includes factors that derive from different approaches. The eight factors that are included at the classroom level of the dynamic model of educational effectiveness are presented and elaborated on in the following section.

The classroom level factors of the Dynamic Model

Drawn from the main findings of TER (e.g., Brophy & Good, 1986; Darling-Hammond, 2000; Doyle, 1990; Muijs & Reynolds, 2000; Rosenshine & Stevens, 1986; Scheerens & Bosker, 1997), the dynamic model refers to the following eight factors that describe teachers' instructional role and are associated with student outcomes: a) orientation, b) structuring, c) questioning, d) teaching-modeling, e) application, f) teacher role in creating a learning environment, g) management of time, and h) classroom assessment. As aforementioned, these eight factors do not derive from a single approach of teaching such as the constructivist approach (Von Glasersfeld, 1995; Duffy & Cunningham, 1996; Schoenfeld, 1998) or the direct teaching approach (Joyce, Weil & Calhoun, 2000) but they are in line with the new theories of teaching (Creemers, 2007). A short description of each teacher factor follows based on the definition of the factors provided by Creemers & Kyriakides (2008).

A) Orientation: Refers to teacher behavior in providing the students with explanations in regard to the reason(s) for which a particular activity or lesson or series of lessons occur and/or actively involving students to the identification of the reason(s) for which a lesson includes a specific task. Through this process it is expected that the activities that take place during lessons and lessons in general, will become meaningful to students and consequently increase their motivation for participating actively in the classroom (e.g., De Corte, 2000; Paris & Paris, 2001). It is therefore supported that orientation tasks should take place in not only one part of the lesson but be evenly distributed among the different parts of a lesson or series of lessons (e.g., beginning,

ANASTASIA PANAYIOTOU

middle, and end). Moreover, in order to claim that orientation tasks can contribute to student engagement in the lesson, it is important that they are understandable and unambiguous to students and that when students' input is asked to identify the significance of being involved with a task, all student views are encouraged and taken into account.

B) Structuring: is a factor for which research in the field of educational effectiveness has had early indications in regard to its contribution to student learning. Even from the mid 80s attention was called to the fact that student learning is positively influenced when teachers actively present materials and structure them by: (a) beginning with overviews and/or review of objectives; (b) outlining the content to be covered and signaling transitions between lesson parts; (c) calling attention to main ideas; and (d) reviewing main ideas at the end (Rosenshine & Stevens, 1986). In addition, research has shown that student outcomes can be amplified when teachers provide them with summary reviews, as they are expected to contribute to the grouping and outlining of main points (Brophy & Good, 1986). The fore mentioned structuring tasks aim at assisting students' memorizing of information and developing links between the different parts of lessons, instead of dealing with them as isolated units. Finally, the structuring factor is not limited to the mere linkage among the different parts of lessons or series of lessons, but also refers to the gradual increase of the lessons' difficulty level which is expected to provide all students, irrespective of their abilities, with the opportunity to engage in the lesson's processes (Creemers & Kyriakides, 2006).

C) Questioning: Taking into account the fact that research has shown that teacher questioning skills are closely associated with student achievement; this factor was included in the dynamic model and is defined according to five elements. Firstly, it is noted that effective teachers are expected to not only provide a large amount of product questions which require students to respond in a single way, but also focus on expecting students to elaborate on their answers and provide details indicating the mental course they followed

ANASTASIA PANAYIOTOU

to reach their answer (i.e., by also posing process questions) (Askew & William, 1995; Evertson, Anderson, Anderson, & Brophy, 1980). Secondly, it is anticipated that teachers grant students with enough time to think before calling for their answers with the amount of time given depending on each question's level of difficulty. Thirdly, it should be established that the questions posed by the teacher are clear and easily understandable by the students so that no misconceptions or misinterpretations are caused. Fourthly, when posing a question the teacher should take in mind the students' given ability to respond, avoiding too difficult questions that will inevitably cause complete failure to respond (Brophy & Good, 1986). Finally, it is outlined that an important aspect of this factor is the way teachers deal with student responses. Specifically, correct responses should be acknowledged so that it is established that all students are aware of the correct answer at the end of the discussion. In case a student's answer is not fully correct then the teacher should acknowledge whatever part may be correct, and assist the student in discovering the correct answer or provide an improved response, through the provision of clarification or helpful guidelines (Rosenshine & Stevens, 1986).

D) Teaching-modeling: An aspect of education that has received increased attention in the last two decades is that of self-regulated learning (SRL) due to the extensive policy emphasis given on the achievement of the new goals of education (Muijs et al., 2014). However, even though SRL has received major attention in educational research (Winne, 2005), it has still received less attention than other aspects. Taking the above into consideration, the teaching-modeling factor which is related to SLR is included among the classroom level factors of the dynamic model. This factor anticipates that effective teachers are promoting students' use of strategies and/or development of their own strategies in order to address different types of problems (Grieve, 2010) and develop skills promoting active learning. Thus, depending on the problem addressed, teachers may follow two alternative approaches. The first approach concerns the teacher's presentation of a

ANASTASIA PANAYIOTOU

problem-solving strategy without asking for any student input. The second approach demands more active student participation and begins in a rather backward manner, since students are encouraged to describe ways of how they themselves would address a specific problem. Then the teacher is expected to make use of that information for promoting the idea of modeling and encourage the development of the students' own problem-solving strategies (Aparicio & Moneo, 2005; Gijbels, Van de Watering, Dochy, & Van den Bossche, 2006).

E) Application: It is assumed that providing students with practice and application opportunities can enhance learning outcomes (Borich, 1992). Learning new information cannot be a constant process, since according to the Cognitive Load Theory the working memory can only process a limited amount of information at each given time (Kirschner, 2002; Paas, Renkl, & Sweller, 2003). It is also argued that application tasks should not only constitute a repetition of the material that students were taught in classroom but should move a step forward adding more complex and mentally stimulating elements. Thus, application activities should provide the trigger for further knowledge, contributing to the linkage of the units taught in one lesson or series of lessons with the following. Effective teachers are expected to not only observe students engaging in application tasks, but also contribute by supervising their progress and providing them with corrective and constructive feedback (Brophy & Good, 1986).

F) The classroom as a learning environment: This factor as described in the dynamic model consists of five components: a) teacher-student interaction, b) student-student interaction, c) students' treatment by the teacher, d) competition between students, and e) classroom disorder. These five elements have not been included by chance since classroom environment research has evidence showing that these can be considered as important aspects of this factor. Specifically, the first two of these elements can be seen as important for measuring classroom climate (for example, see Cazden, 1986; Den Brok,

ANASTASIA PANAYIOTOU

Brekelmans, & Wubbels, 2004; Harjunen, 2012), while the other three elements refer to teachers' efforts to create a well organized and accommodating environment for learning in the classroom (Walberg, 1986). The dynamic model also supports that the types of interactions in a classroom and teachers' actions in promoting such interactions that relate to learning (on-task behavior) is what should be examined. The classroom can be established as a learning environment by looking at the teacher's behavior in developing and maintaining rules, and by ensuring student respect and co-operation.

G) Management of time: To address this factor the amount of time used per lesson for on-task behavior is investigated. It is anticipated that effective teachers are able to organize and manage the classroom environment reducing any purposeless loss of learning time, maximizing engagement rates. Thus, the main interest of this factor is whether students are on task or off task and whether the teacher is able to deal effectively with any kind of classroom disorder without wasting the teaching time. It is also important to investigate whether the teacher manages to decrease loss of time for different groups of students by taking into consideration their different learning needs and abilities (e.g., by allocating supplementing work to gifted students that finish work earlier than others).

H) Assessment: Assessment is seen as an essential and interconnected part of teaching (Stenmark, 1992). Especially formative assessment has been shown to be one of the most important factors associated with effectiveness at all levels, especially at the classroom level (e.g., De Jong, Westerhof, & Kruiter, 2004; Kyriakides, 2008; Shepard, 1989). In fact, several studies (e.g., Brookhart, 2001; Tunstall & Gsipps, 1996; Wiliam et al., 2004) as well as the meta-analysis of Bangert-Drowns, Kulik and Kulik (1991) regarding research on formative assessment, have shown that the frequency of assessment is related to students' academic achievement (Marzano, 2007). Therefore, the dynamic model places emphasis on student assessment and assumes that the information collected though assessment is expected to be used by the teacher for at least two reasons. The first

ANASTASIA PANAYIOTOU

reason is related to the identification of particular student needs so as to proceed with the provision of feedback and corrective measures where needed. The second reason lies on the teachers' self-evaluation since student results may reflect possible weaknesses in teaching practice and indicate areas for improvement. It is thus stressed that assessment data should be examined in terms of quality (i.e., whether they are reliable and valid) in order to promote the formative rather than the summative purpose of assessment. The dynamic model also examines whether teachers possess the necessary skills allowing them to respond efficiently to each of the main phases of the assessment process (planning/construction of tools, assessment administration, recording, reporting) (Black & Wiliam, 2009), stating at the same time the relations among them.

Table 2.1 provides an overview of the basic elements of the eight teacher factors translated into teacher behaviors (retrieved from Creemers, Kyriakides & Antoniou, 2013).

School Effectiveness Research

As described by Goldstein (1997) "the term 'school effectiveness' has come to be used to describe educational research concerned with exploring differences within and between schools" (p.369). It is therefore important to note that along with the search for factors that operate at the classroom level and are related to student achievement, the search for factors operating at the school level and may have a possible impact on student learning, explaining variation among schools was also given emphasis, leading School Effectiveness Research (SER) to meet a significant growth during the early 1980s (Luyten, Visscher & Witziers, 2005). Specifically, an effort was originally made to link, what can be seen as mostly structural school features, with students' cognitive outcomes (e.g., Rutter et al., 1979) which lead researchers to identify as possible influential factors, school climate and school culture (Hargreaves,1995).

ANASTASIA PANAYIOTOU

Table 2.1

The main elements of each teacher factor included in the dynamic model

Factors	**Main elements**
1) Orientation	a) Providing the objectives for which a specific task/lesson/series of lessons take(s) place; and b) challenging students to identify the reason why an activity is taking place in the lesson.
2) Structuring	a) Beginning with overviews and/or review of objectives; b) outlining the content to be covered and signaling transitions between lesson parts; and c) drawing attention to and reviewing main ideas.
3) Questioning	a) Raising different types of questions (i.e., process and product) at appropriate difficulty level; b) giving time for students to respond; and c) dealing with student responses.
4) Teaching modelling	a) Encouraging students to use problem solving strategies presented by the teacher or other classmates; b) inviting students to develop strategies; and c) promoting the idea of modeling.
5) Application	a) Using seatwork or small group tasks in order to provide needed practice and application opportunities; and b) using application tasks as starting points for the next step of teaching and learning.
6) The classroom as a learning environment	a) Establishing on task behaviour through the interactions they promote (i.e., teacher-student and student-student interactions); and b) Dealing with classroom disorder and student competition through establishing rules, persuading students to respect them and using the rules.
7) Management of time	a) Organising the classroom environment; and b) Maximising engagement rates.
8) Assessment	a) Using appropriate techniques to collect data on student knowledge and skills; b) analysing data in order to identify student needs and report the results to students and parent; and c) evaluating their own practices.

ANASTASIA PANAYIOTOU

Even though the concept of school climate came earlier than school culture in the forefront and was dealt as a rather meaningful way in providing clues as to what works in education (Reynolds & Teddlie, 2000), it was supported that the two terms are closely interrelated and are possibly referring to the same concepts (Hoy, 1990; Denison, 1996; Glisson, 2000). Thus, in later research these two concepts are more broadly referred to as the School Learning Environment (SLE) (Creemers & Kyriakides, 2008).

Taking into consideration the lack of conceptual clarity in many of the school factors investigated in the early 80s and the broad definitions provided (Van Houtte & Van Maele, 2011), researchers aimed at identifying more specific factors that could be associated with student achievement and explain variation. Thus, several studies followed, aiming at achieving consensus so as to what constitutes school effectiveness (Holdaway & Johnson, 1993), in spite of the criticism placed in regard to the weaknesses in the methodology followed and the restricted focus of investigation (e.g., Clark, Lotto, & Astuto, 1984; Mackenzie, 1983).

In particular, the criticism placed on the early studies of SER, as summarized by Mortimore (1993), was based on issues such as the need for theory building, the use of more elaborate statistical techniques of data analysis, the need for improvement in the sampling methods used, the choice of appropriate outcomes and the enhancement of methods allowing to relate outcomes to process data. Even though the discussion on the methodological weaknesses in SER has led to advancements, the need for theory development in the field was also recognized by a number of researchers (Mortimore, 1991; Reynolds and Packer, 1992; Scheerens, 1992, 1993; Slater & Teddlie, 1992; Thrupp, 2001).

One of the first models that led the way for theory development in SER was the model developed by Edmonds (1979) which referred to five distinctive characteristics of effective school performance: (a) high expectations of instructional effectiveness among

ANASTASIA PANAYIOTOU

staff; (b) strong leadership by the principal; (c) an orderly, quiet and work-oriented atmosphere at school; (d) an emphasis on academic activities and development and (e) frequent monitoring of student achievement. However, unlike the equivalent process-product studies that were conducted in the same year (e.g., Brookover, et al., 1979; Rutter, et al., 1979) Edmonds' model could be considered as a rather "static model lacking a dynamic quality" (Slater & Teddlie, 1992, p. 243).

Later, Scheerens and Creemers (1989b) challenged the work of earlier uni-level models such as the one of Edmonds and argued that models of school effectiveness and improvement should contain at least two levels (school and class) with student background variables as a third "context level". In an attempt for identifying factors that can be held accountable for school effectiveness and could contribute to theory building, several studies were conducted that revealed factors which could be seen as indicators of effective schools. Such factors which received relatively large consensus among researchers were: (a) focus on school-based management, (b) strong instructional leadership, (c) stability of staff, (d) goal consensus, (e) school-wide staff development, (f) parental support, (g) approval of academic success, (h) effective use of time, (i) district-level support (j) collegial relationships and planning and (k) organizational commitment (Purkey & Smith, 1983).

However, it should be acknowledged that studies have shown that the school level has a relatively small effect on students' achievement since speaking in terms of variance the school level explains approximately 0.10 to 0.30, depending on variables such as educational level (i.e., primary and secondary) and subject domain (Guldemond & Bosker, 2009). However, factors operating at the school level are expected to have mostly indirect effects on student outcomes through their influence on the classroom level factors (Creemers & Kyriakides, 2008; Kyriakides, Creemers, Antoniou, Demetriou & Charalambous, 2015).

ANASTASIA PANAYIOTOU

Taking the above into consideration, along with the criticism placed on SER on the basis of not granting enough attention on the expansion of its theoretical grounds (Mortimore, 1992; Reynolds, Sammons, De Fraine, Townsend, & Van Damme, 2011); the dynamic model which is considered to be "the most up-to-date multilevel model of educational effectiveness" (Scheerens, 2013, p.10), attempts to encapsulate the main findings of SER and refers to factors operating at the school level that are considered to be related to student learning.

Moreover, as also mentioned in an earlier section of this chapter, even though steps towards theory building have been made in the last three decades, a large number of school effectiveness studies do not make reference to any theories upon which to be based. Specifically, a study conducted by Nordenbo et al. (2009) revealed that the field of SER does not make sufficient use of the theories developed, especially in comparison to TER, since out of 111 empirical school effectiveness studies only 23 were found to be based on a theory or conceptual model. Thus the dynamic model of educational effectiveness places emphasis not only at the lower levels of education and the two main actors of the teaching and learning processes (i.e., teacher and student), but also aims to identify school factors that have effects on student outcomes; thus providing a complete framework upon which research in EER can be based. The factors that are included at the school level of the dynamic model of educational effectiveness are presented in detail in the following section.

The school level factors of the Dynamic Model

In regard to the school level, the dynamic model refers to two main aspects of school policy which are expected to have both direct and indirect effects on student outcomes: a) school policy for teaching and b) school policy for creating a learning environment in the school. Additionally, since it is expected that schools that wish to

ANASTASIA PANAYIOTOU

become or remain effective should take actions to assess the existing policies and take corrective measures where needed (Hopkins, 2001; Kyriakides & Campbell, 2004), the dynamic model also examines the actions taken to evaluate the school policy for teaching and the school learning environment (SLE). The factors concerned with the school policy mainly refer to whatever actions the school may undertake to facilitate clear understanding among teachers and other stakeholders in regard to what is expected from them to do. Support offered to teachers and other stakeholders to implement the school policy is also an aspect of these two school factors.

Taking the above into consideration the following four overarching factors at the school level as described by Creemers and Kyriakides (2008) are included in the model:

1. school policy for teaching and actions taken for improving teaching practice,
2. evaluation of school policy for teaching and of actions taken to improve teaching,
3. policy for creating a SLE and actions taken for improving the SLE, and
4. evaluation of the SLE.

In regard to the school factor concerned with teaching, the dynamic model is concerned with aspects of school policy for teaching associated with:

a) quantity of teaching,

b) provision of sufficient learning opportunities, and

c) quality of teaching.

Actions taken for improving the above three aspects of teaching practice, such as the provision of support to teachers for improving their teaching skills, are also taken into account. Specifically, the following aspects of school policy on quantity of teaching are taken into account: a) school policy on the management of teaching time (e.g., lessons start on time and finish on time; there are no interruptions of lessons for staff meetings and/or for preparation of school festivals and other events), b) policy on student and teacher absenteeism, c) policy on homework, and d) policy on lesson schedule and timetable.

ANASTASIA PANAYIOTOU

School policy on provision of learning opportunities is measured by looking on whether the school has a policy concerning the provision of learning opportunities which is reflected in its policy on curriculum.

Additionally, the school policy on long-term and short-term planning and school policy on providing support to students with special needs is examined. It is also investigated whether any extra-curricular activities (such as school trips), are effectively exploited in order to promote student learning and eliminate loss of time. Finally, school policy on the quality of teaching is closely related to the classroom-level factors of the dynamic model, which refer to teacher behavior in the classroom and deals with the same key concepts. In addition, the school is expected to support teachers in their efforts of applying effective teaching practices.

In regard to the factor concerned with the school learning environment, the dynamic model investigates school policy on the following five aspects which define the environment of the school:

1. student behaviour outside the classroom,
2. collaboration and interaction between teachers,
3. partnership policy (i.e., the school's relations with the community, the parents, and the advisors),
4. provision of sufficient learning resources to students and teachers, and
5. values in favor of learning

The first three aspects refer to the rules which the school has developed for establishing a learning environment inside and outside the classroom. Here the term learning does not refer exclusively to student learning but to any form of action that endorses learning among the school stakeholders. For example, collaboration and interaction between teachers may add to teacher professional development and contribute to their pedagogical knowledge, consequently having an indirect effect on student learning.

ANASTASIA PANAYIOTOU

The fourth aspect refers to the school policy on providing resources for learning. The availability of learning resources in schools may not only have an effect on student learning but may also encourage the learning of teachers. The last aspect of the factor concerned with the SLE refers to measures taken by the school so as to evoke positive attitudes towards learning, both among students as well as among teachers. Figure 2.3 provides an overview of the Dynamic model of Educational Effectiveness as proposed by Creemers and Kyriakides (2008), where not only the factors included in each level are made apparent, but also the relationships assumed across and within levels; and their relationship with student outcomes.

Dimensions of measuring effectiveness factors as proposed by the Dynamic model

Finally, as suggested by Creemers and Kyriakides (2008), the functioning of each effectiveness factor, operating at either level, can be made cleared by looking at not only quantitative characteristics but also qualitative. Most studies in the area of EER attempting to measure the impact of specific factors on student learning have paid exclusive attention to their quantitative characteristics (i.e., the number of times an activity, task or action takes place in an educational setting), treating them as rather uni-dimensional constructs (Creemers & Kyriakides, 2008). Thus, the dynamic model proposes five measurement dimensions for examining each effectiveness factor: *frequency, focus, stage, quality*, and *differentiation*.

- *Frequency* is a quantitative means of measuring the functioning of each effectiveness factor which corresponds to the measurement methods used in previous studies (Scheerens & Bosker, 1997). However, as mentioned above, the sole use of this dimension is not in a position to provide a clear view of the way a factor is associated with student outcomes, since it does not provide researchers with the opportunity to have a more in depth analysis of its complex functioning.

ANASTASIA PANAYIOTOU

The other four dimensions refer to the qualitative characteristics of the functioning of the factors revealing that effectiveness is more complicated that what assumed by previous theoretical models and studies.

- *Focus* can be defined by taking into account two different facets. The first one refers to the specificity of the activities associated with the functioning of a factor, namely whether they can be considered as specific in terms of solid activities or policies; or more general, in terms of not providing adequate details to the different stakeholders on the application processes of an activity. The second aspect refers to the purpose for which an activity takes place by looking whether an action aims at achieving one or several purposes.
- *Stage* is closely related to the time at which tasks associated with a factor take place. It is assumed that the application of a factor in only one point in time may not constitute an effective way of dealing with the factor in terms of increasing the positive effects resulting from its implementation. Thus, it is supported that the factors need to take place over a long period of time to ensure that they have a continuous direct or indirect effect on student learning.
- *Quality* refers to the properties of the specific factor itself, as they are discussed in the literature. For instance, in regard to the assessment factor included at the classroom level, as it is stated through literature, formative assessment is expected to be more beneficial to students than summative and facilitate both learning and teaching (Black & Wiliam, 1998; Hattie & Temperley, 2007; Wiliam, Lee, Harrison, & Black, 2004). Supporting these indications, a study conducted in Cyprus demonstrated that primary school teachers, who conduct assessment for formative reasons, are more effective in terms of promoting student learning outcomes (both cognitive and affective) than those who conduct assessment for summative reasons (Kyriakides, 2005). Therefore, when measuring assessment in

ANASTASIA PANAYIOTOU

terms of its quality dimension it is expected – among others – to observe that teachers make use of formative assessment providing students with constructive feedback that will facilitate and assist learning (Bennetts, 2005; Brookhart, 2004; Muijs & Reynolds, 2011).

- *Differentiation* refers to the extent to which activities associated with a factor are applied without any digression for all the subjects involved with it (e.g., all the students, teachers, schools) irrespective of their needs and/or abilities. It is expected that adaptation to the specific needs of each subject or group of subjects will increase the successful implementation of a factor and will ultimately maximize its effect on student learning outcomes (Creemers & Kyriakides, 2006). Especially when referring to the classroom context it is widely acceptable that students differ in terms of abilities and readiness, with the teacher however attempting to correspond to the different needs by teaching at an average level (Konstantinou-Katzi et al., 2013).

Teaching however without accommodating to the students' specific needs raises issues of not only regarding the quality of education offered, but most importantly of the equal opportunities that the students receive during their school life (Ainscow, 2010; Klees & Qargha, 2014; Kyriakides & Creemers, 2011; Muijs et al., 2011). In fact, taking in mind that students learn best when their teachers become accustomed to the differences in their readiness levels, interests and learning needs and make an effort to adjust their teaching in order to satisfy them (Tomlinson, 2005), the need for examining the functioning of the different factors in terms of differentiation is amplified.

ANASTASIA PANAYIOTOU

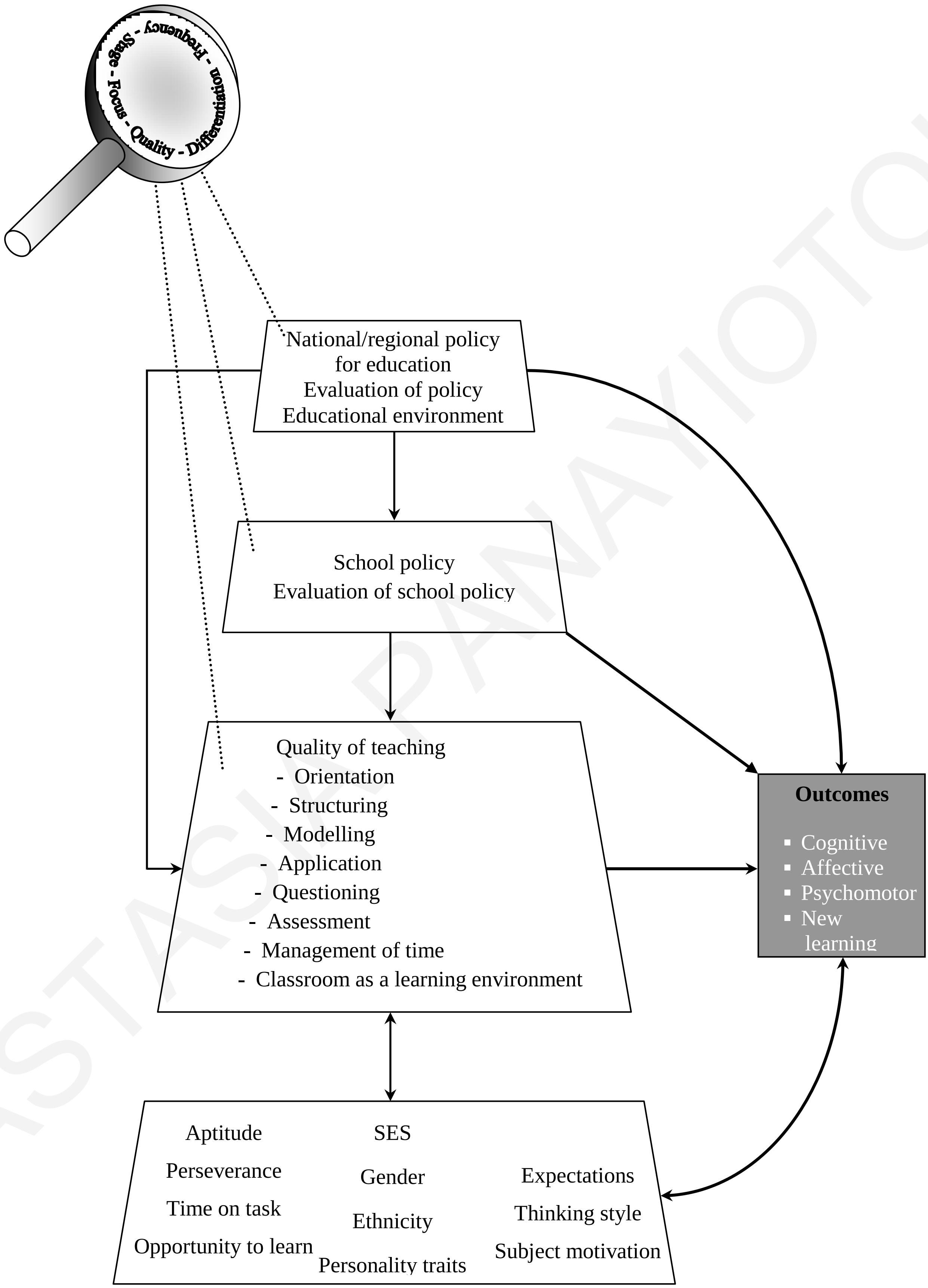


Figure 2.3 The dynamic model of educational effectiveness

National studies supporting the validity of the Dynamic model

So far, only national studies have been conducted to examine the validity of the model at classroom and school level and provide empirical support as for the use of the five dimensions for measuring each factor (Creemers & Kyriakides, 2008). Specifically, five national longitudinal studies testing the effects of classroom and school level factors upon achievement of both cognitive and affective outcomes have attempted to provide empirical support for the validity of the model (Creemers & Kyriakides, 2010a; Kyriakides, Archambault, & Janosz, 2013; Kyriakides & Creemers, 2008, 2009; Kyriakides, Creemers, & Antoniou, 2009). The first study conducted by Kyriakides and Creemers (2008), focused on providing support to each of the five measurement dimensions in regard to the classroom factors included in the model, in an attempt to test their association with student achievement. Providing support to the use of different aspects in regard to the measurement of factors assumed to have an effect on student outcomes is considered important in terms of ensuring consensus among studies in the field of EER and avoiding contradicting results depending on the focus of each study (Kyriakides & Creemers, 2008). Additionally, the study aimed at providing support to the generic nature of the model as far as the different outcomes of schooling (cognitive and affective).

The sample of this study comprised of 50 primary schools in Cyprus and 2503 students. Data on student achievement in three different subjects (i.e., mathematics, Greek language, and religious education) were collected, while data on quality of teaching were collected through external observations and questionnaires administered to students. Even though the study provided support for the construct validity of the five measurement dimensions of most effectiveness factors at the classroom level, due to the national character of the study significant restrictions arise as far as testing whether the model can be considered generic not only for the different goals of education, but also for different

ANASTASIA PANAYIOTOU

educational contexts. Similarly, another limitation of this, as well as other national studies, is the fact that no conclusions can be drawn as far as the differential effects of the factors in different contexts.

The second attempt to verify the importance of the classroom factors and determine their generalizability in two different subjects (i.e., language and mathematics), was through two national studies conducted in Cyprus using data from primary and pre-primary school students (Kyriakides & Creemers, 2009). An essential difference of this national study with the one conducted by Kyriakides and Creemers in 2008 was that it aimed not only to provide support to the generic nature of the model but also to test the differential effects of the effectiveness factors included in the model in different groups of students, teachers and schools since one of the basic criticism of the educational effectiveness models was that a description of teacher behavior following the "one size fits all" approach does not reflect reality in education (Campbell, Kyriakides, Muijs, & Robinson, 2003). The findings of previous research into differential educational effectiveness revealed that there are substantial differences between different groups of students and that the effectiveness of a teacher or a school may vary across different factors, subjects as well as over time (Nutall, Goldstein, Prosser, & Rasbash, 1989; Sammons, Nuttall & Cuttance, 1993; Kyriakides & Tsangaridou, 2008). This study, which collected data from 76 pre-primary and 50 primary schools in Cyprus, was also in a position to reveal differential effects in separate groups of students acknowledging however, the need for further research to test the generalisability of the findings and to provide further explanation for the reasons some factors and dimensions appear to have differential effects (Kyriakides & Creemers, 2009). International longitudinal research may give further evidence not only on the differential effects of the factors in different age groups and subjects, but also provide answers as to the nature of the effectiveness factors across different educational contexts.

ANASTASIA PANAYIOTOU

The third longitudinal study aimed at providing further support to the construct validity of the dynamic model focusing both on the classroom and school level factors as a response to the criticism that could be made to the model as far as its complexity, in comparison to earlier models of effectiveness, and the difficulties occurring in the process of its empirical testing (Creemers & Kyriakides, 2010a). Thus the study investigated the effects of the classroom and school factors on different outcomes of schooling (cognitive and affective), under the light of the five measurement dimensions. The results of this study added to the validity of the model and proved that in spite of the criticism in regard to its complexity it can be put to the test and produce satisfactory results (Creemers & Kyriakides, 2008). The contribution of the classroom and school factors was again identified, however at national context. This kind of methodology has significant limitations as again it is not in a position to produce results concerning the generic nature and the functioning of the factors across countries and therefore does not allow for safe suggestions to be made for improvement purposes in different countries than the one of the study.

Another national study was conducted (Kyriakides, Creemers, & Antoniou, 2009) this time aiming to examine the assumption that the teacher factors of the dynamic model are interrelated and can be grouped so as to be used for the improvement of teaching skills and consequently for the improvement of student outcomes. Thus, to test the assumption concerned with the grouping of factors this study collected data from 50 Cypriot schools. All year 5 students (n=2503) from each class (n =108) of the school sample were chosen to participate in the study. The results of the study demonstrated that the teaching factors of the dynamic model and their dimensions can be grouped into 5 developmental stages moving from more easily acquired teaching skills (such as the structuring of lessons) to more demanding types of teaching behavior such as differentiation of the different aspects of the teacher factors (e.g., differentiation of assessment). Therefore this study was able to

ANASTASIA PANAYIOTOU

provide support to another assumption of the dynamic model regarding the possible relations among the factors included in the classroom level.

The assumption on the grouping of factors was yet further put to the test through another national study conducted in Canada (Kyriakides, Archambault, & Janosz, 2013). This study collected data from seven primary French schools in the suburb area of Montreal. All students of grades 3 - 6 (n=959) from each class (n=42) of the school sample were asked to complete a questionnaire concerned with the teacher factors and their dimensions. The results of this study provided further support to the results of the previous study conducted in Cyprus (Kyriakides, Creemers, & Antoniou, 2009) showing that the teacher factors are interrelated and can be grouped in the same developmental stages. Even though this study was able to provide evidence on the grouping of factors it was again national in character and thus the generalisability of these findings could be further examined through international studies.

Meta-analyses

Generally, a meta-analysis aims to provide a more holistic view of the findings of a large number of studies, so as to facilitate conclusions in regard to an area of research by attempting to provide an estimate of the statistical significance of a relationship (Kyriakides & Creemers, 2010). Even though, criticism, has been placed on meta-analysis on the basis of possible leading to oversimplified conclusions or of being influenced by design errors in the original research studies (Hedges & Olkin, 1985), statistical methods for dealing with such issues have been developed (see Glass, McGaw & Smith, 1981) which have led meta-analysis to become an important tool in comparison to traditional methods of research reviewing and theory development (Creemers & Kyriakides, 2008).

Thus, apart from the abovementioned national studies, an attempt was also made to get a more complete view of the functioning of the teacher and school factors through two

ANASTASIA PANAYIOTOU

recent meta-analyses (Kyriakides, Creemers, Antoniou & Demetriou, 2010; Kyriakides & Christoforou, 2011). These syntheses of studies provided some evidence on the generic nature of the factors but were not in a position to give information as far as the different measurement dimensions are concerned, since in the majority of the studies the only dimension used for measuring the effect of the factors on student achievement was frequency. Therefore, the need for further research which will be in line with the new methodological advances in EER is stressed, so that more qualitative information on the effectiveness factors may be collected.

Furthermore, in these studies, limitation on the specificity of some factors was identified. For example, the general factor of school evaluation was investigated without distinguishing between evaluation of the school policy on teaching and of the school learning environment (Kyriakides, Creemers, Antoniou & Demetriou, 2010). This more specific discrimination of the different aspects of evaluation may produce more specific results which could perhaps be used for policy making purposes. Even though meta-analyses contribute to the better understanding of the results of a plethora of studies conducted in different contexts (Hunter & Schmidt, 1995) and statistical advances have led to the effective dealing with issues such as extracting estimates of effect size from the studies, they are not always in a position to provide a basis for the further development and validation of a theoretical model due to limitations affecting its results, such as sampling or measurement errors (Creemers, Kyriakides & Sammons, 2010).

Therefore, international longitudinal studies may constitute a better basis for identifying generic factors which can be used for the development of educational policies across countries, aiming to improve teaching practices and ultimately enhance student outcomes.

ANASTASIA PANAYIOTOU

International Perspectives of EER

During the first three phases of EER, as described in the first section of this chapter, focus was mostly laid on the identification of factors that contribute to the quality of education through national studies conducted in specific countries. However, during the fourth phase of EER a progressive interest in international perspectives was observed because some of the factors that could explain variance within one country did not seem to have the same impact in other countries (Creemers, 2002a).

This gradually increased interest towards international research arises since it is acknowledged at a large extend that EER lacks cross-cultural perspectives and has been criticized for showing strong ethnocentric tendencies (Reynolds 2000). Although EER publications have acknowledged the influential studies by Coleman et al. (1966), Edmonds (1979), Brookover, Beady, Flood, Schweitzer, and Wisenbaker (1979), Rutter, Maughan, Mortimore, and Ouston (1979), and Mortimore, Sammons, Stoll, Lewis, and Ecob (1988), a science of EER has not been developed. The absence of cross-national perspectives and inter-cultural relationships between educational effectiveness researchers that may explain deviation of findings among countries is seen as intellectually damaging (Kyriakides, 2006a).

Specifically, over the last two decades, a trend has begun to emerge whereby education policy-makers have attempted to apply simplistic suggestions for raising standards based on the notion of 'transplanting' knowledge and ideas from one country, or one culture, to another. Examples of this practice as stated by Reynolds (2000) include the discussion in some American states regarding the lengthening of the school day and shortening of school holidays based on approaches followed in Japan and the trend in British primary schools towards whole-class direct instruction, which is based on approaches used in the Pacific Rim countries. Many EER researchers are voicing concern about the potential hazards of this practice of transplanting educational policies from one

ANASTASIA PANAYIOTOU

country or culture to another (e.g., Reynolds, 2006; Scheerens, 2013). The concerns are largely based on effectiveness studies that have shown how factors that seem successful in some countries may not be effective elsewhere. For example, some U.S. studies have found a positive association between 'assertive principal instructional leadership' and student achievement gains (e.g., Hallinger & Heck, 2011; Louis, Dretzke, & Wahlstrom, 2010), but within-country research in Europe (e.g., The Netherlands, Cyprus) has not found the same level of support for this factor (e.g., Kyriakides, 2008; Scheerens & Bosker, 1997). In addition, most research projects in the past decades have been carried out in countries with similar characteristics, such as decentralized educational systems, sufficient educational resources and common level of parent participation (Murillo & Rincon, 2002). These context similarities cause doubts on whether the results occurring from such studies can be easily understood and applied in fundamentally diverse situations.

Despite the fact that research in the first phases of EER has managed to provide evidence regarding the functioning of factors operating at different levels, the worries mentioned above in regard to the restricted range of research have started rising even as early as a decade ago (Murillo & Rincon, 2002). As found by Reynolds and colleagues (1994), the great majority of studies conducted in the field of EER have only taken place in eight countries (Australia, Canada, Great Britain, Hong Kong, Norway, Taiwan, the Netherlands and the United States). This finding provides clear implications in regard to both the practical as well as the theoretical aspects of EER, demonstrating possible bias in regard to the transfer of knowledge in countries with different social, economic and educational contexts.

Nevertheless, in times when educational policy appears to be following the international route, it is unfortunate that EER appears to be lingering in a uni-culture research dimension, pursuing within-country studies rather than cross-national research. Arguably, cross-national and multicultural studies on educational effectiveness are

ANASTASIA PANAYIOTOU

required in order to develop understanding about the complex structures of educational policy across different countries and cultures, and to explain how policies affect student outcomes in different settings. Reservations to conducting international studies may occur both because of the interest observed from policy-makers within countries to develop national policies promoting quality of education, as well as financial and practical limitations in conducting large-scale international educational effectiveness research (Creemers, 2006). However, it is assumed that an alteration in focus through a shift in across country studies may provide evidence regarding the consistency of outcomes found in national studies; which may evidently assist in answering the question of whether the observed effects consist a generic or specialized phenomenon (Reynolds et al., 2014).

Thus, the recognition of the importance of escaping national borders and expanding research in social sciences in order to meet advancements in other scientific fields (Reynolds, 2000), has led to the foundation of the International Association for the Evaluation of Educational Achievement (IEA) in the late 1950's (Postlethwaite 1995; Postlethwaite & Ross, 1992) followed by the foundation of the Organization for Economic Co-Operation and Development (OECD). IEA has initiated the international comparative studies of TIMSS and PIRLS and OECD has developed the international study of PISA; all aiming to move a step forward from the restrictive knowledge being offered through national studies and provide an equal basis for measuring student outcomes, thus promoting discussion among researchers (Drent, Meelissen & van der Kleij, 2013).

However, in spite of the fact that the increasing participation of countries in these international studies can be seen as an encouraging step for promoting international research, Beaton and Robitaille (2002) revealed that at least in regard to TIMSS, most secondary analyses conducted to date referred to the mere ranking of participating countries focusing solely on achievement results. This comparison between more and less successful countries through these international studies (for example see Danju, Miralay &

ANASTASIA PANAYIOTOU

Baskan, 2014; Grek, 2009) has led to confusions on how the results of such studies can be used. Namely, differences in outcomes between educational systems which inevitably produce competition among countries (Pereyra, Kotthoff & Cowen, 2011) are not in a position to equally reflect any social and other differences that could be held accountable for these differences observed in achievement.

In addition, as mentioned in the introductory chapter of this thesis, even though these international studies provide a common basis upon which different countries can be compared (Beaton et al., 1996), they portray several important limitations; the first one being their cross-sectional character which does not allow for comparisons to be made among the same students across-time. Specifically, it should be acknowledged that these comparative studies are neither in a position to allow comparisons between educational systems nor provide a basis for identifying causal effects (Goldstein, 2004). This is argued since in cross-sectional studies the direction of causality cannot be made explicitly clear, possibly leading to misinterpretation of results (Creemers, Kyriakides & Sammons, 2010).

Finally, it can be argued that even though the aim of these international comparative studies was to locate factors that are related to student learning and can be altered through proper interventions (e.g., curriculum changes, resource allocation, teaching methodology etc.) (Martin, 1996; Yang, 2003) certain factors have so far not been given enough attention. More specifically, the richness of data collected through these international studies, in regard to the factors operating at the classroom and school level is not considered sufficient to provide an insight as to the existing differences in school effectiveness in different countries (Reynolds, 2006).

History of International Longitudinal Studies

In this context, the first attempt for an international longitudinal study that took in mind the methodological and conceptual weaknesses of the other international studies was

ANASTASIA PANAYIOTOU

in 1991 with the International School Effectiveness Research Project (ISERP) (Creemers, Reynolds & Swint, 1994; Reynolds, Creemers, Stringfield, & Teddlie, 1998). This study was based on earlier findings which showed that factors which were found to be important for student outcomes in one context were not found to have an effect in countries with a different context, such as the role of principals whose importance was outlined by studies in the USA (Levine & Lezotte, 1990) in opposition to the results of European studies (Van de Grift, 1990). Therefore for ISERP nine countries with very different contexts were chosen to participate (i.e., USA, Canada, Australia, Hong Kong, Taiwan, The Netherlands, Norway, UK and Ireland). The ISERP study measured student achievement and student social outcomes and used direct methods for the measurement of quality of teaching (i.e., classroom observations).

However, even though it was clear that there was variation across countries (Reynolds, 2006), and through this study the importance for conducting longitudinal studies was amplified, due to practical difficulties in the implementation of the research design and the use of the instruments for the measurement of the classroom factors, further research is needed to investigate which effectiveness factors can be considered generic. Specifically, despite that the methods used for collecting data regarding the classroom factors could be considered innovative, it was recognized that collecting observational data in different countries is more demanding, time consuming and overall more difficult than other means, in terms of efficiently training observers that come from various contexts so as to ensure common understanding and reliability. One of the main contributions of this study was that through the lessons learned it was recognized that efficient methods of collecting data in regard to effectiveness factors in international studies should be seen as a prerequisite to ensure success. Thus, the need for establishing internationally valid observation instruments that are able to be interpreted the same way in diverse educational contexts (e.g., Taiwan and the UK) in order to capture a true image of teacher behavior

ANASTASIA PANAYIOTOU

was stressed and received as one of the most important lessons drawn from this international study (Teddlie et al., 2006).

Consequently, the attempt made through ISERP to conduct an international, longitudinal study and the questions raised from this first effort, provided the basis for a second international study aiming to address the main methodological weaknesses of this first study and shed more light to the factors that can be considered generic in different contexts. The second study aimed to develop an International System for Teacher Observation and Feedback (ISTOF) which would serve in not only assessing teacher effectiveness but also providing feedback to teachers for improvement purposes (Teddlie et al., 2006). Even though the rationale of ISTOF originated from the weaknesses of the previous, first large-scale longitudinal study (i.e., ISERP) the difficulties of using direct measures of teacher effectiveness so as to contribute to the development of a theory, adding to the further development of the field of EER, was again stated.

These early attempts to model effectiveness and to create a theory upon which educational improvement actions can be based, generated the need for further international research that would enhance the theoretical contributions that appeared in the beginning of the 90s and build on their assumptions in order to take them a step further (Creemers 1992; Creemers, 2006). Methodological and statistical advances as well as the increasing use of mixed methods are now providing researchers with the opportunity to decode differences in outcomes between countries (Sammons, 2006).

As a result, the international longitudinal study presented in this thesis aims to contribute to the identification of generic factors that are located at two different levels of education that are seen to have significant effects on student academic outcomes (i.e., the classroom and school level). Taking also in mind that the Dynamic model of educational effectiveness (Creemers & Kyriakides, 2008) can be considered as one of the most prominent theoretical models in the field (Sammons, 2009; Scheerens, 2013), and has also

ANASTASIA PANAYIOTOU

been tested and validated through national studies (Creemers & Kyriakides, 2010a; Kyriakides & Creemers, 2008, 2009) and meta-analyses (Kyriakides, Creemers, Antoniou & Demetriou, 2010; Kyriakides & Christoforou, 2011), this study aims at providing further empirical support to its assumptions and search for more complex relations (i.e., curvilinear) of some factors with achievement. Finally, the study aims to contribute to the establishment of internationally valid instruments for assessing teacher and school effectiveness as described in the dynamic model. The research design, the participants and the research methods used in this study are presented in the following chapter.

ANASTASIA PANAYIOTOU

CHAPTER 3

METHODOLOGY

This chapter presents the methodology used to examine the main research questions set as part of this study. Specifically, the research design employed is described in detail and its selection is being justified in opposition to other research designs. Then the processes of sampling and data collection are described in detail so as to provide a better insight to the methodology used. The different statistical techniques employed, during the analysis process to establish the validity and reliability of the data as well as to provide answers to the research questions posed in regard to the generic and differential nature of the teacher and school factors of the dynamic model are also described. Finally, possible limitations, related to the methodological design of the study, are discussed.

Justification of the Research Method Chosen

Longitudinal vs. Cross-sectional Research Design

As discussed in Chapter 2, the international studies that take place and collect data on several factors concerned with student background, achievement as well as factors that operate at different levels (i.e., teacher and school), face an important methodological limitation that does not allow for assumptions to be made in regard to the contribution of education; namely the effect of these factors upon student achievement gains. These studies follow a cross-sectional research design, meaning that they collect data at only one time point. Collecting data at only one time point may allow for comparisons among student populations and different countries/districts but not for more in depth analyses and/or investigating cause and effect relations. It is also not possible to examine changes

ANASTASIA PANAYIOTOU

over time and the way they are related with changes in the functioning of specific factors (Creemers, Kyriakides & Sammons, 2010).

Apart from the cross-sectional design we can also distinguish the longitudinal design which can be applied so as to conduct multiple measurements from the same sample of students, teachers etc. Generally studies in which data collection takes place in at least two time points can be considered as longitudinal even though it is considered that in order to perceive a better realization of what is being measured more than two measurements should take place (Gustafsson, 2010). Thus, when conducting a longitudinal study instead of a cross-sectional it is necessary to consider several design issues such as the number of measurements, the duration of the study and the size of the sample that will provide enough statistical power so as to identify the effects of the independent variables on the dependent (Moerbeek, 2008).

The selection and use of the two fore mentioned research designs depends primarily on the purpose of each study and of the specific research questions that different researchers aim to address. However it is generally accepted that even though the longitudinal research design demands greater resources and the commitment of the sample to participate at all stages of data collection (Galbraith, Stat & Marschner, 2002; Moerbeek, 2008), it portrays several advantages in comparison to the cross-sectional – especially in the field of education and educational effectiveness. As earlier mentioned by Goldstein, (1968, p. 98) "it should be noted that the main advantage of longitudinal over cross sectional studies lies in the efficiency of the estimation of change" thus allowing researchers to identify changes in the variables under investigation (e.g. student achievement) and determine causal effects. The value of longitudinal studies for the search of factors leading to development in skills and knowledge, and consequently the limitations of the cross-sectional design, have also been acknowledged by OECD (2015a). Of course it should be taken into consideration that experimental studies are the most safe

ANASTASIA PANAYIOTOU

and widely acceptable research method for examining causal relations since, among others, it is easier to determine whether changes in the variable that is considered as the cause may lead to changes to the variable that is considered the possible outcome (Cook, Shadish & Wong, 2008; Slavin, 2010).

Even though longitudinal designs may allow the realization as to which variable precedes, the main issue that arises in non-experimental studies is the possibility of having other, omitted variables which may affect both the dependent and the independent variables leading to misconceptions in regard to causal effects (Fox, 1987; Gustafsson, 2010). However, longitudinal studies may provide a basis for the identification of causality and also for determining the direction of causality (Cook & Campbell, 1979). Additionally, statistical advances have led to the development of methods of data analysis which allow researchers to draw conclusions in regard to causal effects through observational data (Heckman, 1976, 1979; Rosenbaum, 1986). Longitudinal studies may therefore be used to assist hypothesis development relating to the factors that may lead to differences in student outcomes or theory validation in regard to the factors that may be responsible for student achievement gains, wherever an experimental study may not be possible (e.g., when dealing with large samples, examining a different range of variables or having issues of random selection and allocation to groups) (Creemers, Kyriakides & Sammons, 2010).

Moreover, the question of whether effects observed are sustainable over the course of time can only be examined through longitudinal studies with at least three measurement points which may demonstrate their long-term impact. Recent studies in the area of teacher professional development raise this issue of whether the positive impact of one year of professional development can be sustained and specifically whether the duration of the program plays an important role in improving teaching skills (Panayiotou, Kyriakides & Christoforidou, 2015). Such questions could not have been answered merely by using a

ANASTASIA PANAYIOTOU

cross sectional research design since it would neither be possible to examine the effect of the intervention upon the improvement of teaching skills nor its stability.

Longitudinal research design was chosen for conducting the present study since one of its main aims was to examine the effect of the classroom and school factors included in the dynamic model of educational effectiveness on student achievement gains. Collecting data in more than one phase may allow seeing how the functioning of the factors may affect student progress in two different outcomes (i.e., mathematics and science) and in six different countries, testing their generic nature. However, it should be acknowledged that only two measurement points have been used in this study due to limited resources and the practical difficulties in including more measurements in a large-scale international study and data on the teacher and school factors were only collected at one time point. Therefore, due to this limitation, only direct relations could be examined as part of this study since in order to be easier to seek for the indirect effect of the school factors on student achievement at least three measurements are required. More than two measurements may allow to see how changes in school policy may affect changes in teaching practice and through that, student outcomes (Hendriks, 2014; Heck & Moriyama, 2010; Kyriakides et al., 2015).

In addition, even though as was discussed earlier in this section, randomized experimental studies are considered the highly preferred solution for causal inferences (Cook, Shadish & Wong, 2008), their selection instead of observational studies also lies on the specific purpose of each study. The present study aimed at providing further support to one of the most recent theoretical frameworks in the field of educational effectiveness (i.e., the dynamic model of educational effectiveness) at classroom and school level, attempting to validate its assumptions on the generic nature of the factors included in these levels. Observational longitudinal research can also provide valuable data allowing for theory building and validation, especially when a wide range of factors is under investigation

ANASTASIA PANAYIOTOU

(Slavin, 2010). Results from such studies may provide evidence upon which further experimental studies can be based as previous experience has shown (see Antoniou & Kyriakides, 2011; Demetriou & Kyriakides, 2012). An experimental design was therefore not chosen to provide a basis for this study due to the large amount of factors under investigation, the number of participating countries and the main aim of the study which is the further validation of an existing theoretical model at international level.

Research Design

The research design of the study involved two main steps. The first step comprised of actions in regard to the preparation of the study; namely the aim of this first stage was the development of the measurement instruments, their adaptation to the context of the six participating countries and their validation, as well as the selection of the sample. Then, the second step concerned the main study where the data collection took place. The data collection was realized in two phases. In the first phase which took place at the beginning of the school year 2010-2011 data were collected from the student sample (n=10742) in regard to their achievement in mathematics and science. To assess student knowledge in mathematics and science written forms of assessment were developed.

The second phase of the data collection process aimed to collect data on student final achievement in the two fore mentioned subjects so as to examine student progress over time and determine the effect of the teacher and school factors on student achievement gains. To examine teacher behavior inside the classroom questionnaires were also administered to the student sample. To collect information concerning the different aspects of school policy as they are described in the dynamic model, questionnaires were administered to the teacher sample (n=2923). The teacher and student questionnaires were administered at the end of the school year, rather at the beginning, so that the teachers on one hand would have sufficient time to come to term with all the aspects of their school's

ANASTASIA PANAYIOTOU

policy and also, students would have the time to get accustomed with their teacher's classroom behavior.

Finally, after the completion of the main study, the testing of the validity and reliability of the data collected and the data analyses were conducted. Across- and within-country analyses were conducted in order to identify the extent to which each classroom- and school-level factor is associated with achievement in each outcome revealing generic and differential factors operating at different levels. Table 3.1 presents the timeframe of the study.

Table 3.1

Study timeframe

Study steps	Timeframe	Actions
Step 1: *Preparation of the study*	January 2010	• Development of the student and teacher questionnaires (translation and back translation procedures).
	February 2010	• Final version of the questionnaires
	March – May 2010	• Construction of the achievement tests
	May 2010	• Sample selection
Step 2: *Main study*	September – October 2010	• Mathematics and science tests administration to the student sample (prior achievement data)
	April – May 2011	• Mathematics and science tests administration to the student sample (final achievement data) • Student questionnaire administration • Teacher questionnaire administration
Step 3: *Analysis of data*	September 2011 – January 2012	• Across- and within country CFA analysis: Student/ Teacher questionnaire data
	February – May 2012	• Across- and within country multilevel analysis
	June – August 2012	• Multivariate analysis

ANASTASIA PANAYIOTOU

Preparation of the Study

a) Development of the measurement instruments

Classroom Level Factors

Regarding classroom factors, all grade 4 students were asked to complete a questionnaire concerned with the behavior of their teacher in the classroom in regard to the eight factors of the dynamic model (see Appendix A). For the development of the student questionnaire, an adapted version of the instrument that was developed for the national studies mentioned in Chapter 2 (i.e., Kyriakides & Creemers, 2008; Creemers & Kyriakides, 2010a), which covered all eight factors and their dimensions was used. Specifically, with the use of a five-point Likert scale (Likert, 1931), students were asked to reveal the extent to which their teacher behaves in a certain way in their classroom. For example, an item concerned with the stage dimension of the structuring factor asked students to indicate whether at the beginning of the lesson the teacher explains how the new lesson relates to previous ones; another item asked whether at the end of each lesson the teacher spends time reviewing the main ideas of the lesson. Another example is an item that was used to measure the differentiation dimension of the factor concerned with the teacher's skills in questioning: "Our teacher praises all pupils the same when we answer a question correctly".

The original instrument was thoroughly discussed by the members of each country team, who were asked to express their views on the applicability and relevance of each item to their educational context and to evaluate whether young students in primary schools in their country would be in a position to adequately provide information in regard to each item. This process resulted in a substantial number of items being dropped from the original questionnaire. Specifically, items of the questionnaire which referred to teaching materials and/or strategies that are used in the country where the original study took place

ANASTASIA PANAYIOTOU

but are not necessarily used in the other five countries had to be removed to ensure that the questionnaire would be relevant to each context.

Additionally, some of the items measuring the differentiation dimension had to be removed which can most likely be interpreted by the fact that differentiation may be used or understood differently in each country. Specifically, in some countries the differentiation items were perceived as inconsistencies to teaching behavior, instead of being recognized as a response to the different educational needs of students. For example, in some countries it was considered positive for the teacher to provide more time to certain students to complete their assessment assignment (e.g., slow learners), whereas in countries with a more centralized system, teachers may not be enabled to differentiate their assessment. As a result, the new revised instrument was able to measure all eight factors, but not all five measurement dimensions of each factor (see Appendix B). In order to take into account the new questionnaire's inability to measure the five dimensions, the items of each factor were classified into two broad categories concerned with the quantitative and qualitative characteristics of the functioning of each factor.

The frequency and stage dimensions were considered as indicators of the quantitative characteristics of each factor, while the focus, quality and differentiation dimensions were seen as indicators of the qualitative characteristics of the factors (this classification of items was examined through CFA analysis and the factorial structure of the student questionnaire is presented in the next chapter). The working version of the questionnaire was in English; the next step was to translate and back-translate the questionnaire into four versions, i.e., Dutch, German, Greek, and Slovenian. To provide a better insight as to the items used in the student questionnaire, a representative item for each teacher factor is provided in Table 3.2.

ANASTASIA PANAYIOTOU

Table 3.2

Indicative items used to measure each factor in regard to the teacher factors of the dynamic model

Teacher Factor	Indicative items per factor
1) Orientation (n=1)	When doing an activity in Mathematics and Science I know why I am doing it.
2) Structuring (n=8)	When the teacher is teaching, I always know what part of the lesson (beginning, middle, end) we are in.
3) Application (n=7)	At the end of each lesson, the teacher gives us exercises on what we have just learned.
4) Management of Time (n=3)	There are times we do not have the necessary materials for the lesson to take place (e.g., dienes, unifix, test tubes, thermometers, calculators, rulers)
5) Questioning (n=8)	When we do not understand a question, our teacher says it in a different way so we can understand it.
6) Modeling (n=4)	Our teacher encourages us to find ways or tricks to solve the exercises or work s/he gives us.
7) Classroom as a learning Environment / Teacher – Student Interaction (n=6)	The teacher gives all pupils the chance to take part in the lesson.
8) Classroom as a learning Environment / Dealing with Misbehaviour (n=7)	When the teacher talks to a pupil after they have been naughty, sometimes after a while, that pupil will be naughty again.
9) Assessment (n=5)	When we are having a test I finish up within the time given to us.

ANASTASIA PANAYIOTOU

School Level Factors

At the end of the school year 2010-2011, data regarding the school level factors included in the dynamic model of educational effectiveness were collected. School factors were measured by asking all the teachers of the school sample to complete a questionnaire which was designed in such a way that information about the school-level factors of the dynamic model on school policy for teaching and on policy for improving the school learning environment could be collected (see Appendix C). To encapsulate school policy, teachers were asked to provide information about a number of activities that take place in their school in relation to the aspects of policy for teaching and policy for school learning environment mentioned in the previous chapter. For example, to measure the aspect of school policy on quantity of teaching, items were used relating to issues discussed during staff meetings (e.g., dealing with teacher and student absenteeism, amount and type of homework etc.).

Teachers were also asked to refer to the extent to which these issues were discussed in documents and other materials which are distributed by the school management team to school stakeholders (i.e., teachers, students, and parents). Teachers were both asked whether these issues were covered in the policy documents, and also whether the documents made clear to the teachers what they are expected to do. Finally, teachers were asked whether the school management team provides support to the teachers in order to help them implement the policy. The items measuring this factor and all the other factors discussed above have been used in previous studies testing the validity of the dynamic model of educational effectiveness. Through these studies, support to the construct validity of the teacher questionnaire has been provided (see Creemers & Kyriakides, 2010a). A representative item for each subscale is given in Table 3.3 (for specification table see Appendix D).

ANASTASIA PANAYIOTOU

Table 3.3

Indicative items used to measure each subscale in regard to the school factors of the dynamic model

Scales	Indicative items per subscale
Scale 1: School policy on teaching	
1) Quantity of teaching (n=8)	Discussions at staff meetings help me to improve my practice in making good use of teaching time
2) Provision of learning opportunities (n=12)	There is material on notice-boards in the school relevant to provision of learning opportunities beyond the ones provided by the formal curriculum
3) Quality of teaching (n=14)	At school staff meetings we discuss and take decisions on issues concerned with methods to effectively teach pupils (e.g., questioning, application, modeling)
Scale 2: Policy on the school learning environment (SLE)	
1) Student behavior outside the classroom (n=4)	In my school, we have taken the decision to organize fun activities during break time that may help students to achieve specific learning goals (e.g., games, dance, sports)
2) Teacher collaboration (n=4)	The teachers in my school cooperate with each other by exchanging ideas and material when teaching specific units or series of lessons.
3) Partnership: relation with the school community and parents (n=15)	In parent-teacher meetings organized by our school, we discuss ways in which parents can help to deal with the following issues: a) homework, b) ….

ANASTASIA PANAYIOTOU

4) Provision of learning resources to students and teachers (n=4)	There is material on notice-boards in the school relevant to: The use of different educational tools for teaching provided by the school
Scale 3: Evaluation of the school policy on teaching (n=11)	The principal and/or other members of the school staff observe the way the teaching policy is put into practice and presents the results of their observations to staff
Scale 4: Evaluation of the SLE (n=9)	My school keeps systematic records concerned with: a) the use of educational tools for teaching supplied by the school (e.g., maps, software etc.), b)…

A Likert scale was used to collect data on teachers' perceptions of the school level factors. Data on school factors are based on teacher questionnaires only and the limitations of using perceptual methods to measure school factors should be acknowledged. Nevertheless, using the G theory the generalizability of the data was found to be appropriate. In addition, reliability was computed for each subscale measuring school policy for teaching, policy for SLE and school evaluation showing that for each factor the Cronbach alpha was satisfactory (for all factors alpha was greater than 0.7). The findings of these analyses provided further support to the reliability and generalizability of the data emerged from the teacher questionnaire.

Student Achievement in Mathematics and Science

For the construction of the tests used to measure student achievement, permission was obtained from IEA to use the released items of TIMSS 2007. Because not all test booklets from the TIMSS 2007 cycle were released for public use, booklets 1 to 4 were administered for assessing student achievement in mathematics and science. In terms of TIMSS notation this meant that the test blocks M01 to M05 for mathematics and S01 to

ANASTASIA PANAYIOTOU

S05 were part of the achievement tests (Mullis, Martin, Ruddock et al., 2005, p. 101). The properties of each item and its relation with the curricula of grades 3 and 4 in each country were taken into account for developing the two types of test. However, because concerns were raised regarding the difficulty level of the TIMSS mathematics items for grade 3 students, the TIMSS test blocks were complemented by 13 items originating from the Belgian *Schoolloopbanen in het BasisOnderwijs* (SIBO) (see Appendix E). These SIBO items have previously been used in a Belgian longitudinal study (see Hendrikx, Verhaege, Ghesquière et al., 2006) and like the TIMSS items, have shown good fit indices.

Both types of tests (i.e., mathematics and science) comprised of tasks classified under different learning domains (i.e., knowledge, application, and reasoning), as well as different content domains. Tables 3.4 and 3.5 provide an overview of the distribution of test items by content and learning domain as classified in TIMSS 2007 for mathematics and science, respectively, as well as the SIBO items.

For the estimation of the achievement data in each of the two subjects, data scaling was conducted by applying the Partial Credit Model (Masters & Wright, 1997) as implemented in the software ConQuest (Wu, Adams & Wilson, 1998). As estimation algorithm an adaptation of the quadrature method described by Bock and Aitken (1981) was used. Thus, for each student two different scores for his/her achievement at national and international level in each subject at the beginning and the end of grade 4 were generated (Vennemann & Wendt, 2012). Prior to the estimation of the achievement scores several other steps took place so as to clean and recode the raw achievement data collected from each country. The task of data preparation can be differentiated into three steps: (1) recoding the data delivered by all participating countries, (2) cleaning the data, and (3) checking student IDs to assure unique usage. As part of the first steps of data cleaning, descriptive statistics by item were conducted and checked carefully, in that respect the response patterns were investigated within and across countries. The third important step

ANASTASIA PANAYIOTOU

of data preparation was to identify mismatching IDs or replicative usage of student IDs and duplicates had to be deleted.

Table 3.4

Classification of the TIMSS and SIBO Mathematics items by Content and Learning Domain

	Items			
	n		% (of total number of items)	
Content Domain	*TIMSS*	*SIBO*	*TIMSS*	*SIBO*
Numbers	31	10	46	18
Geometric shapes and measures	13	3	19	5,5
Data display	11	-	16	-
Learning Domain				
Knowing	17	5	31	9
Applying	23	8	42	15
Reasoning	15	-	27	-

Table 3.5

Classification of the Science items by Content and Learning Domain

	Items	
Content Domain	n	%
Life Science	24	40
Physical Science	23	38
Earth Science	13	22
Learning Domain		
Knowing	21	35
Applying	25	42
Reasoning	14	23

ANASTASIA PANAYIOTOU

b) **Selection of the research sample**

In each participating country (Belgium/Flanders, Cyprus, Germany, Greece, Ireland, and Slovenia), stratified sampling procedure (Cohen, Manion, & Morrison, 2000) was used to select a sample of at least 50 primary schools. All grade 4 students of the school sample participated in the study (n= 10742) and tests in mathematics and science were administered to them at the beginning and at the end of school year 2010-2011.). Students were also asked to complete a questionnaire concerned with the behavior of their teacher in the classroom in regard to the eight factors of the dynamic model included at the classroom level. Table 3.6 shows an overview of the sample of schools and students used to measure student achievement gains in mathematics and science and the effect of the teacher factors across countries.

Table 3.6
Overview of the sample of schools and students used to measure student achievement gains in mathematics and science and the effect of the teacher factors across countries

Country	**Schools**	**Students**
Belgium	50	1908
Cyprus	59	1899
Germany	54	1228
Greece	49	1128
Ireland	63	2424
Slovenia	59	2155
Total sample	***334***	***10742***

For the measurement of the school policy on teaching and the school learning environment, all the teachers of the school sample (n=2923) were given a questionnaire. However, in Greece this was not possible and less than four questionnaires per school were

ANASTASIA PANAYIOTOU

administered. For this reason, the data collected from the teacher questionnaires in Greece were not used for any analyses concerning the school factors and therefore the results on the impact of school factors on student achievement in the other five countries will be presented in the next chapter. In addition, schools with a teacher response rate below 30% were removed from the analyses. The percentage of 30% was chosen, since having at least 70% response rate per school was considered satisfactory. Table 3.7 shows the number of teachers and schools that were used in the multilevel analyses which were conducted in order to identify the impact of the school factors on student achievement. It is therefore shown that from the remaining five countries, 22 out of 265 schools in total were removed from the analyses due to low teacher response rates. Specifically, 5 schools were removed from Belgium, 2 schools were removed from Cyprus, 4 schools were removed from Germany and 11 schools were removed from Ireland. In Slovenia no schools had to be removed as for all the schools the response rate was greater than 30%. In total, data concerning the school factors were collected from 2888 teachers of the sample.

Table 3.7

Number of teachers and schools per country from which data were collected and number of teachers and schools per country used in the analyses concerned with the impact of school factors on student achievement

Country	**Data collected**		**Data used in the analyses**	
	Teachers	**Schools**	**Teachers**	**Schools**
Belgium/Flanders	535	45	523	40
Cyprus	294	59	291	57
Germany	363	45	358	41
Ireland	342	57	327	46
Slovenia	1389	59	1389	59
Total sample	***2923***	***265***	***2888***	***243***

ANASTASIA PANAYIOTOU

Analysis of Data

a) Generalizability analysis

After the cleaning and preparation of the data, a Generalisability Study on the use of students' and teachers' ratings was separately conducted (Shavelson, Webb & Rowley, 1989; Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Generalizability Theory was applied since data have been collected by students and teachers so as to draw conclusions on factors belonging to a higher level (i.e. teacher and school, respectively). As stated by Shavelson, Webb and Rowley (1989, p.922), "a score's usefulness, largely depends on the extent to which it allows us to generalize accurately to behavior in some wider set of situations, a universe of generalization".

Firstly, in regard to the results on the use of students' ratings, the ANOVA analysis showed that the data can be generalized at the classroom level, as for all the items of the questionnaire, the between-group variance was higher than the within-group variance ($p<0.05$). Between-group variance is considered to be related to the independent variable by demonstrating a systematic variance; on the contrary within-group variance is not related to group differences and is considered to be a result of error. Likewise, a Generalisability Study on the use of teachers' ratings showed that the data obtained through the teacher questionnaires can be generalized at the school level, as again for all the items of the questionnaire, the between group variance was higher than the within group variance ($p<0.05$).

b) Confirmatory Factor Analysis

Having in mind the categorization of the student and teacher questionnaire items as presented in the two respective specification tables (see Appendices B & D), separate Confirmatory Factor Analysis (CFA) was conducted for each of the teacher and school

ANASTASIA PANAYIOTOU

factors of the dynamic model which were measured by the student and teacher questionnaire respectively, by using the EQS software for Structural Equation Modeling (Bryne, 1994). CFA was used, as the objective was to test whether the data fit a hypothesized measurement model (Harrington, 2008); in this case the assumptions of the dynamic model in regard to the two broader dimensions of each teacher factor and the school factors. Additionally, CFA was applied since the sample size was greater than 200 both in the case of the student and the teacher questionnaire data which is considered satisfactory for most models (Kline, 1998). For both questionnaires, two sets of CFA were conducted: across countries (i.e., using the full dataset) and within countries (i.e., separate analysis for each country).

The Construct Validity of the Student Questionnaire

The results of the across-country CFA confirmed the construct validity of the questionnaire. Although the scaled chi-square was statistically significant, the values of RMSEA were smaller than .05 and the values of CFI were greater than .95, thus meeting the criteria for an acceptable level of fit. Moreover, the standardized factor loadings were all positive and moderately high, ranging from 0.48 to 0.84, with most of them higher than 0.65.

In the case of the time management factor, exploratory factor analysis was conducted and provided satisfactory results. CFA was not used to test the validity of the questionnaire measuring this factor, as there were only three items measuring the frequency dimension and the one-factor model was just identified (i.e., degrees of freedom=0). The results of the EFA showed that the first eigenvalue was equal to 1.40 and explained almost 50 per cent of the total variance, whereas the second eigenvalue was less than 1 (i.e., 0.81). These results showed that these three items could be treated as belonging to one factor, especially since they had relatively high loadings (i.e., >.67).

ANASTASIA PANAYIOTOU

The within-country CFA analyses revealed that nine of the 49 questionnaire items had to be removed, in order to keep items with relatively high factor loadings. Specifically, four items measuring the differentiation dimension of the eight factors and five of the negative items had to be removed. Finally, the items concerned with the classroom as a learning environment were found to belong to two different one-factor models which measured the type of interactions in the classroom and teacher's ability to deal with student misbehaviour.

Since one of the main assumptions of the dynamic model is that the teacher factors are interrelated (see Kyriakides, Creemers, & Antoniou, 2009), the next step of the analysis of data was to examine how these effectiveness factors were related to each other. The assumption was that the factors concerned with: a) management of time, b) teacher ability to deal with student misbehaviour, and c) the quantitative dimension of the questioning factor (i.e., measuring the extent to which teachers raise appropriate questions and avoid loss of teaching time) belonged to one second-order factor, whereas the other factors could be grouped together as another second-order factor. This assumption was initially tested by conducting across-country SEM analysis and later by conducting six separate within-country analyses. The results of both the across- and the within-country analyses are presented in Chapter 4.

The Construct Validity of the Teacher Questionnaire

Following the same approach as with the student questionnaire, an attempt was made to examine the construct validity of the teacher questionnaire. The CFA models which were conducted for the school factors, showed that some of the items that were included in the teacher questionnaire had to be removed from the analyses due to low loadings on the corresponding factors. Specifically, 30 out of the 81 items that were included in the questionnaire had to be removed during the CFA analysis. These items

ANASTASIA PANAYIOTOU

belong to four categories which are: a) items concerned with staff meetings, b) items measuring the differentiation dimension, c) negative items and d) items concerned with school participation in programs/projects. Some explanations of the possible reasons why some items had to be removed from the analyses are presented in more detail below and the removed items are classified based on assumptions about the reasons for which they might have had to be removed.

a) Items concerned with staff meetings:

As mentioned, the results from the analyses showed that some of the items concerned with staff meetings had to be removed. A possible explanation could be that the questionnaire was developed in Cyprus where schools have staff meetings every week and that may not be the case in some countries, were staff meetings may not take place that often and therefore decisions about the school policy may not be taken at the staff meetings.

b) Items measuring the differentiation dimension:

Moreover, as it resulted from the CFA analyses all the items measuring the differentiation dimension had to be removed. This can probably be explained by the fact that differentiation dimension is not used in all the countries the same way and also by the fact that it is difficult to be located.

c) Negative items:

The one negatively worded item that was included in the questionnaire (item 35) had to be removed for which the translation may have caused some problems. A similar problem with the negative items and the items measuring the differentiation dimension was also found in the student questionnaire measuring the teacher factors.

d) Items concerned with school participation in programs/projects:

Items concerned with school participation in projects (i.e. Comenius etc) had to be removed. This may be due to differences to the system in some countries as perhaps in

ANASTASIA PANAYIOTOU

some countries it is more common for schools to participate in projects than others. For instance, schools in more centralized systems are not expected to have a large participation in projects whilst in more decentralized systems schools are expected to have a greater participation. The items that remained in each of the school factors in the CFA models are presented in Table 3.8.

For factors concerned with teacher collaboration and relation with community, the CFA results are not presented in Table 3.8 as only 3 items remained measuring each factor and the one-factor model is just identified (i.e., its degrees of freedom are 0). Therefore, for teacher collaboration and relation with community exploratory factor analysis was conducted with satisfactory results. Specifically, for teacher collaboration the first eigenvalue was equal to 1.57 and explained more than 50% of the variance whereas the second eigenvalue was much smaller than 1 (i.e., 0.77). These results show that we can treat these three items as belonging to one factor especially since all three items had relatively high loadings (i.e., higher than 0.69). For relation with community the first eigenvalue was equal to 1.83 and explained more than 60% of the variance whereas the second eigenvalue was again much smaller than 1 (i.e., 0.65). In addition, all three items had relatively big loadings (i.e., bigger than 0.73) which shows that we can treat these three items as belonging to one factor.

In the case of Quality of teaching, items 8I, 8J, 8K, 8L, 8M, 8N,8O were grouped as one variable as they all concerned aspects of quality of teaching through staff meetings and they had a high correlation (greater than 0.5). Likewise, in the case of evaluation of school policy on teaching items 32A, 32B and 32C were grouped as one variable and items 5A, 5B and 5C were grouped as another variable. For the factor concerned with the evaluation of the SLE, items 5F and 5G were also grouped as one variable to be used in the CFA analysis.

ANASTASIA PANAYIOTOU

At the next step, SEM analysis was conducted to see whether the items of the teacher questionnaire can be grouped according to the assumptions of the theoretical model upon which they are based (i.e., the dynamic model). Separate SEM analyses were therefore conducted for the three overarching factors included at the school level of the dynamic model: a) School policy on teaching, b) Policy on the School Learning Environment and c) Policy on Evaluation. The main attempt was to develop three models for these overarching factors based on the data from all the countries and then to replicate these models in the within country analyses. The results of both the across- and the within-country analyses are presented in Chapter 4.

c) Reliability Analysis

The reliability of each scale measuring the teacher and school factors was also separately calculated using Cronbach's alpha. Cronbach's alpha was used since "it is a test reliability technique that requires only a single test/questionnaire administration to provide a unique estimate of the reliability for a given test/questionnaire" (Gliem & Gliem, 2003, p.84). Thus, reliability analysis was conducted and the Cronbach alpha was calculated for the entire scale of the student questionnaire (49 items included in Part A of the questionnaire). The results of the reliability analysis showed that the Cronbach alpha was satisfactory (a= 0.79). In addition, the calculation of the value of the Alpha "if item deleted" revealed that none of the items had to be removed.

Then, the reliability of each scale measuring the teacher factors was also calculated and the results show that for each factor the Cronbach alpha was relatively satisfactory (given the small number of items of each subscale). As mentioned by Cronbach, (1990) and Gliem and Gliem, (2003) even though the values of Cronbach alpha range from 0 to 1 and a high value for Cronbach's alpha (i.e., close to 1) indicates good internal consistency of the items in the scale, one should take into consideration the number of items in each

ANASTASIA PANAYIOTOU

scale whose reliability is being tested, since it is expected to affect the value of Cronbach alpha. The only exception was the factor of questioning where Chronbach's alpha was small (a<0.5). In addition, the calculation of the value of the "Alpha if item deleted" revealed that none of the items had to be removed from each factor. The results of the reliability analysis per factor, across countries are presented in Table 3.9.

The same approach was followed to measure the reliability of teacher responses in the questionnaire measuring the school factors of the dynamic model. First, the reliability of all the items of the teacher questionnaire was computed (n=81) and Cronbach's alpha was found to be very high (a= 0.97) (George & Mallery, 2003). Then reliability was computed for each subscale measuring school policy for teaching, policy for the SLE and school evaluation showing that for each factor the Cronbach alpha was satisfactory (for all factors alpha was greater than 0.7). In addition, the calculation of the value of the "Alpha if item deleted" revealed that none of the items had to be removed from each factor. The results of the reliability analysis per factor, across countries are presented in Table 3.8.

d) Multilevel Analysis

Through research in the field of educational effectiveness and the criticism placed upon the earlier studies which led to the formation of uni-level theoretical models, such as Edmond's (1979) 5 factor model, researchers have come to the realization that it is unrealistic to not take into consideration educational systems' hierarchical structure (Rhoads, 2011). To better capture the complexity of education and of the relations among the different factors affecting student outcomes it is a prerequisite to address the educational systems, both methodologically as well as theoretically, as multilevel (Creemers, 1994b). Thus, with respect to the multilevel character of education, methodological advancements have led to the development of relevant statistical software aiming to address the clustered nature of data such as HLM (Raudenbush & Bryk, 1991,

ANASTASIA PANAYIOTOU

2002), Mlwin (Rabash et al., 2000), Mplus (Muthén & Muthén, 2008) and Stata (Rabe-Hesketh & Skrondal, 2008).

Table 3.8

Items of the CFA models and across-country results of the SEM and reliability analysis for the teacher questionnaire measuring school factors

School Factors		SEM and Reliability Analysis Results
Scale 1: School policy on teaching		
1) Provision of Learning opportunities	8F, 8G, 15B, 22B	One – Factor Model: (X^2= 12, df= 1 CFI=0.995, RMSEA= 0.063) a = 0.70
Scale 2: Policy on the school learning environment (SLE)		
1) Partnership Policy	21A, 21B, 21C, 21D, 23, 26, 27	One – Factor Model: (X^2= 78, df=8 CFI=0.987, RMSEA= 0.055) a = 0.78
2) Provision of sufficient learning resources	8E, 11C, 15E, 22D	One – Factor Model: (X^2= 16, df= 2 CFI=0.993, RMSEA= 0.051) a = 0.71
Scale 3: Policy on Evaluation		
1) Evaluation of the school policy on teaching	5ABC, 31, 32ABC, 33, 34, 37	One – Factor Model: (X^2= 47, df= 5 CFI=0.994, RMSEA= 0.054) a = 0.84
2) Evaluation of the learning environment	5FG,36, 38, 40, 41, 42, 43	One – Factor Model: (X^2= 41, df= 11 CFI=0.995, RMSEA= 0.031) a = 0.82

Note: Quality and quantity of teaching were measured with only 2 indicators and therefore they are not presented in Table 3.8.

ANASTASIA PANAYIOTOU

Table 3.9

Items of the CFA models and across-country results of the SEM and reliability analysis for the student questionnaire measuring teacher factors.

Teacher Factors	Quantitative Characteristics	Qualitative Characteristics	SEM and Reliability Analysis Results
1) *Modeling*	44, 47	45, 46	One – Factor Model: (X^2= 47, df= 2 CFI=0.991, RMSEA= 0.049) a = 0.67
2) *Structuring*	3, 2, 34, 38	1, 4, 7, 10	Two – Factor Model: (X^2= 298, df= 18 CFI=0.967, RMSEA= 0.040) a1 =0.62, a2= 0.52
3) *Application*	11, 12	13, 26	One – Factor Model: (X^2= 2, df= 1 CFI=0.99, RMSEA= 0.012) a = 0.58
4) *Questioning*	25, 39	24, 37, 40, 42	One – Factor Model: (X^2= 166, df= 8 CFI=0.970, RMSEA= 0.045) a = 0.46
5) *CLE / Teacher – Student Interaction*	16, 17	19, 20, 21	One – Factor Model: (X^2= 21, df= 5 CFI=0.997, RMSEA= 0.018) a = 0.63
6) *CLE / Dealing with Misbehaviour*	29	27, 33 , 30	One – Factor Model: (X^2= 230, df= 2 CFI=0.965, RMSEA= 0.10) a = 0.69

Multilevel regression techniques (Muthen & Satorra, 1995) allow researchers to determine at once the variation at each level when dealing with complex data, by using sets of predictors/independent variables at each level (Snijders, 2011; Thomas & Heck, 2001). On the contrary, using a uni-level approach may lead to false interpretation of results due

ANASTASIA PANAYIOTOU

to a higher probability of falsely rejecting the null hypothesis (i.e., committing Type I errors) (Hill & Rowe, 1996). Namely, by disaggregating data the sample size is increased and therefore one may incorrectly identify statistically significant results. On the contrary, by aggregating data at a higher level may inevitably decrease the sample size leading to loss of statistical power. This misleading data treatment is also recognized and described by Hox (2010) who mentions that analyzing multilevel data using ordinary multiple regression analysis may cause both statistical as well as conceptual problems (i.e., aggregating or disaggregating data to one level and then falsely drawing conclusions for another level). Educational settings are usually described using four distinct levels; country, school, classroom, and the lower level - the student, with the lower levels being nested in the upper levels. Thus, multilevel modeling techniques may assist the identification of the contribution of each level to the variance of the outcome variable; in this case student achievement (Luyten & Sammons, 2010).

Taking the hierarchical structure of the data into consideration, separate multilevel modeling techniques (Goldstein, 2003; Snijders & Bosker, 1999) were used in this study, to identify the impact of the teacher and school factors on student achievement in mathematics and science. In each case a two-level regression or "random coefficient" model (Hofmann, 1997; Longford, 1993) was used. According to Plewis and Hurry (1998, p. 14) "The level at which the intervention is assigned should influence the way in which the data from the study are analyzed". Therefore, to test the effect of the teacher factors on students' achievement in mathematics and science a two-level model with the classroom/teacher as the higher and the student as the lower level was used. While to test the effect of the school factors on student achievement gains in mathematics and science the two levels used were the school and student.

Due to the fact that most of the sample's schools in the participating countries were comprised of only one classroom per year group, it was not possible to distinguish the two

ANASTASIA PANAYIOTOU

levels (i.e., classroom and school) and therefore a three-level model (i.e., school, classroom and student) was not possible to be used. Hence, when examining the impact of school factors, the school level was used for the analyses instead of the classroom/teacher level. Likewise, due to the fact that in each classroom only one teacher was assigned it was not possible to distinguish two different levels (i.e., classroom and teacher). Moreover, the country level was not used as an additional level in the models as the number of countries participating in this study was very small (see Maas & Hox, 2005).

The null model used in the two-level regression analysis is expressed by the following equation:

$$Y_{ij} = \beta_{00} + u_{0j} + r_{ij}$$

Where:

> *Y*: is the dependent variable; in this case student achievement in mathematics or science respectively
> *j*: is level-2 units (i.e., number of classes or schools, respectively)
> *i*: is level-1 units (i.e., number of students)
> $\beta_{00:}$ is the intercept
> u_{0j}: is the random part
> r_{ij}: is the residual

In the next equation, model 1 is presented where an explanatory variable measured at the student level (e.g., student initial achievement) is introduced:

$$Y_{ij} = \beta_{00} + \beta_1 X_{ij} + u_{0j} + r_{ij}$$

In this case β_1 represents the regression coefficient (regression slope) for the explanatory variable *X*.

Since the aim of this study was not only to reveal generic factors operating at classroom and school level, but also to identify possible differential effects of the factors in

ANASTASIA PANAYIOTOU

the six participating countries, both across- and within-country multilevel analyses were conducted for the teacher and school factors. The results of the across- and within-country multilevel analyses conducted to test the impact of the teacher and school factors on student achievement in mathematics and science are presented in Chapter 4.

Multivariate Multilevel Analysis

Multivariate multilevel analysis is seen to comprise an extension of simple multilevel analysis in terms of providing the opportunity to insert more than one dependent variable to the model (Luyten & Sammons, 2010). Specifically, “multivariate data analysis refers to all statistical methods that simultaneously analyze multiple measurements on each individual respondent or object under investigation” (Hair, 2011, p. 905). Thus, as described by Hox (2002), multivariate multilevel analysis can be compared to the uni-level MANOVA where more than one continuous outcome measures can be used (instead of ANOVA).

Multivariate multilevel regression models portray some significant advantages in comparison to conducting multiple multilevel regression models. First, similarly to comparing MANOVA and ANOVA, when simultaneously adding all dependent variables in one multivariate model the possibility of committing Type I error is reduced (Tabachnick & Fidell, 2011). Second, when adding more than one outcome variable in one model, it is possible to identify correlations between the dependent variables which often exist and are not taken into consideration when conducting separate multilevel analysis. For example, multivariate analysis allows us to examine whether teachers that are effective in mathematics are also effective in science and/or whether students’ achievement gains in the two subjects are related. Finally, multivariate analysis provides the advantage of assisting the identification of differential effects of the independent variables on the dependent variables (e.g., establishing whether a classroom level factor has a statistically

ANASTASIA PANAYIOTOU

significant effect both in mathematics and science and whether the effect is bigger in one subject rather than the other) (Luyten & Sammons, 2010).

Taking in mind the contribution of multivariate analysis and aiming to provide further support to the results of the multilevel analyses described in the previous section in regard to the identification of the generic and differential effects of the classroom and school factors of the dynamic model in two different outcomes, across- country multivariate multilevel analysis was also conducted.

The random intercept, three-level model for dependent variable Y_h without any explanatory variables (null model) used in the multivariate analysis is expressed by the following equation:

$$Y_{hij} = \beta_{0h} + U_{hj} + R_{hij}$$

To represent the multivariate data in the multilevel approach three nesting levels are used.

The first level is that of the dependent variables indexed by *h*=1 and 2, where 1= mathematics and 2= science.

i: is the second level and it is that of the individuals (i.e., number of students)

j: is the third level and it is that of groups (i.e., number of classes or schools, respectively).

β_{0h}: is the intercept for each outcome

u_{hj}: is the random part

r_{hij}: is the residual

The results of the multivariate multilevel analysis conducted to compare the impact of the teacher and school factors on student achievement in mathematics and science and identify possible differential effects in the two subjects, are presented in Chapter 4.

ANASTASIA PANAYIOTOU

Research Limitations

Every study that takes place involves a particular research design, the selection processes of a sample that is considered suitable so as to assist the provision of answers to specific research questions and the researchers' attempts to establish the validity and reliability of the data collected. However, the choices made by researchers during the research design of the study, as well as the data analyses processes may lead to research limitations which should be reported and elaborated on to assist in depth comprehension of the new knowledge and comprise a base for discussions that may lead to the further development and understanding of a scientific field (Price & Murnan, 2004). As described in the previous sections of this Chapter, this study was designed so as to collect data from teachers and students in six European countries in regard to the school and classroom level factors of the dynamic model, respectively, in order to contribute – among others – to the establishment of a solid theoretical framework which can provide a basis for further studies in the field of EER. However, some possible limitations can be acknowledged.

First, it should be acknowledged that this cross-national study was not able to provide data about each measurement dimension of the teacher and school factors. This could be attributed to the difficulties which occurred in the attempt of developing a student and teacher questionnaire addressing the functioning of these factors in quite diverse educational contexts. By removing items that were not in line with the context of all the participating countries, the final specification table of the questionnaires did not cover the five dimensions of each factor.

Additionally, to collect data in regard to the teacher factors of the dynamic model only one measurement method was used (i.e., student questionnaires). Even though it was possible to establish the construct validity of the student questionnaire and identify that primary students in grade 4 are able to provide valid data on their teachers' behavior in the classroom in relation to the teacher factors included in the dynamic model when they are

ANASTASIA PANAYIOTOU

asked to report on observable behavior and to teaching actions instead of inferences or opinions, different mechanisms of measuring teacher behavior are required to support the validity and reliability of data. Thus a methodological limitation that may be acknowledged is that, mostly due to limited resources, the study was not designed to collect data through the use of observation instruments, using well-trained observers that may generate valid and reliable data about all five measurement dimensions (Reynolds, 2006; Teddlie et al., 2006). Similarly, it was not possible to collect data on the school factors through other means than teacher questionnaires (e.g., documents) and the limitations of using perceptual methods to measure school factors should be acknowledged.

Another limitation of the study relates to the test used to measure student achievement in mathematics and science. Specifically, to assess student knowledge in these two subjects only the released items of TIMSS 2007 were used. One could however suggest that these items are not in a position to cover the whole spectrum of material taught in grades 3 and 4. Having, access to more items or constructing tests based on the curriculum of each participating country, may have provided the opportunity to explain more of the variance in these two outcomes and thus to identify larger effects of the teacher and school factors on achievement. In addition, the focus of this comparative study was only aimed towards students' cognitive outcomes in two subjects. Such international studies may also be conducted to provide evidence about the impact of teacher and school factors in different learning outcomes other than mathematics and science, in other domains (e.g., affective and psychomotor), and in meta-cognition.

Finally, the study was conducted in six different European countries and one could suggest that the findings could be attributed to basic similarities of European cultures. In order to explore this argument, further comparative studies collecting data from countries within and outside Europe are required, to test the generic nature of these factors. These

ANASTASIA PANAYIOTOU

studies may provide evidence about the cultural impact on education and quality of teaching and thereby the factors and dimensions contained in the dynamic model.

This Chapter has outlined the research design and methods used in this study. The next section presents how research data, were analyzed in order to address the research questions set.

ANASTASIA PANAYIOTOU

CHAPTER 4

RESEARCH RESULTS

This chapter presents the analysis of the data collected throughout the study. Research results are presented addressing the research questions set in Chapter 1. Specifically, this Chapter is divided into four sections. The first section presents the analyses conducted to test the construct validity of the student and teacher questionnaire as well as the assumption of the dynamic model that the teacher and school factors, respectively, are related to each other, both across and within countries. The second section describes the results of multilevel modeling that was performed to explore the impact of the teacher and school factors on student achievement gains in mathematics and science. The third section presents the results of multivariate analysis conducted to identify possible differential effects of the teacher and school factors on student achievement in mathematics and science as well as correlations between the two dependent variables (i.e., achievement in mathematics and science). Finally, the fourth section presents the results of multilevel analyses searching for curvilinear relations of the teacher and school factors with achievement in the two outcomes.

Using Structural Equation Modeling to test the Construct Validity of the Student and Teacher Questionnaire

This section presents the results of the SEM analyses conducted to provide support to the construct validity of the questionnaires used to measure quality of teaching (i.e., student questionnaire) and the school policy in regard to the quality of teaching, the SLE and the policy on the evaluation of these two aspects (i.e., teacher questionnaire). Both the across- and within- country results are presented and discussed.

ANASTASIA PANAYIOTOU

The Construct Validity of the Student Questionnaire

Across-country results

In the previous Chapter, the results of the Confirmatory Factor Analysis that was separately conducted for each of the teacher factors of the dynamic model were presented (see Table 3.9). To examine the factorial structure of the student questionnaire, the EQS software for Structural Equation Modeling (Bryne, 1994) was used. Three separate fit indices were examined to evaluate the extent to which the data fitted the tested models: a) the scaled chi-square, b) Bentler's (1990) Comparative Fit Index (CFI), and c) the Root Mean Square Error of Approximation (RMSEA) (Brown & Mels, 1990). Additionally, the factor parameter estimates for the models with acceptable fit were observed to facilitate the interpretation of the models.

Initially across-country analysis was conducted to examine whether the main assumptions of the dynamic model in regard to the grouping of factors could be confirmed. Then six separate within-country analyses were conducted so as to examine the fit of the across-country model to the data obtained by each of the six participating countries. Even though the dynamic model assumes that all eight factors located at the classroom level are part of an overarching factor relating to quality of teaching, when taking into consideration the main notions of the classroom level factors, another assumption can be made. Namely, that the eight factors can be grouped into two overarching factors according to whether they are referring to the "quality" or to the "quantity" of teaching. This assumption was tested and supported by the across-country data collected in this study; thus two second order factors were identified.

The first second order factor consists of the factors measuring management of time, teacher ability to deal with student misbehaviour and the quantitative aspects of questioning (i.e., teacher's ability to raise appropriate questions so that teaching time is not

ANASTASIA PANAYIOTOU

lost). This overarching factor can be treated as the factor measuring the ability of teachers to maximize the use of teaching time and can be seen as an indicator of the quantity of teaching. All the other factors were found to load on the second overarching factor which can be treated as an indicator of the quality of teaching. V1 is an indicator of the factor concerned with orientation. Due to the adaptation of the student questionnaire to fit the context of each country no other items measuring orientation were used and therefore this item is regressed directly to the second order factor of quality of teaching. It is also important to note that the correlation between these two overarching factors was found to be very small (see Figure 4.1), implying that teachers who are able to maximize the use of teaching time are not also necessarily able to use the teaching time appropriately. The main results of the SEM analysis for testing the factorial structure of the student questionnaire are presented in the first part of this section and the fit indices of the across-country model are shown in Table 4.1.

When testing the construct validity of a measurement instrument one should also examine the so called "social desirability factor" (Robinson, Rush & Head, 1974). Social desirability factor refers to people's tendency in responding to a measurement instrument in a way that depicts commonly acceptable views, thus fitting into the norm (Edwards, 1957). Social desirability may be portrayed in research data through high positive correlations among questionnaire items demonstrating response bias (Christie & Lindauer, 1963). Additionally, a researcher should also take into consideration that solely testing for the goodness-of-fit of a model is not adequate. On the contrary, when comparing alternative models to test their fit to a given data set one should also take the parsimony of the models into perspective (Kline, 1998; Pitt & Myung, 2002). Parsimony principle can be reflected by taking into account the simplicity of a model as well as its ability to accurately account for the structure of the data. Parsimony principle is also commonly described as the "Occam's razor". As explained by Vandekerckhove, Matzke and Wagenmakers (2015,

ANASTASIA PANAYIOTOU

p. 4) "Occam's metaphorical razor symbolizes the principle of parsimony: by cutting away needless complexity, the razor leaves only theories, models, and hypotheses that are as simple as possible without being false".

Thus, two alternative models were examined to compare their fit to the data with the proposed model. In the first of these alternative models (Model 2), all the items that were used for the SEM analysis were considered as belonging to a single first-order factor. This model was an attempt to see if the questionnaire items refer to a social desirability factor and thereby the questionnaire may not provide valid data. In the second alternative model (Model 3), the 19 items concerned with the factors of the dynamic model measuring quality of teaching were considered as belonging to a single first-order factor, while the items concerned with the three factors of the dynamic model measuring quantity of teaching were assumed to belong to another first-order factor. If Model 3 was found to fit to the data, this might cause doubts about the complexity of the proposed model. However, the results show that Model 1 provides the best fit to the data and only the fit indices of Model 1 can be considered satisfactory (see Table 4.1). Figure 4.1 presents the results from the across-country SEM analysis and shows the second-order factor model that best fits the data of the student questionnaire (Model 1).

Table 4.1

Fit indices of the models used to test the factorial structure of the instrument measuring the teacher factors emerged from the across-country analyses

Models	**X^2**	**df**	**χ^2/df**	**P**	**CFI**	**RMSEA**
Across -country sample (N=9967)						
Model 1	3604	325	11.1	0.001	0.929	0.032
Model 2	16507	350	47.1	0.001	0.648	0.068
Model 3	6502	349	18.3	0.001	0.866	0.042

ANASTASIA PANAYIOTOU

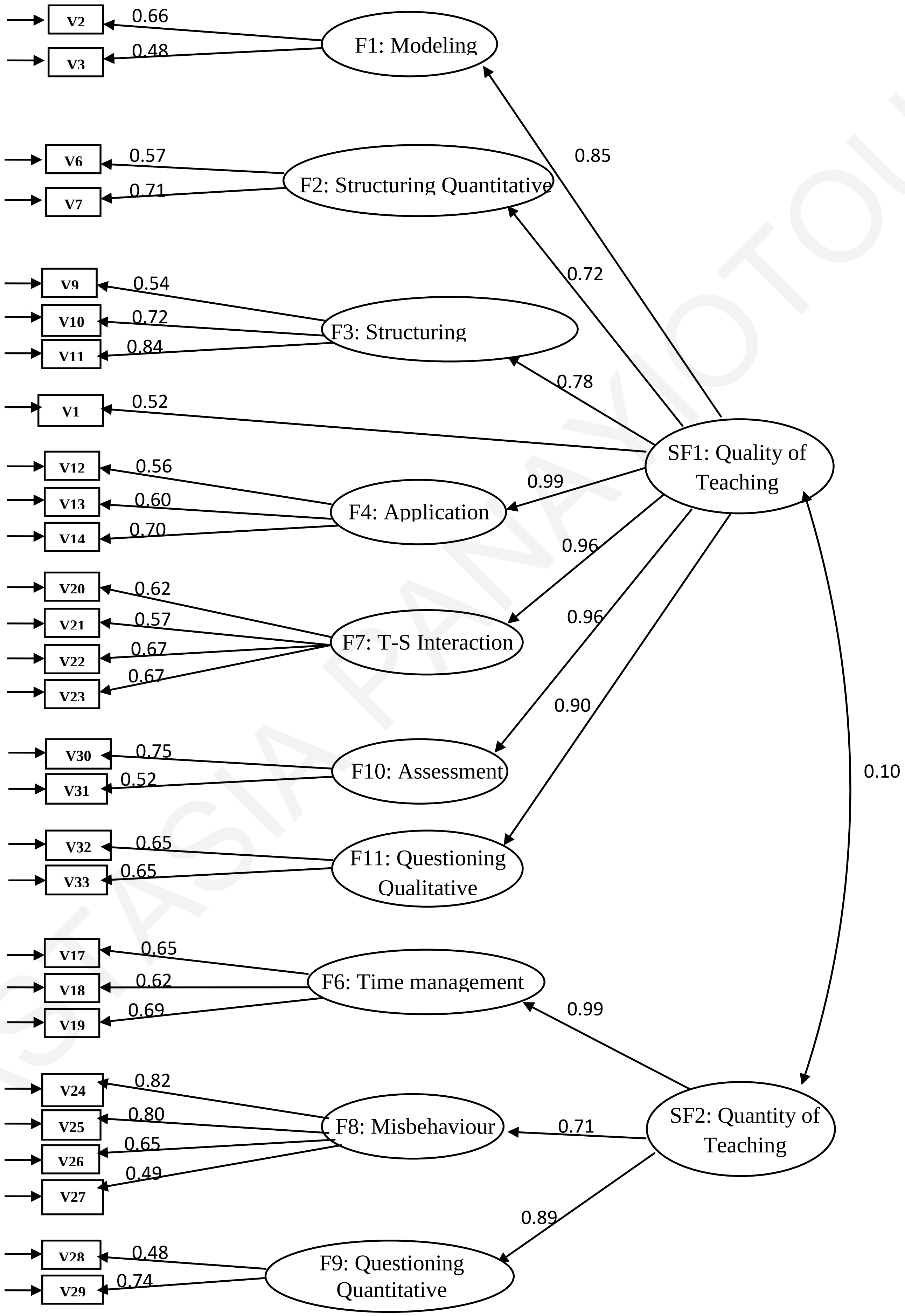


Figure 4.1 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates

ANASTASIA PANAYIOTOU

Within-country results

A similar procedure as the across-country analyses was used in order to conduct within-country SEM analyses where the data of each country were processed separately. For the within-country analyses the assumption was again in line with the main assumptions of the dynamic model in regard to the allocation of factors into the two broader categories of quantity and quality of teaching. In the within-country analyses the two alternative models (i.e., Model 2 and Model 3) that were tested in the across-country analysis were also tested to compare their fit to the data with the proposed model.

The results, provided in Table 4.2, show that the second-order factor model (i.e., the theoretical model) best fits the data from each country separately, while neither of the two alternative models meet the requirements for acceptable fit. Moreover, the within-country analyses revealed that most of the correlations between the two second-order factors are small, which again implies that teachers who are effective at maximizing the use of teaching time may not be as effective in using the teaching time appropriately (see Figures 1-6 in Appendix F).

In certain countries some further items were removed during the SEM analyses due to low loadings on the corresponding factor. For example, item 10 from the factor of structuring (qualitative characteristics) was excluded from the analysis of the Slovenian data as its loading was low. Also, in the analysis of the German data three more items had to be removed in order for the model to best fit the data of the country (see Figure 6 in Appendix F). However, the structure of the model that was found to best fit the across-country data was supported through the within-country analyses, thus providing further support to the construct validity of the student questionnaire.

ANASTASIA PANAYIOTOU

Table 4.2

Fit indices of the models used to test the factorial structure of the instrument measuring the teacher factors emerged from the within-country analyses

Models	X^2	df	χ^2/df	P	CFI	RMSEA
A) Belgium (N=1908)						
Model 1	731	297	2.4	0.01	0.929	0.028
Model 2	2668	324	8.3	0.001	0.616	0.061
Model 3	1395	323	4.3	0.001	0.824	0.042
B) Cyprus (N=1881)						
Model 1	825	317	2.6	0.01	0.943	0.029
Model 2	3441	350	9.8	0.001	0.652	0.069
Model 3	1584	349	4.3	0.001	0.861	0.043
C) Greece(N=905)						
Model 1	560	312	1.8	0.01	0.944	0.030
Model 2	2386	350	6.8	0.001	0.542	0.080
Model 3	1285	349	3.7	0.001	0.789	0.054
D) Ireland (N=2140)						
Model 1	915	327	2.8	0.01	0.929	0.029
Model 2	2416	350	6.9	0.001	0.752	0.053
Model 3	1416	349	4.1	0.001	0.872	0.038
E) Slovenia (N=2049)						
Model 1	1158	281	4.1	0.01	0.926	0.039
Model 2	4573	324	14.1	0.001	0.640	0.080
Model 3	2196	323	6.8	0.001	0.841	0.053
F) Germany (N=1072)						
Model 1	547	219	2.5	0.01	0.959	0.037
Model 2	3472	275	12.6	0.001	0.599	0.104
Model 3	1434	274	5.2	0.001	0.855	0.063

ANASTASIA PANAYIOTOU

The Construct Validity of the Teacher Questionnaire

Across-country results

To examine the validity of the teacher questionnaire both within- and across-country SEM analyses were conducted and evidence supporting the construct validity of the questionnaire was produced. As mentioned in Chapter 3, prior to conducting the SEM analyses, schools with a teacher response rate below 30% (or schools with less than 3 teacher questionnaires) had to be removed from the data set. For that reason, the data collected from the teacher questionnaires in Greece were not used for any analyses concerning the school factors, since the schools in Greece were very small and for all the Greek schools less than 3 questionnaires were obtained. Therefore, in the next section which concerns the within-country analyses, no results are presented regarding the teacher questionnaires administered in Greece.

Separate SEM analyses were then conducted for the three overarching factors: a) School policy on teaching, b) Policy on the School Learning Environment and c) Policy on Evaluation. The first overarching factor which is school policy on teaching consists of the factors measuring: a) quantity of teaching, b) quality of teaching and c) provision of learning opportunities. The second overarching factor is policy on the school learning environment and consists of the factors measuring: a) teacher collaboration, b) partnership policy, c) provision of sufficient learning resources and d) relation with the community. Finally, the third overarching factor, namely policy on evaluation, consists of the factors measuring: a) evaluation of the school policy on teaching and b) evaluation of the SLE. The fit indices of the across-country models are shown in Table 4.3.

Apart from these three models, another model was tested for each of the three overarching factors (i.e., school policy on teaching, policy on the SLE and school policy on evaluation) in order to compare its fit to the data with the three proposed models. In these

ANASTASIA PANAYIOTOU

models (Models 2) all the items that were used for the SEM analysis in each of the three overarching factors were considered as belonging to a single factor. These models were examined in order to take into consideration the possibility of the teacher responses to be a result of a social desirability factor. In this case, high positive correlations among questionnaire items would have been observed causing doubts on the validity of the data collected. The fit indices of each alternative model are also shown in Table 4.3. It can be observed that Model 1 is the model that best fits the data for each of the overarching factors. In case that Models 2 were found to best fit the data, based on the parsimony principle it would not have been possible to have scores per each factor separately.

Figures 4.2, 4.3 and 4.4 reveal the second order factor models that were found to fit the teacher questionnaire data when across-country analysis was conducted. These models demonstrate that the items of the teacher questionnaire can be used to measure the school factors. Figure 4.2 presents the second-order factor model of the teacher questionnaire measuring school factors on the school policy on teaching with factor parameter estimates. Figure 4.3 presents the second-order factor model in regard to the policy on the school learning environment and Figure 4.4 shows the second-order factor model for school policy on the evaluation of quality of teaching and the SLE. By observing the loadings of the items on the first order factors, as well as the loadings of the first order factors on the second order, one may notice that they are all very high and that all the loadings are statistically significant. The only exception is for V1 in the overarching factor concerned with policy on teaching (Figure 4.2) and for V6, V7 and V13 in the overarching factor concerned with evaluation (Figure 4.4) which are around .50. However, the loadings of all the other items are very high (higher than 0.8 for policy on teaching and higher than 0.6 for evaluation).

ANASTASIA PANAYIOTOU

Table 4.3

Fit indices of the models emerged from the across-country SEM analyses of the teacher questionnaire used to measure each overarching school factor

A. School Policy on teaching							
Models	X^2	Df	X^2/ df	P	CFI	RMSEA	Range RMSEA
Model 1	143	16	8.9	0.001	0.984	0.052	0.045 – 0.060
Model 2 (one factor model)	484	20	24.2	0.001	0.943	0.090	0.083 – 0.097
B. Policy on the school learning environment							
Models	X^2	Df	X^2/ df	P	CFI	RMSEA	Range RMSEA
Model 1	669	96	6.9	0.001	0.963	0.045	0.042 – 0.049
Model 2 (one factor model)	3879	135	28.7	0.001	0.759	0.098	0.095 – 0.101
C. School Evaluation							
Models	X^2	Df	X^2/ df	P	CFI	RMSEA	Range RMSEA
Model 1	585	57	10.2	0.001	0.967	0.057	0.053 – 0.061
Model 2 (one factor model)	1600	65	24.6	0.001	0.905	0.090	0.087 – 0.094

ANASTASIA PANAYIOTOU

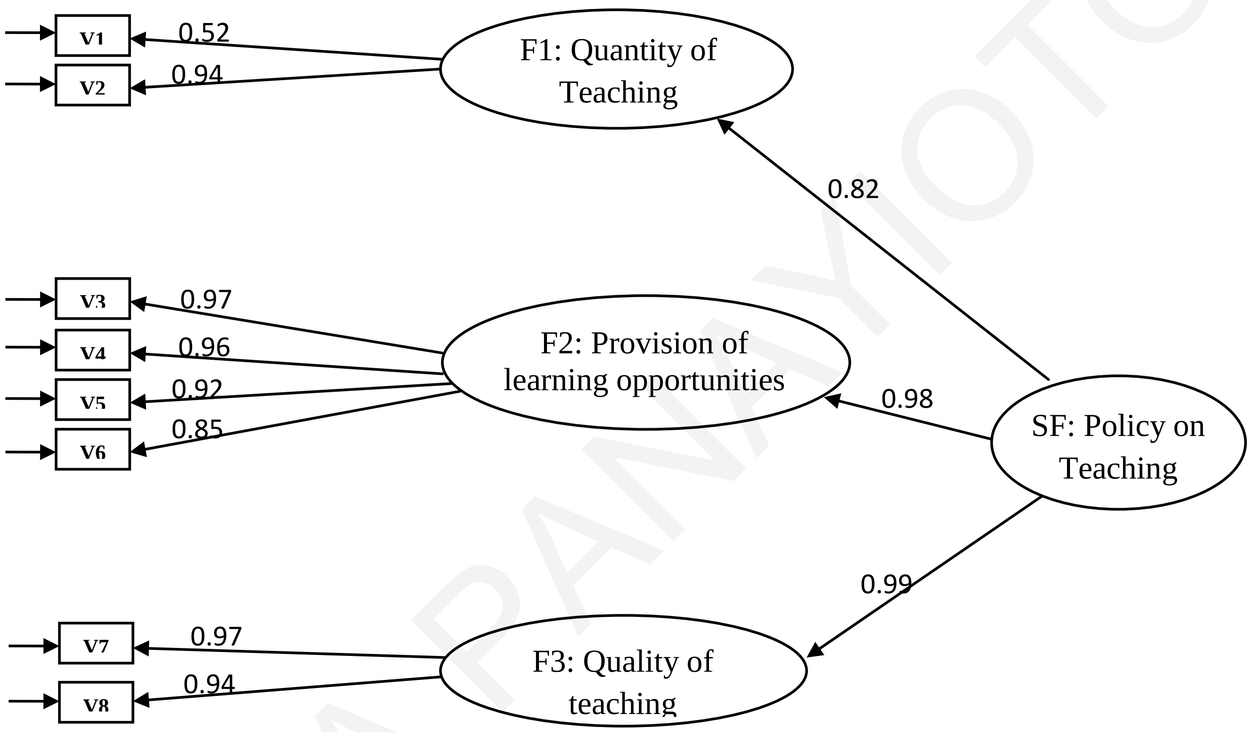
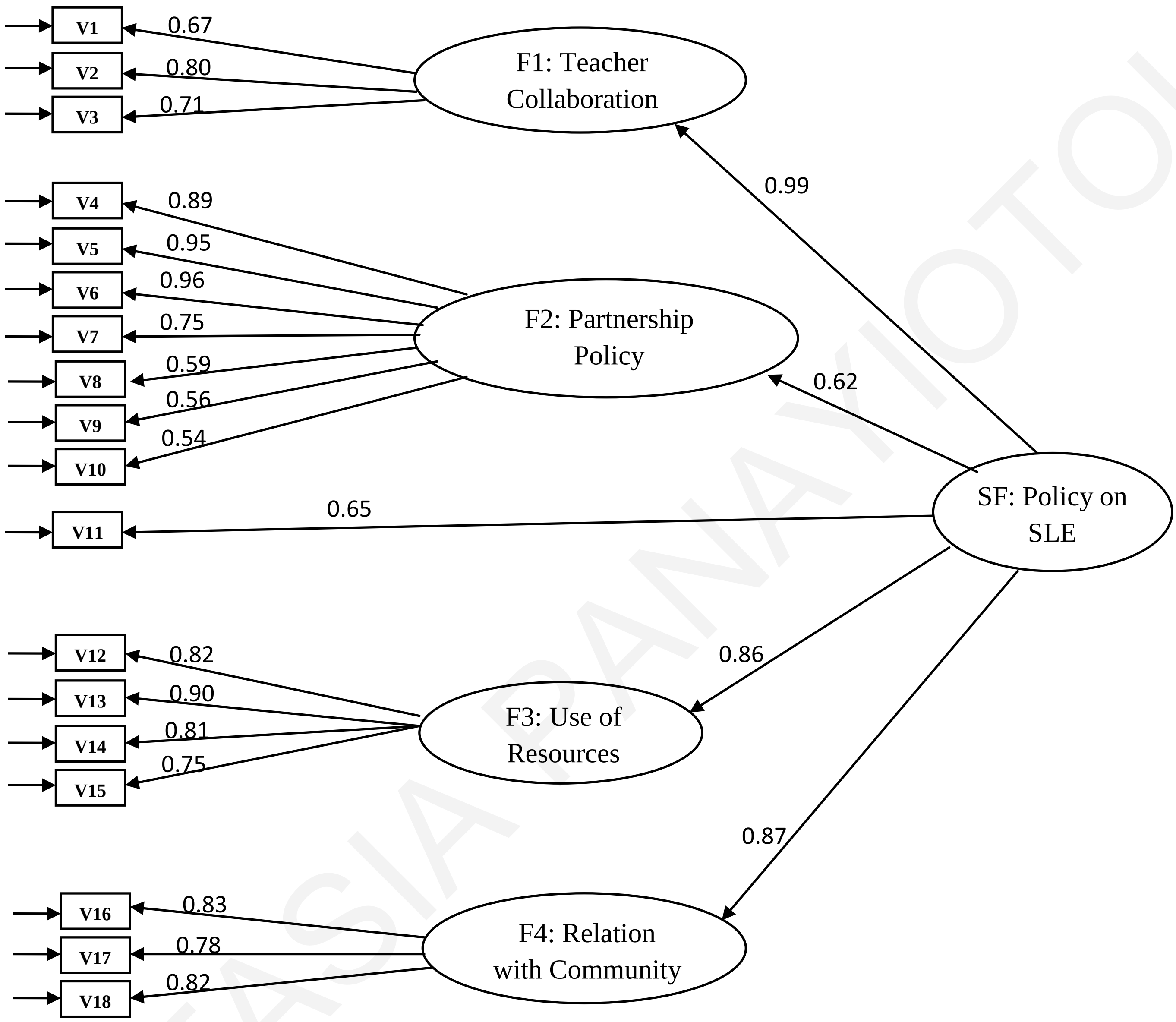


Figure 4.2 The second-order factor model of the teacher questionnaire measuring school factors on the school policy on teaching with factor parameter estimates

ANASTASIA PANAYIOTOU



**Note:* V11 is an indicator of student behavior outside the classroom

Figure 4.3 The second-order factor model of the teacher questionnaire measuring school factors on the school learning environment with factor parameter estimates

ANASTASIA PANAYIOTOU

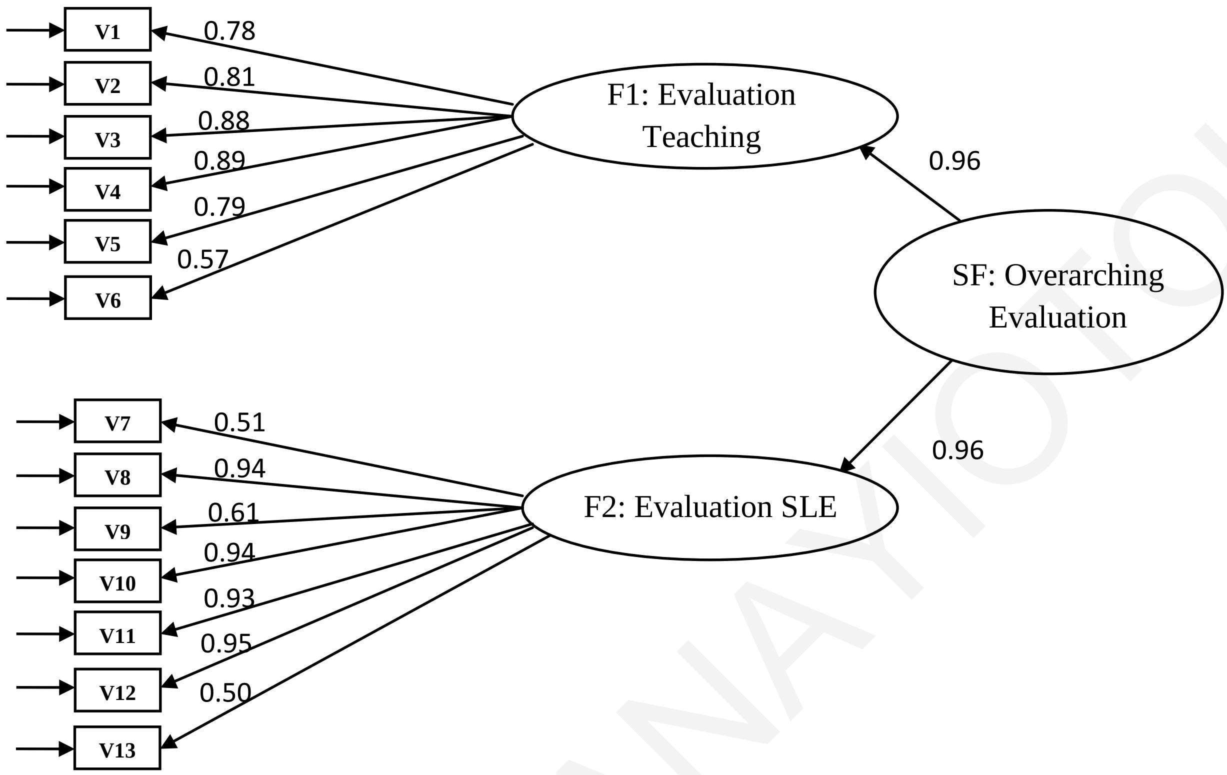


Figure 4.4 The second-order factor model of the teacher questionnaire measuring school factors on evaluation with factor parameter estimates

Based on the loadings of the items that occurred from the SEM analyses as they are shown in Figures 4.2-4.4, factor scores were estimated. These factor scores were aggregated at the school level and were used for the multilevel analyses presented in the following sections of this Chapter in order to identify the impact of the school factors on student achievement gains in mathematics and science.

The next section presents the within-country analyses performed to test whether the three models that were estimated using the across country data fit well to the data emerged by each country separately.

ANASTASIA PANAYIOTOU

Within-country results

For the within-country analyses separate SEM analyses were again conducted for the three overarching factors: a) School policy on teaching, b) Policy on the School Learning Environment and c) Policy on Evaluation. The aim was to develop three models for these overarching factors based on the data from each country separately in order to replicate the across-country models. From the within-country SEM analysis, it was found that the models that were produced by the across-country analysis fit well to the data emerged by each country separately. More specifically, three models were developed for each country and three second order factors were identified. The three overarching factors for each country consist of the same factors that were developed in the across country analyses. However, in analyzing the Irish data, for the overarching factor concerned with the policy on the SLE, items 17 and 29 from the factor concerning the relation with the community were excluded from the analysis as their loadings were low. Therefore question 30 (i.e., V16) is regressed to the overarching factor. For the overarching factor concerned with the policy on evaluation the structure of the model for Ireland is the same as the across country model for evaluation and no further items had to be removed. Figures 1.1 - 5.3 which are included in Appendix G present the second-order factor models of the teacher questionnaire measuring school factors with factor parameter estimates for each country separately.

In the within-country analyses the other model which was tested in the across-country analyses (Model 2) was again examined to compare its fit to the data with the proposed model for each of the three overarching factors (i.e., school policy on teaching, policy on the school learning environment and school evaluation). In these models (Models 2) all the items that were used for the SEM analysis in each of the three overarching factors were considered as belonging to a single factor. The fit indices of each model are shown in

ANASTASIA PANAYIOTOU

Tables 4.4 – 4.8. It can be seen that Model 1 is again the model that was found to best fit the data in each country for each of the overarching factors.

Table 4.4

Fit indices of the models emerged from the SEM analyses of the teacher questionnaire used to measure each overarching school factor in Belgium

A. School Policy on teaching							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	21	15	1,4	0.001	0.994	0.029	0.001 – 0.054
Model 2 (one factor model)	82	20	4,1	0.001	0.941	0.077	0.060 – 0.095
B. Policy on the school learning environment							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	140	83	1,6	0.001	0.974	0.036	0.025 – 0.046
Model 2 (one factor model)	1448	135	10,7	0.001	0.546	0.120	0.113 – 0.126
C. School Evaluation							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	90	53	1,7	0.001	0.979	0.037	0.023 – 0.049
Model 2 (one factor model)	347	65	5,3	0.001	0.846	0.091	0.082 – 0.101

ANASTASIA PANAYIOTOU

Table 4.5

Fit indices of the models emerged from the SEM analyses of the teacher questionnaire used to measure each overarching school factor in Cyprus

Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
A. School Policy on teaching							
Model 1	27	16	1,7	0.001	0.989	0.049	0.011 – 0.080
Model 2 (one factor model)	61	20	3,5	0.001	0.957	0.085	0.061 – 0.109
B. Policy on the school learning environment							
Model 1	83	88	0,94	0.001	0.999	0.001	0.001 – 0.028
Model 2 (one factor model)	431	135	3,2	0.001	0.81	0.087	0.078 – 0.096
C. School Evaluation							
Model 1	74	54	1,4	0.001	0.99	0.036	0.011 – 0.055
Model 2 (one factor model)	187	65	2,9	0.001	0.937	0.081	0.067 – 0.094

ANASTASIA PANAYIOTOU

Table 4.6

Fit indices of the models emerged from the SEM analyses of the teacher questionnaire used to measure each overarching school factor in Germany

A. School Policy on teaching							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	24	15	1,6	0.001	0.992	0.041	0.001 – 0.070
Model 2 (one factor model)	110	20	5,5	0.001	0.919	0.112	0.092 – 0.133
B. Policy on the school learning environment							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	105	74	1,4	0.001	0.981	0.035	0.018 – 0.049
Model 2 (one factor model)	695	119	5,8	0.001	0.659	0.117	0.108 – 0.125
C. School Evaluation							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	115	55	2,1	0.001	0.969	0.056	0.041 – 0.070
Model 2 (one factor model)	285	65	4,4	0.001	0.886	0.097	0.086 – 0.109

ANASTASIA PANAYIOTOU

Table 4.7

Fit indices of the models emerged from the SEM analyses of the teacher questionnaire used to measure each overarching school factor in Ireland

Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
A. School Policy on teaching							
Model 1	8	5	1,6	0.001	0.995	0.044	0.001 – 0.096
Model 2 (one factor model)	18	9	2	0.001	0.985	0.057	0.018 – 0.094
B. Policy on the school learning environment							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	58	66	0,9	0.001	0.999	0.001	0.001 – 0.026
Model 2 (one factor model)	791	104	7,6	0.001	0.631	0.142	0.133 – 0.151
C. School Evaluation							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	84	49	1,7	0.001	0.98	0.047	0.030 – 0.064
Model 2 (one factor model)	322	65	4,9	0.001	0.859	0.110	0.098 – 0.122

ANASTASIA PANAYIOTOU

Table 4.8

Fit indices of the models emerged from the SEM analyses of the teacher questionnaire used to measure each overarching school factor in Slovenia

A. School Policy on teaching							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	53	12	4,4	0.001	0.987	0.050	0.037 – 0.064
Model 2 (one factor model)	277	20	13,9	0.001	0.921	0.096	0.086 – 0.107
B. Policy on the school learning environment							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	388	115	3,4	0.001	0.963	0.041	0.037 – 0.046
Model 2 (one factor model)	1790	135	13,3	0.001	0.774	0.094	0.090 – 0.098
C. School Evaluation							
Models	X^2	Df	X^2/ df	p	CFI	RMSEA	Range RMSEA
Model 1	275	57	4,8	0.001	0.972	0.053	0.046 – 0.059
Model 2 (one factor model)	710	65	10,9	0.001	0.917	0.085	0.079 – 0.090

The next section presents the results of the multilevel analyses that were conducted to identify the impact of the teacher and school factors of the dynamic model on student achievement gains in mathematics and science. To perform the multilevel analyses, factor

ANASTASIA PANAYIOTOU

scores were estimated for each teacher and school level factor. These factor scores were based on the loadings of the items that occurred from the SEM analyses. Two different sets of factor scores were developed; across-country factor scores (using the loadings of the items as they occurred from the across-country SEM analyses) and within-country factor scores (using the loadings of the items as they occurred from SEM analyses performed with the data of each country separately).

Using Multilevel Analysis to Search for the Impact of the Teacher Factors on Student Achievement

Across-country results

To identify the impact of the teacher factors on student achievement in mathematics and science, separate multilevel modeling analyses were conducted for each of the two subjects. The results of these analyses are presented in Tables 4.9 and 4.10. The first step was to run a two-level model (classroom/teacher level and student level) with no explanatory variables (i.e., empty model) to determine the variance at each level. The variance for mathematics was 24.4% at the classroom level and 75.6% at the student level and was statistically significant in each level. The variance for science was 31.4% at the classroom level and 68.6% at the student level and was again statistically significant in each level. Thus, comparing the empty models in mathematics and science one may realize that the effect of the classroom level is more pronounced on achievement in science rather than in mathematics.

Then, in Model 1 the context variables of the students' prior achievement and the students' prior achievement at the classroom level were added to the empty model. Both variables had a statistically significant effect at level .05. In addition, as can be seen in Tables 4.9 and 4.10, Model 1 explains more than 40% of the total variance of student

ANASTASIA PANAYIOTOU

achievement in each outcome (mathematics and science), and most of the explained variance is at the student level which is consistent with results previously found in other studies (see Muijs et al., 2014; Reynolds et al., 2014).

At the next step of the analysis, for each student outcome (i.e., mathematics and science), different versions of Model 2 were established (i.e., Models 2a-2i for mathematics, as shown in Table 4.9, and Models 2a-2g for science, as shown in Table 4.10). In each version of Model 2, the first-order factor scores of the SEM models which refer to the teacher factors of the dynamic model were added one by one to Model 1. Thus, the fitting of each of these models was tested against Model 1, and the likelihood statistic (X^2) shows a significant change between Model 1 and each version of Model 2 ($p <.001$). This implies that variables measuring the teacher factors have significant effects on student achievement in mathematics and science. Table 4.9, shows that in mathematics all the first order factors were found to have a statistically significant effect on achievement, except of the factor concerned with the quantitative characteristics of questioning. Table 4.10 which depicts the results of the multilevel analysis for science shows that all the factors were found to have a statistically significant effect on student achievement except three factors concerned with: a) modeling, b) the quantitative characteristics of questioning, and c) qualitative characteristics of structuring. The factors which were not found to have a statistically significant effect on achievement are not included in Tables 4.9 and 4.10, respectively.

Due to the fact that the first order teacher factors are highly correlated with each other, it was not possible to create a model with all the first order teacher factors of the quality and/or the quantity factors since in that case the problem of multicollinearity would have not allowed for the true effect of the factors to be identified, possibly leading to an incorrect interpretation of results (Farrar & Glauber, 1967; Graham, 2003; Gunst & Webster, 1975). Therefore, in Model 3, the two second-order, or overarching, factors (i.e.,

ANASTASIA PANAYIOTOU

Quality of Teaching and Quantity of Teaching) were added to Model 1 to test their impact on student achievement. The fitting of Model 3 was again tested against Model 1, and the likelihood statistic (X^2) shows a significant change between Model 1 and Model 3 (p <.001). The likelihood statistic also shows that Model 3 fits the data better than any of the Models 2 (both in mathematics and science) where only one teacher factor is added. Model 3 explains approximately 50% of the total variance of student achievement in mathematics and approximately 45% of the total variance of student achievement in science.

Table 4.9.

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics (Students within classes)

Teacher Factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
Fixed part							
(intercept)	330.3(1.5)	39.2(8.28)	27.1(11.8)	21.6(10.0)	28.6(10.3)	12.2(11.5)	14.9(8.9)
Student Level							
Context							
Prior achievement		0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)
Class Level							
Context							
Prior achievement		0.30(0.03)	0.30(0.03)	0.30(0.03)	0.30(0.03)	0.30(0.03)	0.24(0.03)
Modeling			5.8(3.8)				
Structuring Quant.				8.4(2.9)			
Structuring Qual.					4.5(2.9)		
Application						11.8(3.7)	
Time Management							19.2(2.9)
Variance components							
Class	24.4%	6.3%	5.6%	5.6%	5.8%	5.6%	5.1%
Student	75.6%	47.1%	46.7%	46.8%	46.8%	46.8%	46.7%
Explained		46.6%	47.7%	47.6%	47.4%	47.6%	48.2%
Significance test							
Log-likelihood	103422	98606	96963	96963	96969	96961	96759
Reduction		4816	1643	1643	1637	1645	1847
Degrees of freedom		2	1	1	1	1	1
p value		0.001	0.001	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2e) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.9.

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics (Students within classes) (continued)

Teacher Factors	Model 0	Model 1	Model 2f	Model 2g	Model 2h	Model 2i	Model 3
Fixed part							
(intercept)	330.3(1.55)	39.2(8.28)	-3.8(12.4)	31.8(8.4)	12.3(12.7)	7.8(11.6)	-7.4(13.5)
Student Level							
Context							
Prior achievement		0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)
Class Level							
Context							
Prior achievement		0.30(0.03)	0.30(0.03)	0.25(0.03)	0.30(0.03)	0.30(0.03)	0.22(0.03)
T-S Interactions			18.4(4.1)				
Misbehavior				10.3(2.4)			
Questioning Qual.					10.7(3.7)		
Assessment						12.1(3.2)	
Overarching Quality							13.7(5.3)
Overarching Quantity							21.1(4.4)
Variance components							
Class	24.4%	6.3%	5.5%	5.4%	5.5%	5.6%	4.7%
Student	75.6%	47.1%	46.8%	46.8%	46.8%	46.8%	46.7%
Explained		46.6%	47.7%	47.8%	47.7%	47.6%	48.6%
Significance test							
Log-likelihood	103422	98606	96951	96781	96791	96958	96765
Reduction		4816	1655	1825	1815	1648	1841
Degrees of freedom		2	1	1	1	1	2
p value		0.001	0.001	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2f up to 2i) and for Model 3 the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.10. *Parameter Estimates and (Standard Errors) for the analysis of student achievement in science (Students within classes)*

Teacher Factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2g	Model 3
Fixed part (intercept)	317.3(1.6)	49.5(8.59)	33.4(10.24)	19.1(12.2)	35.7(9.5)	2.3(12.8)	45.4(8.9)	35.3(13.8)	20.6(12.2)	9.29(14.7)
Student Level										
Context										
Prior achievement		0.54(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)
Class Level										
Context										
Prior achievement		0.36(0.03)	0.35(0.03)	0.35(0.03)	0.32(0.03)	0.32(0.03)	0.33(0.03)	0.35(0.03)	0.35(0.03)	0.31(0.03)
Structuring Quant.			8.8(3.2)							
Application				13.7(4.1)						
Time Management					11.7(3.5)					
T-S Interactions						21.9(4.6)				
Misbehavior							5.1(2.6)			
Questioning Qual.								5.8(4.1)		
Assessment									11.8(3.6)	
Overarching Quality										
Overarching Quantity										14.5(5.9)
										12.9(5.0)
Variance components										
Class	31.4%	9.8%	9.1%	9.0%	8.9%	8.7%	9.1%	9.1%	9.0%	8.1%
Student	68.6%	48.4%	48.0%	48.0%	48.0%	48.0%	48.0%	48.0%	48.0%	48.0%
Explained		41.8%	42.9%	43.0%	43.1%	43.3%	42.9%	42.9%	43.0%	43.9%
Significance test										
Log-likelihood	99531	95955	94344	94341	94201	94330	94209	94211	94342	94198
Reduction		3576	1611	1614	1754	1625	1746	1744	1613	1757
Degrees of freedom		2	1	1	1	1	1	1	1	2
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2g) and for Model 3 the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Within-country results

As mentioned in the first Chapter of this thesis, this study aimed not only at providing evidence to the assumption regarding the generic nature of the teacher and school factors of the dynamic model, but also to test their differential effectiveness. Therefore, separate within-country multilevel analyses (students within classrooms) have been conducted so as to examine the effect of these factors on student achievement gains in six rather different educational systems. To conduct the within-country analyses the same approach as the across-country multilevel analyses has been followed.

Namely, for each subject, the first step was to run a two-level model (classroom/teacher level and student level) with no explanatory variables (i.e., empty model) to determine the variance at each level. Then prior achievement and student prior achievement at classroom level were, added to the empty model (i.e., Model 1). Next, the first order factor scores of the within-country SEM models which refer to the classroom-level factors of the dynamic model were added one by one to Model 1. Finally, in Model 3, the two second-order, or overarching, factors (i.e., Quality of Teaching and Quantity of Teaching) were added to Model 1 to test their impact on student achievement.

Based on these results, in order to estimate the relative importance of each factor, the effect sizes of each factor on student achievement in mathematics and science emerged from both the across- and within- country analyses were estimated (see Tables 4.11 and 4.12). The effect sizes have been calculated since they provide the opportunity to compare the level of impact (i.e., strength) of each factor on student achievement on an equal basis and assist the interpretation of results (Rosenthal, 1994; Rosenthal, Rosnow & Rubin, 2000; Thalheimer & Cook, 2002). By converting the unstandardized factor effects into standardized and using Cohen's d it is possible to determine the magnitude of the effect of each factor on the outcome measure (i.e., student achievement in mathematics and science) in each of the six participating countries. It is suggested that effect sizes of .20 are

ANASTASIA PANAYIOTOU

considered as small, .50 are considered as medium, and .80 are considered as large Cohen (1992).

Specifically, the fixed effects obtained with multilevel analysis were converted to standardized effects or Cohen's d by following the approach proposed by Elliot & Sammons (2004) which expresses the change on the outcome measure that will be produced in terms of a standard deviation on the continuous independent variable, standardized by the within school standard deviation adjusted for covariates in the model. Thus, the relative strength of the effects of each factor can be compared more easily across the two outcome measures and across the participating countries assisting the investigation of differential effects. The effect sizes of each teacher factor on student achievement in mathematics and science are presented in Tables 4.11 and 4.12.

Based on Tables 4.11 and 4.12, the following observations can be made. First, it is shown that the teacher factors of the dynamic model were found to have an effect ranging from 0.11 to 0.23. These values can be considered as small, however they can be explained by the fact that he variance in the functioning of the factors in the participating countries was also very small (i.e., smaller than 0.80) and lack of enough variance may not allow for the effect of the factors to be more evident.

Secondly, the results presented in Tables 4.11 and 4.12, seem to provide further support to the generic nature of the teacher factors in regard to their functioning in two different subjects, since only small differences in the effect sizes of the factors were found in mathematics and science achievement. The generic nature of the teacher factors also seems to be supported in terms of their effectiveness in the six countries of the study since the effect sizes for each factor do not significantly vary across the six participating countries and no patterns are observed (i.e., a factor to systematically not have an effect in all or most countries). However, in Germany there were more cases of factors where it was not possible to identify statistically significant effects looking at both subjects. Someone

ANASTASIA PANAYIOTOU

could therefore argue that the teacher factors have differential effects, being more important in some countries than others and that fewer factors are important for Germany. However, one should bear in mind that the variance of the factors in Germany was small (i.e., smaller than 0.40) and did not provide enough statistical power to identify the effect of all factors at level .05 in both subjects.

Table 4.11

Effects of each teacher factor on student achievement in mathematics (as expressed by Cohen's d) in each country and across countries.

Teacher Factors	Across Countries	Belgium	Cyprus	Germany	Greece	Ireland	Slovenia
Modeling	0.17	0.16	0.19	N.S.S.	0.19	0.21	0.18
Structuring Quant.	0.18	0.19	N.S.S.	N.S.S.	0.20	N.S.S.	N.S.S.
Structuring Qual.	0.20	0.18	0.17	N.S.S.	0.18	0.18	N.S.S.
Misbehavior	0.21	0.18	0.19	0.17	0.19	0.22	0.23
Application	0.22	0.19	0.20	N.S.S.	0.18	N.S.S.	N.S.S.
Management of Time	0.19	0.18	0.20	0.21	0.18	0.22	0.20
Assessment	0.18	N.S.S.	0.20	0.14	0.23	0.19	N.S.S.
Questioning Qual.	0.20	0.22	0.19	N.S.S.	0.21	0.20	N.S.S.
T-S Interactions	0.19	0.17	0.19	0.16	0.20	0.18	N.S.S.
Questioning Quant.	N.S.S.	0.17	0.16	N.S.S.	0.15	0.17	0.15
Overarching Quality	0.21	0.19	0.20	0.14	0.19	0.22	0.21
Overarching Quantity	0.20	0.20	0.18	0.22	0.18	0.21	0.20

ANASTASIA PANAYIOTOU

In some cases the factors were found to have a statistically significant effect at level .10, however these factors were noted as not statistically significant (N.S.S.) in Tables 4.11 and 4.12. It should also be noted that no factor was found to have a negative effect on student achievement in mathematics and science in either country.

Table 4.12

Effects of each teacher factor on student achievement in science (as expressed by Cohen's d) in each country and across countries.

Teacher Factors	Across Countries	Belgium	Cyprus	Germany	Greece	Ireland	Slovenia
Modeling	N.S.S.	N.S.S.	0.14	0.15	0.12	0.14	0.16
Structuring Quant.	0.19	N.S.S.	N.S.S.	N.S.S.	0.18	0.19	0.18
Misbehavior	0.21	0.19	N.S.S.	0.21	N.S.S.	0.20	0.19
Application	0.18	N.S.S.	0.20	N.S.S.	0.17	0.16	0.19
T-S Interactions	0.18	0.17	0.21	N.S.S.	0.19	0.19	0.20
Management of Time	0.19	0.17	0.22	0.18	0.23	0.17	0.20
Assessment	0.21	N.S.S.	0.18	N.S.S.	0.18	0.20	0.19
Questioning Qual.	0.22	N.S.S.	0.16	0.15	0.20	0.19	0.21
Questioning Quant.	N.S.S.	0.15	0.14	N.S.S.	0.18	0.11	0.17
Overarching Quality	0.21	N.S.S.	0.23	0.17	0.18	0.19	0.22
Overarching Quantity	0.19	0.21	0.19	0.18	0.20	0.17	0.19

ANASTASIA PANAYIOTOU

Using Multilevel Analysis to Search for the Impact of the School Factors on Student Achievement

Across-country results

For the analysis of the impact of the school factors on students' achievement in mathematics and science the same procedure as with the teacher factors was used. The first step was again to run a two-level model (school level and student level) without any explanatory variables (empty model) to determine the variance at each level. The variance for mathematics was 23.7% at the school level and 76.3% at the student level and was statistically significant in each level. The variance for science was 30.9% at the school level and 69.1% at the student level and was also statistically significant in each level. This suggests that multilevel analysis can be used in order to identify the impact of each factor. Also, a comparison of the empty models in mathematics and science reveals that the effect of the school level is more distinct on achievement in science rather than in mathematics.

In Model 1 the context variables of the students' prior achievement and the students' prior achievement at the school level were added to the empty model. Both context variables had a statistically significant effect at .05 level on each outcome. Model 1 explains approximately 48% of the total variance of student achievement in mathematics and 43% in science and most of the explained variance is at the student level.

Then, for each student outcome (mathematics and science), different versions of Model 2 were established (i.e., Models 2a-2i for mathematics and Models 2a-2h for science). In each version of Model 2, the first order factor scores of the SEM models which refer to the school-level factors of the dynamic model were added one by one to Model 1. Thus, the fitting of each of these models was tested against Model 1, and the likelihood statistic (X^2) shows a significant change between Model 1 and each version of Model 2 ($p < .001$). This suggests that variables measuring the school factors have significant effects

ANASTASIA PANAYIOTOU

on student achievement in mathematics and science. As can be seen in Table 4.13, all the first order factors were found to have a statistically significant effect on achievement in mathematics. In science only the factor concerning policy on quantity of teaching was not found to have a statistically significant effect and thus it is not included in Table 4.14. Each version of Model 2 explains approximately 51% of the total variance of student achievement in mathematics and approximately 49% of the total variance of student achievement in science.

Finally, in Models 3a-3c the three overarching factors were added separately in Model 1 to examine their effect on student achievement. All three overarching factors were found to have a statistically significant effect on student achievement in both subjects. By comparing the fitting of each version of Model 3 (models 3a-3c) with Model 1 and the likelihood statistic, one may observe a significant change between Model 1 and each version of Model 3 ($p < .001$). Each version of Model 3 explains approximately 52% of the total variance of student achievement in mathematics and approximately 50% of the total variance of student achievement in science. However, it was not possible to add the three overarching factors together in one model due to multicollinearity issues. When adding the three overarching factors simultaneously in one model only the overarching factor of the policy on SLE was found to have a significant effect.

ANASTASIA PANAYIOTOU

Table 4.13 *Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics (Students within schools) (models 0, 1, 2a-2i)*

School Factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2g	Model 2h	Model 2i
Fixed part (intercept)	330.5(1.9)	33.7(9.9)	31.9(12.6)	34.1(14.4)	29.5(13.5)	29.8(13.3)	32.2(12.1)	26.5(13.04)	39.1(12.2)	37.1(13.5)	30.6(12.8)
Student Level											
Prior achievement		0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)
School Level											
Prior achievement		0.32(0.03)	0.22(0.04)	0.23(0.04)	0.23(0.04)	0.23(0.04)	0.22(0.04)	0.22(0.04)	0.23(0.04)	0.23(0.04)	0.23(0.04)
Eval. of Teaching			11.2(3.4)								
Eval. of SLE				9.3(4.2)							
Teacher Collab.					12.3(4.1)						
Resources						10.6(3.4)					
Relation with community							11.3(2.9)				
Partnership								14.6(3.9)			
Quantity of Teaching									8.04(3.1)		
Learning opp.										6.9(3.1)	
Quality of Teaching											9.7(2.9)
Variance components											
School	23.7%	4.9%	3.2%	3.3%	3.2%	3.2%	3.4%	3.2%	3.2%	3.3%	3.2%
Student	76.3%	47.5%	45.3%	45.3%	45.4%	45.3%	45.3%	45.3%	45.3%	45.3%	45.3%
Explained		47.6%	51.5%	51.4%	51.4%	51.5%	51.3%	51.5%	51.5%	51.4%	51.5%
Significance test											
Loglikelihood	103307	98607	77846	77852	77847	77847	77842	77843	77850	77851	77845
Reduction		4700	20761	20755	20760	20760	20765	20764	20757	20756	20762
Degrees of freedom		2	1	1	1	1	1	1	1	1	1
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2i) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.13

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics (Students within schools) (models 0, 1, 3a-3c) (continued).

School Factors	Model 0	Model 1	Model 3a	Model 3b	Model 3c
Fixed part (intercept)	330.5(1.9)	33.7(9.9)	30.5(13.6)	24.2(14.0)	30.6(13.3)
Student Level					
Prior achievement		0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)
School Level					
Prior achievement		0.32(0.03)	0.23(0.04)	0.22(0.04)	0.2(0.04)
Overarching Evaluation			11.8(4.1)		
Overarching SLE				17.6(5.2)	
Overarching Policy					
Teaching					10.9(3.5)
Variance components					
School	23.7%	4.9%	2.7%	2.4%	2.6%
Student	76.3%	47.5%	45.3%	45.3%	45.3%
Explained		47.6%	52.0%	52.3%	52.1%
Significance test					
Loglikelihood	103307	98607	77848	77845	77847
Reduction		4700	20759	20762	20760
Degrees of freedom		2	1	1	1
p value		0.001	0.001	0.001	0.001

Note: For each alternative Model 3 (i.e., Models 3a up to 3c) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.14 *Parameter Estimates and (Standard Errors) for the analysis of student achievement in science (Students within schools) (models 0, 1, 2a-2h)*

School Factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e	Model 2f	Model 2g	Model 2h
Fixed part (intercept)	318.03(2.1)	40.4(10.2)	29.7(13.1)	35.6(15.4)	28.4(14.1)	34.5(13.9)	34.7(12.0)	31.3(13.8)	38.9(3.5)	34.1(12.9)
Student Level										
Prior achievement		0.54(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)
School Level										
Prior achievement		0.39(0.04)	0.32(0.04)	0.34(0.04)	0.32(0.04)	0.34(0.04)	0.31(0.04)	0.32(0.04)	0.33(0.04)	0.32(0.04)
Eval. of Teaching			12.6(3.9)							
Eval. of SLE				8.7(4.8)						
Teacher Collaboration					13.5(4.7)					
Resources						8.8(3.8)				
Relation with community							11.9(3.5)			
Partnership								12.5(4.6)		
Learning opportunities									6.5(3.5)	
Quality of Teaching										9.3(3.3)
Variance components										
School	30.9%	7.8%	4.8%	5.0%	4.8%	5.0%	4.7%	4.9%	5.0%	4.9%
Student	69.1%	48.9%	46.2%	46.2%	46.2%	46.2%	46.2%	46.2%	46.2%	46.2%
Explained		43.3%	49.0%	48.8%	49.0%	48.8%	49.1%	48.9%	48.8%	48.9%
Significance test										
Loglikelihood	99395	95962	75645	75652	75647	75650	75644	75648	75652	75647
Reduction		3433	20317	20310	20315	20312	20318	20314	20310	20315
Degrees of freedom		2	1	1	1	1	1	1	1	1
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2h) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.14

Parameter Estimates and (Standard Errors) for the analysis of student achievement in science (Students within schools) (models 0, 1, 3a-3c) (continued).

School Factors	Model 0	Model 1	Model 3a	Model 3b	Model 3c
Fixed part (intercept)	318.03(2.1)	40.4(10.2)	29.5(14.3)	24.7(14.5)	36.1(13.7)
Student Level					
Prior achievement		0.54(0.01)	0.56(0.01)	0.56(0.01)	0.56(0.01)
School Level					
Prior achievement		0.39(0.04)	0.33(0.04)	0.32(0.04)	0.33(0.04)
Overarching Eval.			12.5(4.6)		
Overarching SLE				18.5(6.1)	
Overarching Teaching					8.9(4.1)
Variance components					
School	30.9%	7.8%	3.9%	3.5%	4.2%
Student	69.1%	48.9%	46.2%	46.2%	46.2%
Explained		43.3%	49.9%	50.3%	49.6%
Significance test					
Loglikelihood	99395	95962	75648	75646	75650
Reduction		3433	20314	20316	20312
Degrees of freedom		2	1	1	1
p value		0.001	0.001	0.001	0.001

Note: For each alternative Model 3 (i.e., Models 3a up to 3c) the reduction is estimated in relation to the deviance of Model 1.

Within-country results

To examine whether the school factors of the dynamic model have a greater effect on student achievement gains in some countries than others and test their possible differential effects, within-country multilevel analyses have been conducted, the results of which have been used to estimate the effect size of each school factor. To conduct the within-country analyses the same approach as the across-country multilevel analyses has been followed. To that end, for each subject, a two-level model (school level and student level) without any explanatory variables (empty model) was established. Then, prior achievement and student prior achievement at school level were, added to the empty model (i.e., Model 1). Next, the first order factor scores of the SEM models which refer to the school-level factors of the dynamic model were added one by one to Model 1 to establish different versions of Model 2. Finally, in Models 3a-3c the three overarching factors were added separately in Model 1 to examine their effect on student achievement in the two outcomes.

Then, to be in a position to compare the impact of each school factor, the effect sizes of each factor on student achievement in mathematics and science were estimated. Specifically, following again the approach proposed by Elliot & Sammons (2004), Cohen's d was calculated and the fixed effects obtained with multilevel analysis were converted to standardized effects. Tables 4.15 and 4.16 illustrate the effects of each school factor on student achievement in mathematics and science, respectively, in each country and across countries. Based on the results provided in Tables 4.15 and 4.16, the following observations arise. First, it is shown that the effect of the school factors of the dynamic model ranges from 0.13 to 0.21. These values can be considered as small; however they are consistent to the average effect sizes that were shown through two recent meta-analyses searching for the impact of school factors on student outcomes (see, Kyriakides, Creemers, Antoniou & Demetriou, 2010; Scheerens, et al., 2005).

The second observation that arises concerns the question in regard to the generic nature of the factors. The results illustrated in Tables 4.15 and 4.16, provide further support to the generic nature of the school factors in regard to their functioning in two different subjects, since only small differences in the effect sizes of the factors were found in mathematics and science achievement. Namely, slightly higher effect sizes were observed in regard to science which however can be justified by the fact that the variance at the school level was more pronounced in science (see empty models of Tables 4.13 and 4.14). Having more variance may allow for the effect of the factors to be more evident. The generic nature of these factors was also supported in terms of their effectiveness in the five countries of the study. Specifically, it is evident that the effect sizes for each factor do not significantly vary across the five participating countries and no patterns are observed; namely some factors to systematically not have a statistically significant effect in all or some countries. In some cases it was not possible to identify statistically significant effects for the factors in all countries. However, in those cases the variance in the functioning of the factors in the respective countries was very small (i.e., smaller than 0.85).

Table 4.15

Effects of each school factor on student achievement in mathematics (as expressed by Cohen's d) in each country and across countries.

Scales	Across Countries	Belgium	Cyprus	Germany	Ireland	Slovenia
Scale 1: School policy on teaching						
1) Quantity of teaching	0.14	0.14	0.15	0.12	N.S.S	0.11
2) Provision of learning opportunities	0.13	0.11	N.S.S	0.13	0.12	N.S.S
3) Quality of teaching	0.18	0.15	0.17	0.16	N.S.S	0.14
Scale 2: Policy on the school learning environment (SLE)						
1) Partnership policy	0.18	N.S.S.	0.16	0.15	0.15	N.S.S
2) Relation with the school community	0.19	N.S.S	0.19	0.18	0.17	0.17
3) Teacher collaboration	0.15	0.10	0.13	0.12	0.14	0.13
4) Provision of learning resources to students and teachers	0.16	N.S.S	0.18	0.17	0.15	0.16
Scale 3: Evaluation of the school policy on teaching	0.16	0.13	0.14	0.16	N.S.S	0.15
Scale 4: Evaluation of the SLE	0.13	N.S.S	0.12	0.14	0.12	0.11

ANASTASIA PANAYIOTOU

Table 4.16
Effects of each school factor on student achievement in science (as expressed by Cohen's d) in each country and across countries.

Scales	Across Countries	Belgium	Cyprus	Germany	Ireland	Slovenia
Scale 1: School policy on teaching						
1) Quantity of teaching	N.S.S	0.12	N.S.S.	0.11	N.S.S	0.10
2) Provision of learning opportunities	0.14	0.11	0.13	0.12	N.S.S	0.12
3) Quality of teaching	0.18	0.16	0.19	N.S.S	0.18	0.17
Scale 2: Policy on the school learning environment (SLE)						
1) Partnership policy	0.19	0.15	0.17	0.18	0.17	N.S.S
2) Relation with the school community	0.20	N.S.S	0.17	0.18	N.S.S	0.17
3) Teacher collaboration	0.19	0.09	0.12	0.11	0.10	0.10
4) Provision of learning resources to students and teachers	0.15	N.S.S	0.13	0.14	0.12	0.11
Scale 3: Evaluation of the school policy on teaching	0.21	0.17	N.S.S	0.18	0.19	0.17
Scale 4: Evaluation of the SLE	0.14	N.S.S	0.12	0.13	0.13	0.11

Finally, the effect sizes presented in Tables 4.15 and 4.16 may also provide suggestions for policy development. Namely, some factors were found to have a somewhat greater effect on student achievement, such as school policy on the quality of teaching, partnership policy and policy on the relation with the school community. These differences in the effect sizes were small (not greater than 0.07), but could have policy implications. Specifically, based on these results, improvement priorities could be set and in order for

ANASTASIA PANAYIOTOU

schools to be effective they could develop a policy on teaching as well as a policy on establishing a learning environment in both subjects (i.e., mathematics and science).

Using Multivariate Analysis to Search for Differential Effects of the Teacher and School Factors on Student Achievement in two Different Subjects

As mentioned in the last section of Chapter 3, multivariate multilevel analysis was chosen as the next step of data analyses as it would provide the opportunity of not only identifying the effects of the teacher and school factors on student achievement gains but also to serve two more purposes; a) searching for possible differential effects of the factors on the two outcomes and b) estimate the correlations between the two outcomes (i.e., mathematics and science achievement). In the next two sections, the results of the multivariate analyses conducted to examine the joint effects of the teacher and school factors, respectively on student achievement in mathematics and science are presented.

Using Multivariate Analysis to Search for Differential Effects of the Teacher Factors on Student Achievement in Mathematics and Science – Across Country Results

The first step for conducting the multivariate analysis was to run a three-level multivariate model with no explanatory variables (i.e., empty model), to determine the variance at classroom and student level for each of the two outcomes. In this model, the first level was that of the dependent variables indexed by h=1 and 2, where 1= mathematics and 2= science. The second level was that of the individuals (i.e., student level) and the third level was that of groups (i.e., class/teacher level). The results of this analysis are presented in Table 4.17. Then, in Model 1, students' prior achievement was added to the empty model and was found to have a statistically significant effect at level .05 in both subjects.

At the next step, similarly to the separate multilevel analyses, different versions of Model 2 were established (i.e., Models 2a-2i). In each version of Model 2, the first-order factor scores of the SEM models which refer to the teacher factors of the dynamic model were added one by one to Model 1 by using two different approaches. First, each first-order teacher factor was added in Model 1 as "separate coefficients" for mathematics and science in order to identify differential effects in the two subjects. The fitting of each of these models was tested against Model 1, and the likelihood statistic (X^2) showed a significant change between Model 1 and each version of Model 2 ($p < .001$). This implies that variables measuring the teacher factors have significant effects on student achievement in mathematics and science. However, another approach was followed where each first-order teacher factor was added in Model 1 as a "common coefficient" for the two subjects. The fitting of these models was again tested against Model 1 and the likelihood statistic (X^2) showed a significant change between Model 1 and each version of these models. However, the fitting of these models was also compared to the fitting of the models of the first approach where separate estimates were yielded for the factors, for each outcome and by looking at the log likelihood it was shown that in some cases these models had a better fit to the data. Table 4.17 presents the models that were found to best fit the data. The factors which were not found to have a statistically significant effect on achievement in the two subjects are not included in Table 4.17. It may also be observed that no differential effects were found in the two subjects; however the results provide further support to the generic nature of the teacher factors.

Finally, in Model 3, the two second-order, factors (i.e., Quality of Teaching and Quantity of Teaching) were added to Model 1 to test their impact on student achievement in each of the two subjects. Both overarching factors were found to have a statistically significant effect on student achievement, with the effect of the overarching factor related to quantity of teaching being more evident in both subjects. The fitting of Model 3 was

ANASTASIA PANAYIOTOU

again tested against Model 1, and the likelihood statistic (X^2) shows a significant change between Model 1 and Model 3 ($p < .001$). The likelihood statistic also shows that Model 3 fits the data better than any of the Models 2 where only one teacher factor is added.

Table 4.17
Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within classes)

Teacher Factors	Model 0	Model 1	Model 2a	Model 2b
Fixed part (maths)	329.6(1.5)	140.7(2.7)	121.6(7.1)	140.1(2.7)
Fixed part (science)	316.8(1.6)	165.6(2.7)	137.9(7.9)	164.9(2.7)
Student Level				
Context				
Prior achievement (maths)		0.6(0.01)	0.6(0.01)	0.6(0.01)
Prior achievement (sci)		0.5(0.01)	0.5(0.01)	0.5(0.01)
Class Level				
Structuring Quant.(maths)			9.0(3.3)	
Structuring Quant.(sci)			13.3(3.7)	
Structuring Qual.				11.7(3.2)
Random part				
Class Level				
Var (maths)	1148.2(82.3)	391.7(31.6)	378.9(31.0)	374.5(30.7)
Var (sci)	1299.0(88.9)	571.3(41.6)	544.4(40.3)	554.9(40.9)
Cov(maths,sci)	1067.7(78.4)	386.8(31.2)	370.6(30.4)	372.5(30.5)
Student Level				
Var (maths)	3471.7(52.4)	2160.9(32.6)	2147.5(32.7)	2147.6(32.7)
Var (sci)	2744.0(41.8)	1931.9(29.5)	1917.9(29.4)	1917.7(29.5)
Cov(maths,sci)	1655.4(37.6)	540.3(22.8)	529.6(22.8)	530.6(22.8)
Significance test				
Loglikelihood	199515	193938	190711	190710
Reduction		5577	3227	3228
Degrees of freedom		2	2	1
p value		0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2i) and Model 3 the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.17

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within classes) (continued)

Teacher Factors	Model 2c	Model 2d	Model 2e	Model 2f
Fixed part (maths)	139.8(2.7)	71.5(11.9)	139.8(2.7)	139.8(2.7)
Fixed part (science)	164.6(2.7)	68.0(13.3)	164.6(2.7)	164.8(2.7)
Student Level				
Context				
Prior achievement (maths)	0.6(0.01)	0.6(0.01)	0.6(0.01)	0.6(0.01)
Prior achievement (sci)	0.5(0.01)	0.5(0.01)	0.5(0.01)	0.5(0.01)
Class Level				
Application	15.6(4.0)			
T-S interactions (maths)		26.3(4.5)		
T-S interactions (sci)		37.3(5.1)		
Assessment			15.5(3.5)	
Questioning Qual.				13.1(4.1)
Random part				
Class Level				
Var (maths)	375.9(30.9)	356.4(29.6)	373.4(30.8)	367.9(30.5)
Var (sci)	542.3(40.4)	496.8(37.4)	537.7(40.2)	545.0(40.7)
Cov(maths,sci)	366.9(30.4)	338.1(28.4)	363.1(30.2)	363.7(30.2)
Student Level				
Var (maths)	2147.4(32.7)	2147.2(32.7)	2147.1(32.7)	2145.6(32.7)
Var (sci)	1917.9(29.5)	1918.8(29.5)	1918.1(29.5)	1918.7(29.5)
Cov(maths,sci)	530.1(22.8)	530.4(22.8)	530.2(22.8)	530.5(22.9)
Significance test				
Loglikelihood	190709	190671	190705	190405
Reduction	3229	3267	3233	3533
Degrees of freedom	1	2	1	1
p value	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2i) and Model 3 the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.17

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within classes) (continued)

Teacher Factors	Model 2g	Model 2h	Model 2i	Model 3
Fixed part (maths)	141.1(2.7)	140.9(2.7)	110.1(6.9)	141.8(2.7)
Fixed part (science)	165.9(2.7)	165.7(2.7)	123.8(7.7)	166.1(2.7)
Student Level				
Context				
Prior achievement (maths)	0.6(0.01)	0.6(0.01)	0.6(0.01)	0.6(0.01)
Prior achievement (sci)	0.5(0.01)	0.5(0.01)	0.5(0.01)	0.5(0.01)
Class Level				
Time Management	28.9(2.9)			
Misbehaviour		18.6(2.3)		
Questioning Quant. (maths)			14.5(3.2)	
Questioning Quant. (sci)			20.2(3.6)	
Quality of Teaching				14.5(5.5)
Quantity of Teaching				38.8(4.1)
Random part				
Class Level				
Var (maths)	303.3(26.5)	324.5(27.8)	360.0(29.9)	299.0(26.2)
Var (sci)	488.8(37.1)	515.6(38.7)	515.1(38.6)	468.3(35.8)
Cov(maths,sci)	303.3(26.5)	327.3(27.9)	346.7(29.0)	290.7(25.8)
Student Level				
Var (maths)	2146.7(32.7)	2146.6(32.7)	2145.9(32.7)	2146.6(32.7)
Var (sci)	1918.8(29.5)	1918.6(29.5)	1919.2(29.5)	1919.2(29.5)
Cov(maths,sci)	533.4(22.9)	533.1(22.9)	532.4(22.9)	535.2(22.9)
Significance test				
Loglikelihood	190327	190355	190383	190313
Reduction	3611	3583	3555	3625
Degrees of freedom	1	1	2	2
p value	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2i) and Model 3 the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Apart from searching for differential effects, the second aim of the multivariate analysis was to assist the estimation of correlations between the residuals of the two outcome variables (i.e., mathematics and science achievement) at each level (i.e. classroom/teacher level and student level) separately.

The following formula was used to calculate the correlations at the classroom and student level, respectively:

$$r(u1j, u2j) = \frac{Cov(u1j, u2j)}{\sqrt{Var(u1j) * Var(u2j)}}$$

$$r(\text{R1ij}, \text{R2ij}) = \frac{Cov(\text{R1ij}, \text{R2ij})}{\sqrt{Var(\text{R1ij}) * Var(\text{R2ij})}}$$

Where 1=mathematics and 2= science

The correlations between the residuals of the two outcome variables at each level are provided in Table 4.18 for all the models that occurred from the multivariate multilevel analyses (i.e., Models: 0, 1, 2a-2i and 3) and the following observations arise. First, it is shown that in all the models the correlation between the residuals for mathematics and science at the classroom level was higher than .78. This implies that teacher effectiveness is related in the two subjects. Specifically, from the correlation in Model 0 it can be assumed that teachers who are effective in promoting student achievement in mathematics are also effective in promoting student achievement in science. Model 1, where student prior achievement is introduced demonstrates that teachers who are effective in promoting student achievement gains in mathematics are also effective in promoting student achievement gains in science. Finally, looking at the correlations in Models 2a-3 it can be argued that teachers who are effective in terms of using the factors of the dynamic model to promote student achievement gains in one subject are also effective in the other subject.

ANASTASIA PANAYIOTOU

At the student level the correlation between the residuals for mathematics and science for the null model was .54. These results imply that students who achieve well in one subject also achieve well in the other. The correlation in the next models is however reduced which can be explained by the fact that the factors that are introduced concern the teacher level instead of the student.

Table 4.18

Correlations between the residuals of mathematics and science at classroom and student level

	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d
Class Level						
r	0.87	0.82	0.82	0.82	0.81	0.80
Student Level						
r	0.54	0.26	0.26	0.26	0.26	0.26
	Model 2e	**Model 2f**	**Model 2g**	**Model 2h**	**Model 2i**	**Model 3**
Class Level						
r	0.81	0.81	0.79	0.80	0.81	0.78
Student Level						
r	0.26	0.26	0.26	0.26	0.26	0.26

Using Multivariate Analysis to Search for Differential Effects of the School Factors on Student Achievement in Mathematics and Science – Across Country Results

For the analysis of the impact of the school factors on students' achievement in mathematics and science the same procedure as with the multivariate analysis on the teacher factors was used. First a three-level model without any explanatory variables (empty model) was run, to determine the variance at each level for the two outcomes. In

ANASTASIA PANAYIOTOU

this model, the first level was again that of the dependent variables (i.e., mathematics and science). The second level was that of the individuals (i.e., student level) and the third level was that of groups (i.e., school level). The results of this analysis are presented in Table 4.19. In Model 1, students' prior achievement was added to the empty model and was found to have a statistically significant effect at level .05 in both subjects.

Then, different versions of Model 2 were established (i.e., Models 2a-2h). In each version of Model 2, the first-order factor scores of the SEM models which refer to the school factors of the dynamic model were added one by one to Model 1 as separate coefficients for mathematics and science to identify possible differential effects in the two subjects. The fitting of each of these models was tested against Model 1, and the likelihood statistic (X^2) shows a significant change between Model 1 and each version of Model 2 ($p <.001$). This shows that variables measuring the school factors have significant effects on student achievement in mathematics and science. The results from the multivariate analysis in regard to the school factors are also consistent with the results obtained by the separate multilevel analyses for the two subjects.

However, the second approach was also followed where each first-order teacher factor was added in Model 1 as a common coefficient for the two subjects. The fitting of these models was compared to the fitting of the models of the first approach and by looking at the loglikelihood it was shown that in some cases these models had a better fit to the data. Table 4.19 presents the models that were found to best fit the data. The factors which were not found to have a statistically significant effect on achievement in the two subjects are not included in Table 4.19. Finally, in Models 3a-3c the three overarching factors were added separately in Model 1 to examine their effect on student achievement in each of the two subjects. All three overarching factors were found to have a statistically significant effect on student achievement in both subjects, with the effect of the overarching factor related to the policy on the SLE being more evident. By comparing the

ANASTASIA PANAYIOTOU

fitting of each version of Model 3 with Model 1 and the likelihood statistic, one may observe a significant change between Model 1 and each version of Model 3 (p <.001).

Table 4.19
Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within schools)

School Factors	Model 0	Model 1	Model 2a
Fixed part (maths)	330.1(1.9)	142.9(2.8)	137.3(3.0)
Fixed part (science)	317.7(2.1)	168.2(2.9)	164.7(3.1)
Student Level			
Context			
Prior achievement (maths)		0.6(0.01)	0.7(0.01)
Prior achievement (sci)		0.5(0.01)	0.5(0.01)
School Level			
Evaluation Teaching			12.7(3.8)
Random part			
School Level			
Var (maths)	1136.9(101.6)	371.6(36.4)	189.0(24.4)
Var (sci)	1294.6(111.9)	541.3(49.5)	328.1(36.9)
Cov(maths,sci)	1082.2(99.3)	382.1(37.8)	198.4(25.6)
Student Level			
Var (maths)	3542.6(52.8)	2206.6(32.9)	2104.7(35.1)
Var (sci)	2814.8(42.3)	1992.7(29.9)	1876.1(31.7)
Cov(maths,sci)	1706.9(38.1)	577.8(23.2)	516.8(24.5)
Significance test			
Loglikelihood	199330	193915	153046
Reduction		5415	40869
Degrees of freedom		2	1
p value		0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2h) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.19

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within schools) (continued)

School Factors	Model 2b	Model 2c	Model 2d	Model 2e
Fixed part (maths)	137.4(3.0)	137.3(3.0)	107.4(8.4)	137.6(3.0)
Fixed part (science)	164.9(3.1)	164.9(3.1)	119.9(9.9)	165.1(3.1)
Student Level				
Context				
Prior achievement (maths)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)
Prior achievement (sci)	0.5(0.01)	0.5(0.01)	0.5(0.01)	0.5(0.01)
School Level				
Teacher Collaboration	13.7(4.6)			
Use of Resources		9.2(3.8)		
Relation with the community (maths)			12.8(3.4)	
Relation with the community (sci)			19.2(4.1)	
Partnership Policy				17.4(4.4)
Random part				
School Level				
Var (maths)	191.6(24.6)	193.8(24.8)	184.3(23.9)	185.7(24.1)
Var (sci)	331.4(37.2)	343.7(38.4)	309.3(35.2)	329.9(37.1)
Cov(maths,sci)	201.5(25.9)	208.7(26.5)	189.5(24.7)	197.6(25.5)
Student Level				
Var (maths)	2104.8(35.1)	2104.5(35.1)	2104.6(35.2)	2104.3(35.1)
Var (sci)	1875.9(31.7)	1875.8(31.6)	1876.6(31.7)	1875.8(31.6)
Cov(maths,sci)	516.8(24.5)	516.4(24.5)	516.9(24.5)	516.6(24.5)
Significance test				
Loglikelihood	153048	153051	153035	153042
Reduction	40867	40864	40880	40873
Degrees of freedom	1	1	2	1
p value	0.001	0.001	0.001	0.001

Note: For each alternative Model 2 (i.e., Models 2a up to 2h) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.19

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within schools) (continued)

School Factors	Model 2f	Model 2g	Model 2h
Fixed part (maths)	122.8(7.5)	137.3(3.0)	137.4(3.0)
Fixed part (science)	160.2(8.9)	164.9(3.1)	164.9(3.1)
Student Level			
Context			
Prior achievement (maths)	0.7(0.01)	0.7(0.01)	0.7(0.01)
Prior achievement (sci)	0.5(0.01)	0.5(0.01)	0.5(0.01)
School Level			
Policy on Quantity of Teaching (maths)	7.3(3.4)		
Policy on Quantity of Teaching (sci)	2.4(4.3)*		
Provision of Learning Opportunities		6.4(3.4)**	
Policy on Quality of Teaching			10.4(3.2)
Random part			
School Level			
Var (maths)	195.5(24.9)	197.6(25.2)	191.5(24.6)
Var (sci)	348.6(38.8)	343.5(38.3)	331.7(37.2)
Cov(maths,sci)	212.9(26.8)	210.5(26.7)	201.7(25.9)
Student Level			
Var (maths)	2104.7(35.1)	2104.6(35.1)	2104.5(35.1)
Var (sci)	1875.9(31.7)	1875.8(31.6)	1875.8(31.6)
Cov(maths,sci)	516.9(24.5)	516.7(24.5)	516.6(24.5)
Significance test			
Loglikelihood	153051	153053	153047
Reduction	40864	40862	40868
Degrees of freedom	2	1	1
p value	0.001	0.001	0.001

*Not Statistically Significant (NSS)

**Statistically Significant at.10 level

ANASTASIA PANAYIOTOU

Table 4.19

Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics and science (Students within schools) (continued)

School Factors	Model 3a	Model 3b	Model 3c
Fixed part (maths)	137.3(3.0)	137.5(3.0)	137.3(3.0)
Fixed part (science)	164.8(3.1)	164.9(3.1)	164.9(3.1)
Student Level			
Context			
Prior achievement (maths)	0.7(0.01)	0.7(0.01)	0.7(0.01)
Prior achievement (sci)	0.5(0.01)	0.5(0.01)	0.5(0.01)
School Level			
Overarch. Policy on Evaluation	11.8(4.6)		
Overarch. Policy on the SLE		21.6(5.9)	
Overarch. Policy on Teaching			10.3(3.9)
Random part			
School Level			
Var (maths)	193.1(24.8)	188.7(24.4)	193.8(24.8)
Var (sci)	337.3(37.7)	324.5(36.6)	340.9(38.1)
Cov(maths,sci)	205.1(26.2)	196.6(25.5)	207.3(26.4)
Student Level			
Var (maths)	2104.7(35.1)	2104.5(35.1)	2104.5(35.1)
Var (sci)	1876.0(31.7)	1875.9(31.7)	1875.8(31.6)
Cov(maths,sci)	516.7(24.5)	516.7(24.5)	516.5(24.5)
Significance test			
Loglikelihood	153050	153044	153050
Reduction	40865	40871	40865
Degrees of freedom	1	1	1
p value	0.001	0.001	0.001

Note: For each alternative Model 3 (i.e., Models 3a up to 3c) the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Then, the correlations between the residuals of the two outcome variables at each level (i.e. school level and student level) were estimated for all the models that occurred from the multivariate multilevel analyses (i.e., Models: 0, 1, 2a-2h and 3a-3c) using the same procedure described in the previous section (see Table 4.20). Based on these correlations the following observations arise.

First, a high correlation was observed between the residuals for mathematics and science at the school level in Model 0 (*r* =.89), implying that schools who are effective in promoting student achievement in mathematics are also effective in promoting student achievement in science. This correlation was high in all the Models (*r* > .78) thus providing support to the argument that schools which are effective in terms of the different aspects of their school policy in one subject are also effective in the other subject managing to promote student achievement gains.

At the student level the correlation between the residuals for mathematics and science for the null model was .54 showing a relation between student achievement in the two outcomes. The correlation in the next models is however reduced which can be explained by the fact that the factors that are introduced concern the school level instead of the student.

Table 4.20

Correlations between the residuals of mathematics and science at school and student level

	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 2d	Model 2e
School Level							
r	0.89	0.85	0.80	0.80	0.81	0.79	0.80
Student Level							
r	0.54	0.28	0.26	0.26	0.26	0.26	0.26
	Model 2f	Model 2g	Model 2h	Model 3a	Model 3b	Model 3c	
School Level							
r	0.82	0.81	0.80	0.80	0.79	0.81	
Student Level							
r	0.26	0.26	0.26	0.26	0.26	0.26	

ANASTASIA PANAYIOTOU

Using Multilevel Analysis to Search for Curvilinear Relations of the Teacher and School Factors with Student Achievement

After the multilevel analyses presented earlier in this chapter that revealed the teacher and school factors that have an effect on student achievement in mathematics and science (see Tables 4.9 and 4.10) an attempt was made to search for curvilinear relations of the teacher and school factors with student achievement in mathematics and science. Thus, separate multilevel analyses were conducted using the quadratic teacher and school factor scores (i.e., the square values of the factors) instead of the non-quadratic. The results of this analysis showed curvilinear relations of some teacher factors with student achievement in mathematics and science while no curvilinear relation was found regarding the school level factors with student achievement in either outcome.

Specifically, the teacher factors that were found to have a curvilinear relation with student achievement in mathematics are concerned with: a) application and b) teacher ability in dealing with misbehavior, and the factors that were found to have a curvilinear relation with student achievement in science are: a) application, b) teacher-student interactions and c) student assessment. Tables 4.21 and 4.22 present the models for which curvilinear relations were identified in mathematics and science, respectively (see Models 2b and 3b for mathematics and 2b, 3b and 4b for science). For these factors, the models with the non-quadratic factor scores are also presented (see Models 2a and 3a for mathematics and 2a, 3a and 4a for science).

In addition, for the factors where curvilinear relations were identified, an alternative model was also examined where each non-quadratic factor score and its equivalent quadratic factor score were added together in Model 1 (see Models 2c, 3c and 4c in Tables 4.21. and 4.22). The fitting of these models was tested against Models 2a, 3a and 4a and Models 2b, 3b and 4b. Based on the likelihood statistic and the reduction observed, one may conclude that these models which have an extra degree of freedom, have a better fit to

ANASTASIA PANAYIOTOU

the data than both of the previous models where the non-quadratic factor score was not included in the model. For example, comparing Model 2c with Models 2a and 2b a reduction of 9 and 7 points respectively, is observed. In all cases an inverted u curvilinear relation was identified which is in line with the assumptions of the dynamic model and was one of the assumptions that was not tested earlier.

The results of the analyses searching for non-linear relations of the school factors with student outcomes are not presented since no curvilinear relation was identified. However, not identifying curvilinear relations for school factors could be attributed to the fact that the variance of the school factors was relatively smaller than the teacher factors. Implications of findings for theory and practice are discussed in the next chapter.

ANASTASIA PANAYIOTOU

Table 4.21. *Parameter Estimates and (Standard Errors) for the analysis of student achievement in mathematics – Curvilinear relations (Students within classes)*

Teacher Factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 3a	Model 3b	Model 3c
Fixed part (intercept)	330.3(1.55)	39.2(8.28)	12.2(11.5)	23.8(9.1)	154.2(47.3)	31.8(8.4)	41.8(8.2)	79.7(18.8)
Student Level								
Context								
Prior achievement		0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)	0.7(0.01)
Class Level								
Context								
Prior achievement		0.30(0.03)	0.30(0.03)	0.30(0.03)	0.30(0.03)	0.25(0.03)	0.25(0.03)	0.25(0.03)
Application			11.8(3.7)		-115.8(41.3)			
Application (quadratic)				2.8(0.8)	27.8(8.9)			
Misbehavior						10.3(2.4)		-37.7(16.8)
Misbehavior (quadratic)							2.7(0.6)	11.5(4.0)
Variance components								
Class	24.4%	6.3%	5.6%	5.6%	5.5%	5.4%	5.4%	5.3%
Student	75.6%	47.1%	46.8%	46.7%	46.7%	46.8%	46.7%	46.7%
Explained		46.6%	47.6%	47.7%	47.8%	47.8%	47.9%	48.0%
Significance test								
Log-likelihood	103422	98606	96961	96959	96952	96781	96777	96772
Reduction		4816	1645	1647	1654	1825	1829	1834
Degrees of freedom		2	1	1	2	1	1	2
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001

Note: For each of the Models 2a up to 3c the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.22. *Parameter Estimates and (Standard Errors) for the analysis of student achievement in science – Curvilinear relations (Students within classes)*

Teacher Factors	Model 0	Model 1	Model 2a	Model 2b	Model 2c	Model 3a	Model 3b	Model 3c
Fixed part (intercept)	317.3(1.6)	49.5(8.59)	19.1(12.2)	34.6(9.4)	128.4(52.8)	2.3(12.8)	30.3(9.2)	197.5(82.3)
Student Level								
Context								
Prior achievement		0.54(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)
Class Level								
Context								
Prior achievement		0.36(0.03)	0.35(0.03)	0.35(0.03)	0.35(0.03)	0.32(0.03)	0.32(0.03)	0.32(0.03)
Application			13.7(4.1)		-82.2(40.6)			
Application (quadratic)				3.1(0.9)	20.9(9.9)			
T-S Interactions						21.9(4.6)		-131.1(64.0)
T-S Interactions (quadratic)							4.4(0.9)	29.9(12.5)
Variance components								
Class	31.4%	9.8%	9.0%	9.0%	8.9%	8.7%	8.7%	8.6%
Student	68.6%	48.4%	48.0%	47.9%	47.9%	48.0%	48.0%	48.0%
Explained		41.8%	43.0%	43.1%	43.2%	43.3%	43.3%	43.4%
Significance test								
Log-likelihood	99531	95955	94341	94340	94337	94330	94329	94325
Reduction		3576	1614	1615	1618	1625	1626	1630
Degrees of freedom		2	1	1	2	1	1	2
p value		0.001	0.001	0.001	0.001	0.001	0.001	0.001

Note: For each of the Models 2a up to 3c the reduction is estimated in relation to the deviance of Model 1.

ANASTASIA PANAYIOTOU

Table 4.22.

Parameter Estimates and (Standard Errors) for the analysis of student achievement in science – Curvilinear relations (Students within classes)

Teacher Factors	Model 0	Model 1	Model 4a	Model 4b	Model 4c
Fixed part (intercept)	317.3(1.6)	49.5(8.59)	20.6(12.2)	35.4(9.4)	150.9(53.2)
Student Level					
Context					
Prior achievement		0.54(0.01)	0.55(0.01)	0.55(0.01)	0.55(0.01)
Class Level					
Context					
Prior achievement		0.36(0.03)	0.35(0.03)	0.35(0.03)	0.35(0.03)
Assessment			11.8(3.6)		-91.2(41.3)
Assessment (quadratic)				2.4(0.7)	20.2(8.1)
Variance components					
Class	31.4%	9.8%	9.0%	9.0%	9.0%
Student	68.6%	48.4%	48.0%	48.0%	48.0%
Explained		41.8%	43.0%	43.0%	43.0%
Significance test					
Log-likelihood	99531	95955	94342	94341	94336
Reduction		3576	1613	1614	1619
Degrees of freedom		2	1	1	2
p value		0.001	0.001	0.001	0.001

Note: For each of the Models 4a up to 4c the reduction is estimated in relation to the deviance of Model 1.

This Chapter presented the analysis of the data collected in order to provide answers to the research questions set. The first section examined the construct validity of the student and teacher questionnaire as well as the assumption of the dynamic model that the teacher and school factors, respectively, are related to each other. Testing the construct validity of the instruments aimed not only at providing support to the quality of the data collected throughout this study, but also to assist the development of internationally valid

ANASTASIA PANAYIOTOU

instruments that may be used in future international studies aiming to collect data on the teacher and school factors of the dynamic model. Thus, the results from the SEM analyses presented in the first section of Chapter 4 may support the claim that even though this study was not in a position to provide data on all five measurement dimensions of the factors included at the classroom and school level of the dynamic model, it was in a position to collect valid and reliable data on each factor.

The second section illustrated the results of multilevel modeling that was performed, both across and within countries, to explore the impact of the teacher and school factors on student achievement gains in mathematics and science, providing support to the assumption regarding the generic nature of the teacher and school factors. In the third section the results of multivariate multilevel analyses were able to support the results of the multilevel analysis that was conducted for each outcome separately and to demonstrate that teachers and schools that are effective in promoting learning in one subject are equally effective in the other. Finally, in the fourth section the results of the analyses searching for non-linear relations of the effectiveness factors included in the dynamic model with student achievement gains in mathematics and science were presented. In the next Chapter the results presented in Chapter 4 are further discussed and elaborated on.

ANASTASIA PANAYIOTOU

CHAPTER 5

DISCUSSION AND SUGGESTIONS FOR FURTHER RESEARCH

This chapter is based on the findings of this study and draws conclusions in relation to the research questions set, aiming to provide a better insight to the practical and theoretical contribution of the study. First, findings in relation to the measurement of the teacher and school factors of the dynamic model are discussed. Next, findings on the generic and differential nature of the factors are elaborated on and implications for policy development are drawn. Finally, suggestions for further research are made.

Introduction

One of the main questions that concerned researchers in the field of education and gained attention early on, in the field of educational effectiveness was the one concerned with the factors that contribute to student learning. Debates on the improvement of learning outcomes and on the maximization of the contribution of schooling, have led to a wide range of studies (e.g., Campbell, Kyriakides, Muijs, & Robinson, 2003; Opdenakker & Van Damme, 2000; Rutter et al., 1979) and several meta-analyses (see Hyde, Fennema & Lamon, 1990; Kyriakides, Creemers, Antoniou & Demetriou, 2010; Kyriakides & Christoforou, 2011; Scheerens, Witziers & Steen, 2013; Seidel & Shavelson, 2007), searching for a better understanding of what comprises educational effectiveness. In this spirit, different instructional approaches have been examined, such as the direct and active teaching approach (Fitzpatrick, 1982; Good & Grouws, 1979; Rosenshine, 1983) as to their contribution to student learning, and both domain specific and generic factors have been taken into consideration.

National studies were able to shed some light to the factors that affect variation in student achievement gains both in terms of cognitive as well as non-cognitive outcomes

ANASTASIA PANAYIOTOU

(i.e., affective, psychomotor and meta-cognitive). However, the question of whether results of national studies can be considered generic and thus equally effective in different educational contexts and subjects still remains under investigation and requires research attention. Research evidence suggests that the mere transfer of factors across countries may lead to spending valuable educational resources on factors that were found to function in one context but may not have an effect on student achievement gains in another (Mullis et al., 2000; Reynolds, 2000; 2006).

Identifying generic factors that may provide a basis for policy development in different countries, as well as research instruments that are able to provide valid information on these factors is imperative to ensure advancements (Teddlie et al., 2006), both in the field of EER as well as in the field of teacher professional development facilitating the ultimate goal of education; namely student learning (Creemers & Kyriakides, 2008). International studies may add to this attempt by assisting the investigation of factors that promote student learning outcomes regardless of the country's context (Kyriakides, 2006a).

Thus, this study collected data from six European countries aiming to add to current discussions in the field of EER regarding the factors that may cross national borders. This study also took into consideration the criticism placed upon research in education in regard to the infrequent use of existing theoretical models which may provide a more holistic view of the functioning of education (Bosker & Scheerens, 1994; Creemers, 2002b; Scheerens, 1993; Scheerens, 2014). Instead of sporadically selecting factors that could explain variation in student outcomes, this study is based on a specific theoretical framework (i.e., the dynamic model of educational effectiveness) which provides the possibility for establishing an evidence-based and theory-driven approach for policy development that can eventually lead to improvement in educational outcomes in different contexts.

ANASTASIA PANAYIOTOU

The dynamic model examines factors operating at the different levels of education and despite its complexity, its practical use for improvement purposes has been proven both at classroom and school level (see Antoniou, 2013; Antoniou & Kyriakides, 2011, 2013; Antoniou, Kyriakides & Creemers, 2011; Creemers & Kyriakides, 2010b; Kyriakides, Archambault & Janosz, 2013). Hence, by providing further support to its main assumptions in regard to the generic nature of the school and classroom factors at international level, this study may be in a position to contribute to the field of EER and to the field of teacher professional development, claiming that an integrated approach should be followed in professional development efforts.

Summarizing the main results of this study conclusions are drawn in regard to the measurement of the teacher and school factors of the dynamic model, the generic and differential effects of these factors on student achievement in six European countries and two different subjects and the practical contribution of the study.

Measurement of the Teacher and School Factors

One of the main lessons learned from the two early attempts in conducting international research, namely ISERP (Creemers, Reynolds & Swint, 1994; Reynolds, Creemers, Stringfield, & Teddlie, 1998) and ISTOF (Teddlie et al., 2006) was that the successful measurement of effectiveness factors primarily lies on the establishment of instruments that are capable to adapt to the discrepancies of different contexts and provide researchers with accurate measurements of the constructs under investigation. Classroom practices face diversity across countries and depend on the specificities of each context since the realization of certain concepts (such as differentiation, misbehaviour etc.) may differ (Ding, Li, Li & Kulm, 2008; Teddlie, Creemers, Kyriakides, Muijs & Yu, 2006). Thus, internationally valid instruments measuring classroom factors may provide a valuable tool in international teacher effectiveness research.

ANASTASIA PANAYIOTOU

In order to measure the classroom level factors of the dynamic model, questionnaires were administered to all grade 4 students of the sample to collect information on their teachers' behaviour in the classroom in relation to the eight factors included in the dynamic model. Students are considered as a valuable source of information since they are immediate recipients of teaching practices throughout the year and are present in different manifestations of teacher behaviour inside the classroom (Creemers, Kyriakides, & Sammons, 2010; Den Brok, 2001; Fraser, 1998; Lasagabaster & Sierra, 2011; Shuell, 1996; Wubbels & Brekelmans, 1998). Student ratings are commonly used in tertiary education to evaluate several teaching aspects of their instructors (Centra, 1993; Feldman, 1997; Marsh, 1987, 2007; Marsh & Roche, 1994; Watkins, 1994). However in the lower levels of education – and especially in primary education – researchers seem hesitant to use student ratings as a source of information due to doubts on the validity and reliability of such ratings. In particular, as Becker (2000) states "...student evaluations of teaching may be widely used simply because they are inexpensive to administer, especially when done by a student in class, with paid staff involved only in the processing of the results..." (p.114).

The quality of student ratings has thus been a matter of discussion and investigation among researchers (Abrami & d'Apollonia, 1991; Abrami, d'Apollonia, & Rosenfield, 1997; Marsh & Roche, 1997). It should however be acknowledged that research results, were in a position to contradict relevant disapprovals and demonstrate stability in student ratings (e.g., Carle, 2009; Marsh, 2007). Other studies were also in a position to demonstrate the validity and reliability of student ratings not only from students coming from secondary and upper secondary education but also from students at the end of primary education (Driscoll, Peterson, Crow, & Larson, 1985; Peterson & Stevens, 1988; Scriven, 1994; Taba, Tylor, & Smith, 1998).

ANASTASIA PANAYIOTOU

This study provided support to the assumption that grade 4 students are able to provide valid information on their teacher's in-class behaviour in relation to the eight factors of the dynamic model. It should nonetheless be noted that the questionnaire items measuring teacher factors referred to observable actions and not to inferences about teacher behaviour. For example, students were asked to specify whether the teacher uses words that are hard to understand when he/she asks us a question or whether the teacher encourages them to ask questions if there is something that they do not understand during the lesson. The specificity of the questionnaire items provided students with not only a better understanding of the meaning of the items, but the items were also stated in such way that students did not have to assess their teacher in terms of whether he/she is a "good" or "bad" teacher causing social desirability problems. Likewise, students were not asked to assess teacher content knowledge or personality traits, since for that they would have required special knowledge and evaluation skills.

Moreover, taking into consideration that students were asked to report on their teacher, the generalizability of the data was examined so as to demonstrate whether there was consensus in student responses in each class (within-group variance) and to identify whether the object of measurement was the teacher. When collecting data from lower level units regarding units of an upper level it is essential to examine whether the data are generalizable. The results of the generalizability study concerning students' ratings, showed that the data can be generalized at the classroom level, as for all the items of the questionnaire, the between-group variance was higher than the within-group variance ($p<0.05$). Taking the above into consideration it can be claimed that data from this study provided support to the validity of the questionnaire used to measure the classroom level factors of the dynamic model.

However, apart from the theoretical implications drawn, establishing a valid instrument for the measurement of factors operating at classroom level that were shown to

ANASTASIA PANAYIOTOU

have an effect on student achievement, may also have practical implications by providing a basis for policy development. Particularly, teachers and other school stakeholders may use this questionnaire to collect data about quality of teacher behavior in classrooms and develop school improvement projects to address factors found to be associated with student learning outcomes. It should nonetheless be acknowledged that combining the results from student questionnaires with other sources of information (i.e., classroom observations) may provide a better insight as to the functioning of the teacher factors in regard to the five measurement dimensions and support the reliability and validity of student ratings. The contribution of collecting observational data has been recognized by large organizations such as OECD (2015b) and has also been demonstrated by studies outside Europe. For example, a recent study that took place in Ghana collected data from a sample of 77 primary schools regarding quality of teaching through student questionnaires and classroom observations using three observation instruments (i.e., a high and two low inference instruments). Even though confirmatory factor analyses provided support to the construct validity of each instrument, only the low inference observation instruments were able to provide information both on the classroom level factors as well as on all their five measurement dimensions. In addition, only the data collected from the observation instruments were able to show the effect of the teacher factors on student achievement (Azigwe & Kyriakides, 2016).

Additionally, the results of the SEM analyses showed that the eight factors of the dynamic model can be classified into two overarching factors; quantity and quality of teaching. Factors belonging to quantity of teaching are concerned with the teacher's ability to maximize the use of available teaching time and factors belonging to quality of teaching refer to the use of teaching time in an effective way. One of the main findings of this study, that has significant policy implications in regard to teacher professional development efforts, is that the correlation between the two overarching factors was found to be very

ANASTASIA PANAYIOTOU

small. This suggests that teachers who are able to maximize the use of teaching time may not necessarily be able to use the time effectively and thus effectiveness in one overarching factor does not also inevitably imply effectiveness in the other. For instance, teachers who may be efficient in minimizing interruptions during the lesson by effectively dealing with misbehavior or able to manage teaching time so as to finish lessons on time, may not be able to effectively structure the lesson or provide students with appropriate application tasks based on their needs and abilities. Thus, teacher professional development programs should equally address factors belonging to both aspects of teaching (i.e., quality and quantity) to more holistically help teachers improve their teaching skills (Antoniou & Kyriakides, 2011; Creemers, Kyriakides, & Antoniou, 2013).

Research in the field of teacher professional development also indicates that in order for teaching skills to be improved, professional development efforts should be differentiated to meet the specific development needs of each teacher (Antoniou & Kyriakides, 2011; Christoforidou, Kyriakides, Antoniou & Creemers, 2014; Creemers & Kyriakides, 2013; Kyriakides, Archambault & Janosz, 2013). Consequently, by using the student questionnaire, teacher educators may collect valid data on teacher skills in six different countries and identify areas for improvement based on which individual action plans can be formed and training courses based on specific areas can be offered.

Another implication also arises for teacher professional development based on the fact that the results of this study support the assumption of the dynamic model that the teacher factors are inter-related and should not be regarded as separate, isolated concepts when developing teacher professional development courses (Creemers & Kyriakides, 2008). One of the most dominant approaches in teacher professional development that is widely used in all of the participating countries of this study, is the Competency Based Approach (CBA) (Brooks, 2002; Last & Chown, 1996; Robson, 1998; Whitty & Willmott, 1991). The CBA is based on the acquisition of discrete – isolated teaching skills (Gonczi,

ANASTASIA PANAYIOTOU

1994) based on lists of skills which are developed by experts (Sprinthall, Reiman, & Thies-Sprinthall, 1996) and no grouping of factors is taken into perspective. However, based on the results of this study as well as the results of previous national studies (see Creemers, Kyriakides & Antoniou, 2013), rather than using the CBA, pre-service and in-service teacher education programs should adopt a more holistic training approach, based on a solid theoretical framework to assist teachers by reflecting their practical needs in the classroom, both in terms of quantity and quality of teaching.

Despite of providing evidence for the validity of student ratings in regard to the eight classroom level factors of the dynamic model, obtained through a student questionnaire, this study was also in a position to provide evidence regarding the validity of the teacher questionnaire designed to collect data on: a) School policy on teaching, b) Policy on the School Learning Environment and c) Policy on Evaluation. Similarly to the student questionnaire analysis, a Generalisability Study on the use of teachers' ratings was conducted and showed that the data obtained through the teacher questionnaires can be generalized at the school level showing that the teacher questionnaire is in a position to collect data from teachers on school level factors.

Providing support to the validation of the teacher questionnaire through this study again generates certain policy implications as to the practical use of this instrument for improvement purposes. Namely, policy makers and school stakeholders may use this instrument to implement school self-evaluation projects. By identifying specific areas for improvement (e.g., relations with the community, policy on the quality of teaching etc.), schools may follow a theory driven and evidence-based approach to evolve aspects of their policy that function less effectively than others. This will allow them to set priorities for improvement and develop relevant action plans. The questionnaire may then be again used to identify progress and redefine the action plans.

ANASTASIA PANAYIOTOU

The first section of this Chapter discussed the practical and theoretical implications occurred from the international validation of the two instruments measuring the teacher and school factors of the dynamic model. The following sections elaborate on the results of the multilevel analyses in regard to the generic nature of the teacher and school factors. The assumptions in regard to the nature of relationship of some factors with student outcomes; namely curvilinear relationship (Campbell et al., 2004; Kyriakides, 2007; Scheerens & Bosker, 1997) are also discussed based on the results of this study and policy implications are drawn.

The Generic and Differential Nature of the Teacher and School Factors

While several national studies are conducted every year in a large number of countries, research attempts in the past few decades highlight the growing interest in the field of EER for the identification of factors that may affect student outcomes irrespective of the context or subject (Peaker, 1975). Factors such as the family background (Buchmann, 2002), topics such as literature education (Purves, 1973) and issues such as inequality (Van de Werfhorst & Mijs, 2010) and quality in education (Heyneman & Loxley, 1983) are some of the constructs that have started to be seen as a matter of international examination.

Namely, two major associations: a) the International Association for the Evaluation of Educational Achievement (IEA) (Postlethwaite 1995; Postlethwaite & Ross, 1992) and b) the Organization for Economic Co-Operation and Development (OECD) have been developed in an attempt to expand nationally restricted knowledge and provide a basis for comparison among counties (Drent, Meelissen & van der Kleij, 2013). The international comparative studies of TIMSS and PIRLS initiated by IEA and the international study of PISA developed by OECD depict the developing interest of providing an equal basis for measuring student outcomes. However, the methodological nature of these international

ANASTASIA PANAYIOTOU

studies (i.e., their cross-sectional character) does not efficiently allow for their results to be used in terms of policy development for improving student outcomes in the participating countries. Specifically, each time these large-scale international studies take place they only collect data on student achievement at one time-point. This does not allow for comparisons to be made among the same students across-time, thus not providing the opportunity to identify progress in student outcomes and search for factors affecting this progress based on which policy actions can be undertaken.

The international study presented in this thesis took in mind this methodological limitation and examined student achievement gains in mathematics and science by collecting data at two time points (i.e., student initial achievement at the beginning of the school year and student final achievement at the end of the school year). As a result, conclusions can be drawn regarding the effects of the teacher and school factors of the dynamic model on student achievement gains and identify more complex relations (i.e., non-linear relations). This inevitably, has key policy implications for the participating countries since by identifying generic factors being relevant for policy making in each country, policy makers may use these results to implement professional development courses at both the teacher and school level.

As mentioned in Chapter 2, one of the main concerns arising in regard to the use of the results occurring from national studies searching for factors that have an effect on student achievement gains, refers to the potential implementation of reforms in certain educational contexts based on the results of national studies that took place in other countries (Reynolds, 2006; Scheerens, 2013). Research has shown that factors found to be associated with student achievement gains in some countries, such as the principal's leadership characteristics (e.g., Hallinger & Heck, 2011; Louis, Dretzke, & Wahlstrom, 2010) did not have an effect in others. Hence, it would be a fallacy to suggest that national research results can provide a basis for policy development in other countries than the one

ANASTASIA PANAYIOTOU

of the study and produce equally effective results. Across-country studies may provide support to the results of national studies suggesting whether their findings comprise a specialized or generic occurrence (Reynolds et al., 2014).

This study was in a position to provide support to the generic nature of the teacher and school factors of the dynamic model in six different countries since multilevel analyses revealed statistically significant effects of almost all teacher and school factors on student achievement gains in the two subjects. So far, national studies attempted to provide empirical support to the assumption as to the generic nature of the factors included in the dynamic model concerning different outcomes of schooling (i.e., cognitive and affective) (see Creemers & Kyriakides, 2010a; Kyriakides & Creemers, 2008, 2009), different subjects (i.e., language and mathematics) and different levels of education (i.e., primary and pre-primary), where differential effects were identified (Kyriakides & Creemers, 2009). This international longitudinal study adds to previous national efforts by providing further evidence not only on the effects of the factors in different subjects, but also regarding the generic nature of the effectiveness factors across different educational contexts.

Looking at the effect sizes showing the practical significance of the teacher and school factors on student achievement in both subjects which were calculated using the results of the within-country multilevel analyses, it can be observed that neither of the factors was systematically found to not have an effect in some countries. Additionally, only small differences in the effect sizes of the factors were found in mathematics and science achievement and the slight differences that were observed in regard to science can be explained by the fact that the variance at both the classroom and school level was more evident in this subject. These findings suggest that the teacher and school level factors respectively, are equally important for student achievement in mathematics and science in the six participating countries and can be interpreted in terms of suggestions for policy

ANASTASIA PANAYIOTOU

development in each of the participating countries. Precisely, even though the minor differences in the effect of some factors on student achievement (not greater than 0.08 for the school factors and not greater than 0.12 for the teacher factors), is not considered sufficient to suggest that differential effects exist, policy makers may use these results to identify priorities for improvement for factors that are found to have an overall greater effect. For example, based on these results, suggestions could be made to schools in order to develop a policy on teaching as well as a policy on establishing a learning environment in both subjects.

In addition, the results of the multilevel analyses showed that factors belonging to different teaching approaches have an effect on student achievement gains in different learning outcomes. Namely, both, factors associated with the active and direct teaching approach (e.g., structuring, application) and factors that refer to the constructivist approach to learning (e.g., modeling) were found to affect student outcomes. This may lead to the conclusion that this study generates support for using an integrated approach in defining quality of teaching and designing teacher training courses, which is also supported by the results of previous meta-analyses (see Seidel & Shavelson, 2007).

The results of multivariate analyses which were conducted to examine whether the teacher and schools factors had differential effects on student achievement in the two subjects, did not indicate differential effects in the functioning of the factors. The results however, provided further support to the generic nature of the factors since they were found to be equally important for both subjects. It should though be acknowledged that the effect of the factors was examined in two very closely related cognitive subjects and in order to be in a position to better examine the differential effects of the factors one should examine their impact on cognitive and non-cognitive outcomes (Gruehn, 1995; Knuver & Brandsma, 1993; Stankov, Morony & Lee, 2014; Stankov & Lee, 2014) as well as on meta-cognition (Boström & Lassen, 2006; Kuyper, Van der Werf & Lubbers, 2000;

ANASTASIA PANAYIOTOU

Mevarech & Fridkin, 2006). A recent study is now conducted in Cyprus testing the effect of the classroom level factors of the dynamic model on students' metacognitive skills and thus, such national studies should provide a basis for expanding research on alternative outcomes (i.e., other than cognitive) (Kyriakides & Anthimou, 2016).

The correlations between the residuals for mathematics and science at the classroom level were also found to be high (>.78), implying that teacher effectiveness is related in the two subjects in terms of promoting student achievement gains. It can therefore be argued that teacher professional development courses may address generic teaching skills which may assist teachers develop professionally and improve their overall teaching behavior. Similarly, it was shown that schools that are effective in terms of one subject are also effective in the other. Again, a better insight can be provided by comparing the effectiveness of teachers and schools in more diverse subjects such as mathematics and physical education, or such as cognition and meta-cognition. It should also be taken into consideration that the sample of the study comprised of primary school teachers which are considered as generic since they teach both subjects. Thus, equivalent questions could be raised taking a sample of secondary education teachers, regarding the correlation of teachers' effectiveness in different classroom contexts (see Kokkinou & Kyriakides, 2016).

Apart from the assumption on the generic nature of the teacher and school factors of the dynamic model, another hypothesis is made which was also put to the test through this study. Specifically, it is assumed that the relationship of some effectiveness factors with student outcomes may not be linear but curvilinear (Creemers & Kyriakides, 2006, 2008). While national studies may not be in a position to provide enough variance so as to determine such complex relations, longitudinal international studies that are conducted using an international sample could provide enough variance and statistical power allowing for the demonstration of more complex relations (Creemers et al., 1998). Thus, moving a step forward and based on its international sample, this study attempted to identify

ANASTASIA PANAYIOTOU

curvilinear relations of the teacher and school factors of the dynamic model with student achievement gains in mathematics and science.

Multilevel analyses indicated quadratic relationships between some teacher level factors with student achievement in mathematics and science, while it was not possible to identify non-linear relations of the school factors with student achievement in either subject. In regard to the teacher factors, application was found to have a curvilinear relation with achievement in both subjects. Application tasks are considered an essential part of teaching since they help students not only to better comprehend the new knowledge, but also to link new knowledge with previous, already acquired one (Sweller, 1994). In order for students to be able to transfer new knowledge from the short-term memory to the long-term, and relieve working memory from an extensive load, application opportunities are required so as to ultimately improve learning outcomes (Borich, 1992). Application activities may also help students code and more easily retrieve information so as to use it in other contexts (Anderson, 1983; Travers, 1982). Based on the Cognitive Load Theory, students need to frequently apply new information since the working memory can only process a limited amount of information at each given time (Chandler & Sweller, 1991; Kirschner, 2002; Plass, Moreno & Brünken, 2010; Sweller, 2011).

Previous studies and meta-analyses were able to demonstrate the significance of application opportunities for promoting learning by identifying the effect of this factor on student achievement (e.g., Klein & Pridemore, 1994; Kyriakides, Christoforou & Charalambous, 2013). The results of this study are in line with earlier results, however identifying a curvilinear relation with achievement and student achievement gains implies that application may reach an optimal point where it can have no further effect on student achievement. This can be explained by the fact that extensive provision of application activities may act at the expense of teaching new learning content. In case teachers spend most of the teaching time on application of previous knowledge, it is assumed that students

ANASTASIA PANAYIOTOU

are not given the opportunity to progress in terms of acquiring sufficient amount of new knowledge. Thus, suggestions can be made for evenly distributing application activities among lessons and for having a balance between teaching of new content and application tasks.

Similarly, based on the results of the multilevel analyses it can be claimed that assessment is another effectiveness factor that may discontinue having an effect on student achievement when extensively used. In spite of the fact that assessment should be considered as an integral part of everyday teaching practice and teachers need to be able to effectively implement assessment techniques to identify student learning needs and promote further development in students' abilities, skills and knowledge (Gardner et al., 2010; Wylie & Lyon, 2009) too much time spent on assessing already obtained knowledge may deprive students of the time to learn new lesson content. Suggestions can therefore be made for policy makers so as to implement professional development courses focusing on developing teacher skills in using assessment for formative reasons (Black & Wiliam, 1998, 2005; Scriven, 1967) using different assessment methods (e.g., observation) that can be easily integrated into teaching practice.

Teacher actions in dealing with student misbehavior and classroom disorder were also found to reach an optimal point. This suggests that even though teachers are expected to effectively deal with student misbehavior in order to promote learning (Kulinna, 2007), too much time spent on dealing with such behaviors may negatively affect on task teaching time and eventually, student achievement gains. Research has shown that teachers spent a substantial amount of time in order to solve problematic in-class behaviors even when misbehavior appears in its simplest forms (e.g., students talking with each other, not paying attention to the lesson, interrupting others etc.) (Houghton, Wheldall, & Merrett, 1988; Little, 2005). In case teachers spend large amounts of teaching time dealing with student misbehavior, this may imply that they do not obtain the necessary skills allowing

ANASTASIA PANAYIOTOU

them to effectively manage to solve disorder problems as soon as they appear, preventing future misbehavior. This may therefore indicate the need for teachers to participate in relevant training courses dealing with ways of effectively addressing student misbehavior and classroom disorder (Merrett & Wheldall, 1993). Teacher's ability to create and maintain an orderly classroom environment was one of the factors that gained interest early on in the field of EER and was found to have significant effects on student outcomes (Creemers, 1994; Creemers & Reezigt, 1996; Doyle, 1984; Kyriakides, Campbell & Christofidou, 2002). The effects of the classroom learning environment in general were not only indentified in terms of cognitive outcomes but also as mentioned by Baker (1999) in terms of affective outcomes, such as motivation, self-concept, and academic engagement. Thus, policy implications can be drawn for assisting teachers develop an appropriate learning environment by minimizing student inappropriate behavior.

The classroom learning environment, as defined by Creemers and Kyriakides (2008), refers not only to teacher's ability to deal with classroom disorder but also to teacher-student interactions, student-student interactions, student's treatment by the teacher and competition between students. These five elements are considered as integral aspects of this factor and effective teachers are expected to take them all into consideration. In specific regard to the teacher – student and student – student interactions the results of previous studies showed that teachers should promote both types of interactions since they are important parts of establishing a classroom environment that promotes student learning (Cazden, 1986; den Brok, Brekelmans, & Wubels, 2004; Fraser, 1991). This study was yet able to show that in the case of teacher-student interactions an optimal point may be reached. This may possibly be explained by taking into consideration that teaching time is - at a certain degree - restricted and thus when teacher – student interactions comprise a large part of that time, less time remains for other types of interactions to take place, such as interactions between students themselves, which were shown to promote student

ANASTASIA PANAYIOTOU

learning. Namely, students' co-operative work in settings where interpersonal interactions may be developed could be in a position to facilitate students' problem solving skills and raise students' activeness, resulting to an overall increase of knowledge acquisition (Leikin & Zaslavsky, 1997). As a result, suggestions can be made for teachers to maintain a balance between the different interactions that may take place in a lesson and promote both teacher - student, as well as student-student interactions.

Concluding, the findings of this study, as these are discussed above, provide important information in relation to the measurement of the teacher and school factors of the dynamic model at international level, as well as the assumptions made on the nature of the factors (i.e., generic) and their relation with student achievement (i.e., linear and curvilinear). Given the fact that this study aimed at not only contributing to theory and demonstrating the contribution of international studies, but is based on the notion that the results of studies in the field of EER should contribute in improving practice (Creemers & Kyriakides, 2008; Reynolds, Hopkins & Stoll, 1993), implications of the findings for policy and practice especially in the context of Cyprus are presented next.

Implications for Policy and Practice in the Context of Cyprus

The main purpose of this study was the further development and testing of the validity of the dynamic model of educational effectiveness at the classroom and school level, in order to test the extent to which the dynamic model can be used as a starting point for establishing the theoretical framework of EER. Nevertheless, the contribution of providing further support to the dynamic model is not restricted to supporting the theoretical foundations of effectiveness studies but also to the usage of these results to promote an evidence-based and theory driven approach for practical improvements in education. In Cyprus, efforts have been made in the past few years to improve students' cognitive and affective outcomes, setting this as a number one goal of the educational

ANASTASIA PANAYIOTOU

system, especially during the school year 2014-2015. Teachers' education and professional development has constituted a priority for the Cypriot educational system since 2011 in order to effectively use new educational material and implement effective teaching practices (Ministry of Education and Culture, Republic of Cyprus, 2014a).

However, despite the Ministry's efforts to implement teacher professional development courses, the training offered still seems to follow a rather mechanistic approach aiming to develop isolated, unrelated skills (e.g., professional development courses on the development of skills in relation to the use of new technologies) (Ministry of Education and Culture, Republic of Cyprus, 2014b). This study provides evidence supporting the use of an integrated approach for teacher professional development. Given the fact that correlations were identified among the eight teacher factors of the dynamic model (see results of the SEM analyses) and that most of these factors were found to affect student outcomes, suggestions can be made for implementing an integrated approach to teacher professional development efforts. These results support the research findings of previous studies that aimed to compare the widely used Competency-Based Approach with a Dynamic Approach that is based on the assumption that teacher skills are inter-related and should be grouped for teacher development purposes (Creemers, Kyriakides & Antoniou, 2013).

Particularly, previous studies were able to support the assumption that the focus on groupings of factors can be proven more beneficial in terms of teacher skill development and also on student achievement, rather than the development of isolated skills (Kyriakides, Archambault, & Janosz, 2013, Kyriakides, Creemers, & Antoniou, 2009). Hence, this study adds to discussions on the form of teacher training courses, providing support to the viewing of teacher development as an integrated whole. The results of this study may be used to persuade policy makers in Cyprus that the factors included at classroom and school level of the dynamic model are important for promoting student

ANASTASIA PANAYIOTOU

outcomes not only in the Cypriot context but also in other European countries and should thus be considered when designing reform policies.

In addition, based on the view that evaluation – and much more – self-evaluation may constitute an unambiguous process of personal development (MacBeath, 2005), this study's contribution also lies on the validation of instruments that can be used for school and teacher self- evaluation purposes in all participating countries of this study. Since data collection is seen as a fundamental part of any evaluation process (Devos, 1998; Kyriakides & Campbell, 2004), schools in Cyprus may use the teacher questionnaires to identify priorities for improvement and with the support of an advisory team, develop an action plan for improving aspects of school policy that are found to underperform.

Similarly, teachers may use the student questionnaires to get a better view of their practices as they are perceived by their everyday recipients, namely students. A common phenomenon observed in Cyprus, which is also acknowledged by several researchers (e.g., MacBeath, 2005; McNamara & O'Hara, 2005; Vanhoof & Van Petegem, 2007), refers to the disagreement of external evaluation results with those of self-evaluation. Specifically, in Cyprus teachers tend to alter their teaching behavior when observed by an external observer (e.g., an inspector) trying to meet certain standards. However, these "ideal" lessons do not reflect everyday reality as experienced by students. Using the student questionnaires of this study, teachers may be involved in a self-evaluation process, setting priorities for improvement without the extensive concern of judgment by an external actor and the fear of accountability (Nevo, 2001).

Research Limitations and Suggestions for Further Research

The desire of educational systems worldwide for achievement excellence has led countries to conduct studies aiming to establish ways of promoting learning and educational advancements (Plucker, 2015). Identification of factors that promote student

ANASTASIA PANAYIOTOU

learning is assumed to have an impact on national policies, leading to implementation of actions for improvement in all levels of education (i.e., system, school, classroom and student). Since the role of teachers and schools was shown to be central in student learning several decades ago (Lockheed & Komenan, 1989), this international longitudinal study was based on a theoretical framework (i.e., the dynamic model of educational effectiveness) and focused on providing information on factors operating at the classroom/teacher and school level that may have an impact on student achievement gains in two cognitive subjects (i.e., mathematics and science).

Given the fact that international studies may not only contribute to the identification of differences among student outcomes in different countries through the provision of a common basis for comparison (Beaton et al., 1996; Mullis, Martin, Gonzalez, & Chrostowski, 2004), but also to the demonstration of factors that are associated with student achievement progress, this study collected longitudinal data and provided evidence-based information on ways to increase student gains.

However, it should be acknowledged that this study only collected data from countries within Europe which do not significantly vary in terms of their context. Suggestions can therefore be made for expanding this study by gathering data from countries outside Europe (i.e., Asia and/or Africa) where contextual differences are more evident. Namely, researchers raise concerns that factors found to have an effect on student achievement in developed countries may not have the same effect in developing countries due to profound differences such as the provision of resources, teacher education and classroom size (Howie, 2005; Legotlo, Maaga & Sebego, 2002). Even though some factors, such as the provision of sufficient teaching time which may allow students to engage in learning activities, were found to relate with student achievement in both developed and developing countries (Avalos & Haddad, 1979; Fuller, 1987), further

ANASTASIA PANAYIOTOU

research may shed more light as to the functioning of all the classroom and school level factors included in the dynamic model in different contexts.

Moreover, it should be acknowledged that this study was able to provide evidence on the functioning of the teacher and school factors of the dynamic model in relation to student achievement gains, but not also on each of the five measurement dimensions (i.e., frequency, focus, stage, quality and differentiation). This occurred as a result of the adaptation of the student and teacher questionnaire, respectively to correspond to the context of each participating country. Thus, the complementary use of other means on data collection, such as well structured classroom observations conducted by well-trained observers (McCutcheon, 1981; Reynolds, 2006; Teddlie et al., 2006; Williams, 1989) can be suggested since they may not only provide the opportunity to collect data on the five measurement dimensions, gaining a more holistic view of the functioning of the factors, but also to better establish the validity and reliability of the data collected through the questionnaires (Milich & Fitzgerald, 1985). The international study presented in this thesis was not in a position to collect observational data due to financial constrains and the practical difficulties of conducting observations in a large-scale international study, and this limitation should be recognized.

Further international studies may also attempt to collect data on not only cognitive, but also on non-cognitive outcomes (Knuver & Brandsma, 1993; Stankov, Morony & Lee, 2014) and meta-cognition (Boström & Lassen, 2006; Kuyper, Van der Werf & Lubbers, 2000). This way the differential effects of the teacher and school factors may be more evident and more easily identified. However, irrespective of whether only student cognitive outcomes are measured, researchers should concentrate their efforts on the establishment of measurement instruments that correspond to the curriculum of each country. This study collected data on student achievement in mathematics and science using items from the TIMSS 2007 booklets. Yet, one of the most important lessons learned

ANASTASIA PANAYIOTOU

through this international study is that the instruments used to measure student outcomes should be adapted so as to sufficiently correspond to the curriculum of each country. By using a test that is not adapted to meet the differences in the curriculum for all countries, researchers face the risk of students underperforming in certain countries where what is actually taught in classrooms differs from the tasks included in the test. This way lack of enough variance may prevent the identification of the full effect of the factors on student achievement. On the contrary, by comparing the curriculum of each country and examining the actual materials taught in each country, standardized tests can be developed and used in international studies.

Finally, further research may contribute to gaining a better understanding of not only aspects relating to the quality of education but also to issues related to equity (Benadusi, 2001; Gorard & Smith, 2004). The provision of equal educational opportunities to students coming from different economic and social backgrounds has gained increased interest in the past decade since the question of whether schools can adjust for some students' unprivileged background remains under investigation (Lynch, 2001; Lynch & Baker, 2005; Kelly, 2012).

Through this study it was not possible to generate an international scale concerning student background characteristics and socio-economic status (SES). To collect data on SES a questionnaire was developed based on items from TIMSS as well as items that were based on the context of each country. However, due to the fact that different countries had several different items measuring aspects of the SES that were relevant to their context it was not possible to develop an internationally valid instrument measuring SES. Thus this study was not able to address issues related to equity and further research is needed to examine whether the teacher and school factors of the dynamic model have differential effects on students coming from different socio-economic backgrounds and may contribute not only in promoting quality but also equity in education. Finally, the possibility that

ANASTASIA PANAYIOTOU

contextual features such as student composition (in terms of prior achievement at the class level) may influence both outcomes and teaching behaviours could be examined (e.g through interaction effects).

ANASTASIA PANAYIOTOU

References

Abrami, E. C., & d’Apollonia, S. (1991). Multidimensional students’ evaluations of teaching effectiveness – Generalizability of ‘‘N = 1” research: Comment on Marsh (1991). *Journal of Educational Psychology, 30*, 221–227.

Abrami, E. C., d’Apollonia, S., & Rosenfield, S. (1997). The dimensionality of student ratings of instruction: What we know and what we do not. In R. E. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 321–367). New York: Agathon Press.

Acosta, S. T., & Hsu, H. Y. (2014). Negotiating diversity: an empirical investigation into family, school and student factors influencing New Zealand adolescents’ science literacy. *Educational Studies, 40*(1), 98-115.

Ainscow, M. (2010). Achieving excellence and equity: reflections on the development of practices in one local district over 10 years. *School Effectiveness and School Improvement*, *21*(1), 75-92.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261-295.

Anderson, W. L. (1984). *Time and School Learning*. London: Croom Helm.

Antoniou, P. (2013). A Longitudinal Study Investigating Relations Between Stages of Effective Teaching, Teaching Experience and Teacher Professional Development Approaches. *Journal of Classroom Interaction*, *48*(2), 25-40.

Antoniou, P., & Kyriakides, L. (2011). The impact of a dynamic approach to professional development on teacher instruction and student learning: results from an experimental study. *School Effectiveness and School Improvement, 22*(3), 291-311.

Antoniou, P., & Kyriakides, L. (2013). A Dynamic Integrated Approach to Teacher Professional Development: Impact and Sustainability of the Effects on Improving Teacher Behavior and Student Outcomes. *Teaching and Teacher Education, 29*, 1-12.

Antoniou, P., Kyriakides, L., & Creemers, B.P.M. (2011). Investigating the Effectiveness of a Dynamic Integrated Approach to Teacher Professional Development. *Center for Educational Policy Studies Journal*, *1*(1), 13-42.

Aparicio, J. J., & Moneo, M. R. (2005). Constructivism, the so-called semantic learning theories, and situated cognition versus the psychological learning theories. *Spanish Journal of Psychology, 8*(2), 180-198.

ANASTASIA PANAYIOTOU

Askew, M., & William, D. (1995). *Recent research in mathematics education* 5–16. London: Office for Standards in Education.

Avalos, B., & Haddad, W. (1979). *A review of teacher effectiveness research*. Ottawa: International Development Research Centre.

Azigwe, J. B. & Kyriakides, L. (2016). Using observation and student questionnaire data to measure the impact of teaching factors on Mathematics Achievement of primary students in Ghana. *Paper to be presented at the 29th International Congress for School Effectiveness and Improvement (ICSEI) 2016.* Glasgow, Scotland, January 6-9.

Baker, J. A. (1999). Teacher-student interaction in urban at-risk classrooms: Differential behavior, relationship quality, and student satisfaction with school. *The Elementary School Journal*, *100*(1), 57-70.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research, 85*(2), 89-99.

Basagaña, X., & Spiegelman, D. (2010). Power and sample size calculations for longitudinal studies comparing rates of change with a time varying exposure. *Statistics in medicine, 29*(2), 181-192.

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

Beaton, A. & Robitaille, D. (2002) A look back at TIMSS: what have we learned about international studies?. In D. Robitaille and A. Beaton (eds), *Secondary Analysis of TIMSS-Data* (Dordrecht: Kluwer), 409–417.

Becker, W. (2000). Teaching economics in the 21st century. *Journal of Economic Perspectives, 14*, 109–120.

Benadusi, L. (2001). Equity and education: a critical review of sociological research and thought, In W. Hutmacher, D. Cochrane & N. Bottani (Eds) *In pursuit of equity in education*. Dordrecht, Kluwer.

Bennetts, T. (2005). The links between understanding, progression and assessment in the secondary geography curriculum. *Geography,* (2), 152-170.

Bentler, P.M. 1990. Comparative fit indexes in structural models. *Psychological*

ANASTASIA PANAYIOTOU

Bulletin 107(2), 238–246.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Black, P., & Wiliam, D. (2005). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and Learning*, London, UK: Sage.

Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Bloom, B. S. (1976). *Human characteristics and school learning*. McGraw-Hill.

Borich, G.D. (1992) (2nd Ed). *Effective teaching methods*. New York: Macmillan Publishing Company.

Bosker, R. J., & Scheerens, J. (1994). Alternative models of school effectiveness put to test. *International Journal of Educational Research*, *21*(2), 159 ‐ 180.

Boström, L., & Lassen, L. M. (2006). Unraveling learning, learning styles, learning strategies and meta-cognition. *Education + Training*, *48*(2/3), 178-189.

Bowen, N. K., & Guo, S. (2011). *Structural equation modeling*. Oxford University Press.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, *8*(2), 153-169.

Brookhart, S. (2004). Classroom assessment: Tensions and intersections in theory and practice. *The Teachers College Record*, *106*(3), 429-458.

Brooks, R. (2002). The individual and the institutional: balancing professional development needs within further education. In G. Trorey & C. Cullingford (Eds.), *Professional development and institutional needs* (pp. 35–50). Hampshire, UK: Ashgate Publishing.

Brookover, W. B.. Beady, C.. Flood, P., & Schwettzer. J. (1979). *School systems and student achievement: schools make a difference.* New York: Praeger

Brookover, W. B., Beady, C. Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *Schools, social systems and student achievement: Schools can make a difference.* New York: Praeger.

ANASTASIA PANAYIOTOU

Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed, pp. 328‑375). New York: MacMillan.

Brown, M. W., & Mels, G. (1990). *RAMONA PC: User Manual.* Pretoria: University of South Africa.

Brown, B.W. and Saks, D.H. (1986) Measuring the Effects of Instructional Time on Student Learning: Evidence From the Beginning Teacher Evaluation Study, *American Journal of Education, 94*, 480–500.

Buchmann, C. (2002). Measuring family background in international studies of education: Conceptual issues and methodological challenges. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 150–197). Washington DC: National Academy Press

Byrne, B. M., (1994). *Structural Equation Modeling with EQS and EQS/Windows.* Newbury Park, CA: Sage.

Caldwell, J.B., & Spinks, M.J. (1993). *The self-managing school.* London: The Farmer Press.

Campbell, R. J., Kyriakides, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education, 29*(3), 347 – 362.

Campbell, R.J., Kyriakides, L., Muijs, R.D. and Robinson, W. (2004) *Assessing Teacher Effectiveness: A Differentiated Model.* London: Routledge Falmer.

Carlo, R. (2012). Improving Quality in Education – Dynamic Approaches to School Improvement. *British Journal of Educational Studies, 60*(3), 283-284.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record, 64*, 722-733.

Carroll, J. B. (1989). The Carroll model a 25-year retrospective and prospective view. *Educational Researcher, 18*(1), 26-31.

Cazden, C. B. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 432–463). New York: MacMillan.

Centra, J. A. (1993). *Reflective faculty evaluation*. San Francisco: Jossey- Bass.

Cervini, R.A. (2009). Class, school, municipal, and state effects on mathematics achievement in Argentina: A multilevel analysis. *School Effectiveness and*

School Improvement, 20(3), 319-340.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction, 8*(4), 293-332.

Christie, R., & Lindauer, F. (1963). Personality structure. *Annual review of psychology, 14*(1), 201-230.

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B.P.M. (2014). Searching for stages of teacher skills in assessment. *Studies in Educational Evaluation, 40*, 1-11.

Clark, D. L., Lotto, L. S., & Astuto, T. A. (1984). Effective schools and school improvement: A comparative analysis of two lines of inquiry. *Educational Administration Quarterly, 20*(3), 41-68.

Coe, R., & Fitz-Gibbon, C. T. (1998). School effectiveness research: Criticisms and recommendations. *Oxford Review of Education*, *24*(4), 421 – 438.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, L., Manion, L. & Morrison, K. (2000). *Research methods in education*. London: Routledge/Falmer.

Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.

Cools, W., De Fraine, B., Van den Noortgate, W. & Onghena, P. (2009). Multilevel design efficiency in educational effectiveness research. *School Effectiveness and School Improvement*, *20*(3), 357-373.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. New York: Rand McNally.

Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within study comparisons. *Journal of policy analysis and management, 27*(4), 724-750.

Creemers, B. (1992). School effectiveness and effective instruction—the need for a further relationship. In J. Bashi & Z. Sass (Eds.), *School effectiveness and improvement.* Jerusalem, Israel: Hebrew University Press.

ANASTASIA PANAYIOTOU

Creemers, B.P.M. (1993). Towards a theory of educational effectiveness. *Paper presented at the 6th Annual Meeting of the International Congress for School Effectiveness and Improvement (ICSEI).* Norrköping, Sweden, January 3-6.

Creemers, B.P.M. (1994a). The Effective Classroom. London: Cassell.

Creemers, B.P.M. (1994b). The history, value and purpose of school effectiveness studies In D. Reynolds, B.P.M. Creemers, P.S. Nesselrodt, E. C. Shaffer, S. Stringfield, & C. Teddlie (Eds.). *Advances in school effectiveness research and practice* (p. 9-25). London: Pergamon.

Creemers, B.P.M. (1997). Visions that work: A further elaboration of the Comprehensive model of Educational Effectiveness. *Paper presented at the Annual Meeting of the American Educational Research Association*. (Chicago, IL, March 24-28, 1997).

Creemers, B.P.M. (2002a). An International Perspective on Efficiency in School Financing. *School Effectiveness and School Improvement, 13*(2), 247-252.

Creemers, B.P.M. (2002b). From school effectiveness and school improvement to effective school improvement: Background, theoretical analysis and outline of the empirical study. *Educational Research and Evaluation, 8*(4), 343-362.

Creemers, B.P.M. (2007). Combining Different Ways of Learning and Teaching in a Dynamic Model of Educational Effectiveness. *Journal of Basic Education, 17*(1).

Creemers, B.P.M. (2006). The importance and perspectives of international studies in educational effectiveness. *Educational Research and Evaluation*, 12(6), 499-511.

Creemers, B.P.M. & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement, 17*(3), 347-366.

Creemers, B.P.M. & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness. A contribution to Policy, Practice and Theory n contemporary schools.* London: Routledge.

Creemers, B.P.M., & Kyriakides, L. (2010a). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research, 54*(1), 263-294.

Creemers, B.P.M., & Kyriakides, L. (2010b). Using the dynamic model to develop an evidence-based and theory-driven approach to school improvement. *Irish Educational Studies, 29*, 5-23.

ANASTASIA PANAYIOTOU

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. (2013). Using the dynamic model to identify stages of effective teaching: An introduction to the special issue. *Journal of Classroom Interaction, 48*(2), 4-9.

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

Creemers, B.P.M., Kyriakides, L., & Antoniou, P. (2013). *Teacher professional development for improving quality in teaching.* Dordrecht, the Netherlands: Springer.

Creemers, B.P.M., Kyriakides, L., & Sammons, P. (2010). *Methodological Advances in Educational Effectiveness Research.* London and New York: Routledge.

Creemers, B. P., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School effectiveness and school Improvement*, *7*(3), 197-228.

Creemers, B., Reynolds, D., Chrispeels, J., Mortimore, P., Murphy, J., Stringfield, S., Stoll, L., & Townsend, T. (1998). The future of school effectiveness and improvement (A report on the special sessions and plenary at ICSEI 1998 in Manchester,UK). *School Effectiveness and School Improvement*, *9*(2), 125–134.

Creemers, B.P.M., Reynolds, D. & Swint, F.E. (1994). *The International School Effectiveness Research Programme ISERP, First Results of the Quantitative Study.* Paper presented at the British Educational Research Association conference, Oxford, September 1994.

Creemers, B.P.M. & Scheerens, J. (1994). Developments in the educational effectiveness research programme. *International Journal of Educational Research*, *21*(2), 125-140.

Creemers, B.P.M. & Van der Werf, G. (2000). Economic viewpoints in educational effectiveness: Cost-effectiveness analysis of an educational improvement project. *School Effectiveness and School Improvement*, *11*(3), 361-384.

Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements. Theory of Generalizability Scores and Profiles.* New York: Wiley.

Danju, I., Miralay, F., & Baskan, G. A. (2014). The Comparison between Successful and Unsuccessful Countries in PISA, 2009. *Procedia-Social and Behavioral Sciences, 116,* 1636-1640.

Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence. *Education Policy Analysis Archives, 8* (1), http://epaa.asu.edu/epaa/v8n1/.

De Corte, E. (2000). Marrying theory building and the improvement of school practice: a permanent challenge for instructional psychology. *Learning and Instruction, 10*(3), 249–266.

De Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the first year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*(1), 3 – 31.

Demetriou, D., & Kyriakides, L. (2012). The impact of school self-evaluation upon student achievement: a group randomization study. *Oxford Review of Education, 38*(2), 149-170.

Den Brok, P. (2001). *Teaching and student outcomes.* Utrecht, The Netherlands: WCC.

Den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement, 15*(3/4), 407–442.

Denison, D. (1996) What is the difference between organisational culture and organisational climate? A native's point of view on a decade of paradigm wars. *Academy of Management Review, 21*(3), 619–654.

D'Haenens, E., Van Damme, J. & Onghena, P. (2010). Multilevel exploratory factor analysis: illustrating its surplus value in educational effectiveness research. *School Effectiveness and School Improvement, 21*(2), 209-235.

Ding, M., Li, Y., Li, X., & Kulm, G. (2008). Chinese teachers' perceptions of students' classroom misbehaviour. *Educational Psychology, 28*(3), 305-324.

Doyle, W. (1984). How order is achieved in a classroom: An Interim report. *Journal of Curriculum Studies, 16,* 259-277.

ANASTASIA PANAYIOTOU

Doyle, W. (1986). Classroom organization and management. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching, Third Edition* (pp. 392-431). New York: Macmillan.

Doyle, W. (1990). Classroom knowledge as a foundation for teaching. *Teachers College Record, 91*(3), 347-360.

Drent, M., Meelissen, M. R., & van der Kleij, F. M. (2013). The contribution of TIMSS to the link between school and classroom factors and student achievement. *Journal of Curriculum Studies, 45*(2), 198-224.

Driessen, G. & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement, 11*(1), 57-79.

Driscoll, A., Peterson, K., Crow, H., & Larson, B. (1985). Student reports for primary teacher evaluation. *Educational Research Quarterly, 9*(3), 43-50.

Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. J. Jonassen (Ed.), *Handbook of research for educational communication and technology* (pp. 170–198). New York: McMillan.

Edmonds, R.R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15-27.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Elberts, R.W. & Stone, J.A. (1988). Student achievement in public schools: do principles make a difference? *Economics Education Review, 7*, 291-299.

Elliott, 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.6-25). Slough: NFER.

Evertson, C. M., Anderson, C. W., Anderson, L. M., & Brophy, J. E. (1980). Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes. *American educational research journal, 17*(1), 43-60.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics, 49*(1), 92-107.

ANASTASIA PANAYIOTOU

Feldman, K. A. (1997), Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry & J. C. Smart (Eds.), *Effective teaching in higher education: Research and practice* (pp. 368-395). New York: Agathon Press.

Fitzpatrick, K. A. (1982). The effect of a secondary classroom management training program on teacher and student behavior. *Paper presented at the annual meeting of the American Educational Research Association*, New York,

Fox, J. (1987). Statistical models for non-experimental data: A comment on Freedman. *Journal of Educational Statistics*, *12*(2), 161-164.

Fraser, B. J. (1991). Two decades of classroom environment research. In B. J. Fraser & H. J. Walberg (Eds.), *Educational environments: Evaluations, antecedents and consequences.* New York: Pergamon Press.

Fraser, B.J. (1998). Science leaming environments: Assessment, effects and determinants. In B.J. Fraser & K.G. Tobin (Eds.), *Intemational handbook of science education* (Part 1, pp. 527-564). London: Kluwer Academic Publishers.

Fuller, B. (1987). Raising school quality in developing countries: What investments boost learning? *Review of Educational Research, 58*,255-292.

Galbraith, S., Stat, M., & Marschner, I. C. (2002). Guidelines for the design of clinical trials with longitudinal outcomes. *Controlled clinical trials, 23*(3), 257-273.

Galton, M. (1987). An ORACLE chronicle: A decade of classroom research. *Teaching and Teacher Education*, *3*(4), 299-313.

Gardner, J., Harlen, W., Hayward, L., Stobart, G. & Montgomery, M., (2010). *Developing Teacher Assessment.* Maidenhead: Open University Press.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.

Glass, G. V, McGaw, B.,& Smit.h, M.L. *Meta-analysis in social research*. Beverly Hills, Ca.: Sage, 1981.

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Paper Presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education,* Columbus, OH, October 8-10.

Glisson, C. (2000) Organisational climate and culture, in: R. Patti (ed.) *The handbook of social welfare* (Thousand Oaks, CA, Sage), 195–218.

ANASTASIA PANAYIOTOU

Gijbels, D., Van de Watering, G., Dochy, F., & Van den Bossche, P. (2006). New learning environments and constructivism: The students' perspective. *Instructional Science,34* (3), 213-226.

Goldstein, H. (1968). Longitudinal studies and the measurement of change. *Journal of the Royal Statistical Society, 18*(2), 93-117.

Goldstein, H. (1997). Methods in School Effectiveness Research. *School effectiveness and school improvement, 8*(4), 369-395.

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Edward Arnold.

Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: principles, policy & practice, 11*(3), 319-330.

Goldstein, H., & Woodhouse, G. (2000). School effectiveness research and educational policy. *Oxford Review of Education, 26*(3-4), 353-363.

Gonczi, A. (1994). Competency based assessment in the professions in Australia. *Assessment in education, 1*(1), 27-44.

Good, T. L., & Grouws, D. A. (1979). The Missouri mathematics effectiveness project. *Journal of Educational Psychology, 71*, 355-362.

Gorard, S., & Smith, E. (2004). An international comparison of equity in education systems. *Comparative education, 40*(1), 15-28.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology, 84*(11), 2809-2815.

Grek, S. (2009). Governing by numbers: The PISA 'effect'in Europe. *Journal of Education Policy, 24*(1), 23-37.

Grieve, A. M. (2010). Exploring the characteristics of "teachers for excellence": Teachers' own perceptions. *European Journal of Teacher Education, 33*(3), 265–277.

Gruehn, S. (1995). The compatibility of cognitive and noncognitive objectives of instruction. *Zeitschrift Für Padagogik, 41*(4), 531–553.

Guldemond, H., & Bosker, R. J. (2009). School effects on students' progress–a dynamic perspective. *School Effectiveness and School Improvement, 20*(2), 255-268.

Gunst, R. F., & Webster, J. T. (1975). Regression analysis and problems of multicollinearity. *Communications in Statistics-Theory and Methods, 4*(3), 277-292.

ANASTASIA PANAYIOTOU

Gustafsson, J. E. (2007). Understanding causal influences on educational achievement through analysis of differences over time within countries. In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 37–63). Washington, DC: Brookings Institution Press.

Gustafsson, J.E. (2010). Longitudinal Designs In. B.P.M. Creemers, L. Kyriakides & P. Sammons (eds). *Methodological Advances in Educational Effectiveness Research* (pp. 77-101). London and New York: Routledge.

Gustafsson, J. E. (2013). Causal inference in educational effectiveness research: a comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement, 24*(3), 275-295.

Hair, J.F. (2011). Multivariate data analysis: an overview. In M. Lovric (ed). *International Encyclopedia of Statistical Science* (pp. 905-907). Berlin Heidelberg: Springer.

Hanushek, E.A. (1986).The Economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-1177.

Hanushek, E. (1994). *Making schools work. Improving performance and controlling costs.* Washington, D.C.: The Brookings Institution.

Hanushek, E.A. (1997). Outcomes, incentives and beliefs: Reflections on analysis of the economics of schools. *Educational Evaluation and Policy Analysis, 19*(4), 301–308.

Hargreaves, D. H. (1995). School culture, school effectiveness and school improvement. *School effectiveness and school improvement, 6*(1), 23-46.

Harjunen, E. (2012). Patterns of control over the teaching–studying–learning process and classrooms as complex dynamic environments: a theoretical framework. *European Journal of Teacher Education, 34*(2), 139-161.

Harrington, D. (2008). *Confirmatory Factor Analysis.* USA: Oxford University Press.

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* London: Rouledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Haystead, M. W., & Marzano, R. J. (2009). *Meta-analytic synthesis of studies conducted at Marzano Research Laboratory on instructional strategies.* Englewood, CO: Marzano Research Laboratories.

ANASTASIA PANAYIOTOU

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Heck, R. H., & Moriyama, K. (2010). Examining relationships among elementary schools' contexts, leadership, instructional practices, and added-year outcomes: A regression discontinuity approach. *School Effectiveness and School Improvement, 21*(4), 377–408.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Economic and Social Measurement, 5*(4), 475-492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society, 47*(1), 153-161.

Hendriks, M. (2014). The influence of school size, leadership, evaluation and time on student outcomes: four reviews and meta-analyses. *Unpublished PhD thesis, University of Twente*, Enschede, The Netherlands.

Hendrikx, K., Verhaege, J. P., Ghesquiére, P., Maes, F., & Van Damme, J. (2006). *Longitudinal onderzoek in het basisonderwijs Toetsen derde leerjaar (schooljaar 2006-2006).* LOA-rapport nr. 45.

Heyneman, S. P., & Loxley, W. A. (1983). The effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of sociology, 88,* 1162-1194.

Hill, P. W., & Rowe, K. J. (1996). Multilevel modelling in school effectiveness research. *School effectiveness and school improvement, 7*(1), 1-34.

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of management, 23*(6), 723-744.

Holdaway, E. A., & Johnson, N. A. (1993). School effectiveness and effectiveness indicators. *School Effectiveness and School Improvement, 4*(3), 165-188.

Hopkins, D. (2001). *School Improvement for Real*. London: RoutledgeFalmer.

Houghton, S., Wheldall, K., & Merrett, F. (1988). Classroom behavior problems which secondary school teachers say they find most troublesome. *British Educational Research Journal, 4,* 297-312.

Howie, S. J. (2005). Contextual factors at the school and classroom level related to pupils' performance in mathematics in South Africa. *Educational research and evaluation, 11*(2), 123-140.

ANASTASIA PANAYIOTOU

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

Hox, J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge.

Hoy, W.K. (1990) Organisational climate and culture: a conceptual analysis of the school workplace, *Journal of Educational and Psychological Consultation, 1*(2), 149–168.

Hunter, J. E. & Schmidt, F. L. (1995). *Methods of Meta-Analysis: Correcting Errors and Bias in Research Findings.* Sage, Thousand Oaks, California.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological bulletin, 107*(2), 139.

Joyce, B., Weil, M., & Calhoun, E. (2000). *Models of teaching.* Boston: Allyn & Bacon.

Kelly, A. (2012). Measuring 'equity'and 'equitability'in school effectiveness research. *British Educational Research Journal, 38*(6), 977-1002.

Klein, J. D., & Pridemore, D. R. (1994). Effects of orienting activities and practice on achievement, continuing motivation, and student behaviors in a cooperative learning environment. *Educational Technology Research and Development, 42*(4), 41-54.

Kieffer, M. J. (2012). Before and after third grade: Longitudinal evidence for the shifting role of socioeconomic status in reading growth. *Reading and Writing, 25*(7), 1725-1746.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and instruction, 12*(1), 1-10.

Klees, S. J., & Qargha, O. (2014). Equity in education: The case of UNICEF and the need for participative debate. *PROSPECTS,* 1-13.

Kline, R.H. (1998). *Principles and practice of structural equation modelling.* London: Gilford Press.

Knuver, A. W., & Brandsma, H. P. (1993). Cognitive and affective outcomes in school effectiveness research. *School effectiveness and school improvement, 4*(3), 189-204.

Kokkinou, E., & Kyriakides, L. (2016). Assessing teacher effectiveness: Searching for classroom context effects in measuring quality of teaching. *Paper to be presented at the 29th International Congress for School Effectiveness and Improvement (ICSEI) 2016.* Glasgow, Scotland, January 6-9.

ANASTASIA PANAYIOTOU

Konstantinou-Katzi, P., Tsolaki, E., Meletiou-Mavrotheris, M., & Koutselini, M. (2013). Differentiation of teaching and learning mathematics: an action research study in tertiary education. *International Journal of Mathematical Education in Science and Technology, 44*(3), 332-349.

Kulinna, P. H. (2007). Teachers' attributions and strategies for student misbehavior. *The Journal of Classroom Interaction, 42*(2), 21-30.

Kuyper, H., Van der Werf, M. P. C., & Lubbers, M. J. (2000). Motivation, meta-cognition and self-regulation as predictors of long term educational attainment. *Educational Research and Evaluation, 6*(3), 181-205.

Kyriakides, L. (2012). Advances in School Effectiveness Theory. In C. Chapman, P. Armstrong, A. Harris, D. Muijs, D. Reynolds, & P. Sammons (Eds.), *School Effectiveness and Improvement Research, Policy and Practice: Challenging the orthodoxy?* (pp. 44-57). London: Routledge.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19*(4), 429-446.

Kyriakides, L. (2007). Generic and Differentiated Models of Educational Effectiveness: Implications for the Improvement of Educational Practice. In T. Townsend (Ed.) *International Handbook of School Effectiveness and Improvement* (pp. 41–56). Dordrecht, the Netherlands: Springer.

Kyriakides, (2006a). Introduction International Studies on Educational Effectiveness. *Educational Research and Evaluation, 12*(6), 489-497.

Kyriakides, (2006b). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data. *Educational Research and Evaluation, 12*(6), 513-534.

Kyriakides, L. (2006c). Measuring the learning environment of the classroom and its effect on cognitive and affective outcomes of schooling. In D. L. Fisher, & M. S. Khine (Eds.), *Contemporary Approaches to Research on Learning Environments: World Views* (pp. 369-408). Hackensack, New Jersey: World Scientific.

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement, 16*(2), 103 – 152.

ANASTASIA PANAYIOTOU

Kyriakides, L., & Anthimou, M. (2016). The Effects of Teacher Factors on Students' Cognitive and Metacognitive Skills: a study testing the validity of the dynamic model. *Paper to be presented at the 29th International Congress for School Effectiveness and Improvement (ICSEI) 2016.* Glasgow, Scotland, January 6-9.

Kyriakides, L., Archambault, I., & Janosz, M. (2013). Searching for stages of effective teaching: a study testing the validity of the dynamic model in Canada. *Journal of Classroom Interaction, 48*(2), 11-24.

Kyriakides, L., & Campbell, R. J. (2004). School self-evaluation and school improvement: A critique of values and procedures. *Studies in Educational Evaluation, 30*(1), 23–36.

Kyriakides, L., Campbell, R. J., & Christofidou, E. (2002). Generating criteria for measuring teacher effectiveness through a self-evaluation approach: A complementary way of measuring teacher effectiveness. *School Effectiveness and School Improvement, 13*(3), 291-325.

Kyriakides, L., Campbell, R.J. & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers' Comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4), 501-529.

Kyriakides, L. & Charalambous, C. (2005). Using educational effectiveness research to design international comparative studies: Turning limitations into new perspectives. *Research Papers in Education, 20*(4), 391-412.

Kyriakides, L., & Christoforou, Ch. (2011). A Synthesis of Studies Searching for Teacher Factors: Implications for Educational Effectiveness Theory. *Paper presented at the American Educational Research Association (AERA) 2011 Conference.* New Orleans, April 2011.

Kyriakides, L., Christoforou, C., & Charalambous, C.Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education, 36*, 143-152.

Kyriakides, L., & Creemers, B.P.M. (2011). Can Schools Achieve Both Quality and Equity? Investigating the Two Dimensions of Educational Effectiveness. *Journal of Education for Students Placed at Risk, 16*(4), 237-254.

ANASTASIA PANAYIOTOU

Kyriakides, L., & Creemers, B.P.M. (2010). Meta-analyses of effectiveness studies. In B.P.M. Creemers, L. Kyriakides, & P. Sammons, (eds), *Methodological Advances in Educational Effectiveness Research*, 246-276, London and New York: Routledge.

Kyriakides, L., & Creemers, B.P.M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: a study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183-205.

Kyriakides, L., & Creemers, B.P.M. (2009). The effects of teacher factors on different outcomes: two studies testing the validity of the dynamic model. *Effective Education, 1*(1), 61-86.

Kyriakides, L., Creemers, B.P.M. & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12-23.

Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. *British Educational Research Journal, 36*(5), 807-830.

Kyriakides, L., Creemers, B. P., Antoniou, P., Demetriou, D., & Charalambous, C. Y. (2015). The impact of school policy and stakeholders' actions on student learning: A longitudinal study. *Learning and Instruction, 36*, 113-124.

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., & Demetriou, D. (2005, January). *Using international comparative studies for establishing generic and differentiated models of educational effectiveness: The PISA study*. Paper presented at the International Congress for School Effectiveness and Improvement 2005 Conference, Barcelona, Spain.

Kyriakides, L. & Tsangaridou, N. (2004). School effectiveness and teacher effectiveness in physical education. *Paper presented at the 85th Annual Meeting of the American Educational Research Association.* San Diego, CA, April.

Kyriakides, L. & Tsangaridou, N. (2008). Towards the development of generic and differentiated models of educational effectiveness: A study on school and teacher

ANASTASIA PANAYIOTOU

effectiveness in physical education. *British Educational Research Journal, 34*(6), 807–838.

Last, J., & Chown, A. (1996). Competence-based approaches and initial teacher training for FE. In J. Robson (Ed.), *The professional FE teacher, staff development and training in the corporate college* (pp. 20–32). Aldershot, UK: Avebury.

Legotlo, M. W., Maaga, M. P., & Sebego, M. G. (2002). Perceptions of stakeholders on causes of poor performance in Grade 12 in a province in South Africa. *South African Journal of Education, 22*(2), 113-118.

Leikin, R., & Zaslavsky, O. (1997). Facilitating student interactions in mathematics in a cooperative learning setting. Journal *for Research in Mathematics Education, 28*(3), 331-354.

Levine, D.U., & Lezotte, L.W. (1990). *Unusually effective schools: A review and analysis of research and practice.* Madison, WI: National Center for Effective Schools Research and Development.

Likert, R. (1931). A technique for the measurement of attitudes. *Archives of Psychology*. New York: Columbia University Press.

Lindo, E. J. (2014). Family background as a predictor of reading comprehension performance: An examination of the contributions of human, financial, and social capital. *Learning and Individual Differences, 32*, 287-293.

Little, E. (2005). Secondary school teachers' perceptions of students' problem behaviours. *Educational Psychology, 25*(4), 369-377.

Lockheed, M. E., & Komenan, A. (1989). Teaching quality and student achievement in Africa: The case of Nigeria and Swaziland. *Teaching and Teacher Education, 5*(2), 93-113.

Longford, N. (1993). *Random coefficient models*. Oxford: Clarendon Press.

Luyten, H. & Sammons, P. (2010). Multilevel Modelling, In B.P.M. Creemers, L. Kyriakides, & P. Sammons, (eds), *Methodological Advances in Educational Effectiveness Research,* 246-276, London and New York: Routledge.

Luyten, H., Tymms, P., & Jones, P. (2009). Assessing school effects without controlling for prior achievement? *School Effectiveness and School Improvement, 20*(2), 145-165.

ANASTASIA PANAYIOTOU

Luyten, H., Visscher, A. & Witziers, B. (2005). School effectiveness research: From a review of the criticism to recommendations for further development. *School effectiveness and school improvement, 16*(3), 249-279.

Lynch, K. (2001). Equality in education. *Studies: An Irish Quarterly Review, 90*(360), 395-411.

Lynch, K., & Baker, J. (2005). Equality in education An equality of condition perspective. *Theory and Research in Education, 3*(2), 131-164.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86-92.

MacBeath, J. (2005). *Schools must speak for themselves: The case for school self-evaluation.* London: Routledge.

Mackenzie, D. E. (1983). Research for school improvement: An appraisal of some recent trends. *Educational Researcher, 12*(4), 5-17.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research, 11*, 253–388.

Marsh, H. W. (2007). Students' evaluations of university teaching: A multidimensional perspective. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence based perspective* (pp. 319–384). New York: Springer.

Marsh, H. W., & Roche, L. A. (1994). *The use of students' evaluations of university teaching to improve teaching effectiveness. Canberra, Australian Capital Territory*, Australia: Department of Employment, Education and Training.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective. *American Psychologist, 52*, 1187–1197

Martin, M. O. (1996). Third international mathematics and science study. In M. O. Martin & D. L. Kelly (Eds.), *TIMSS technical report* (Vol. 1, pp. 1.1 – 1.19). Boston: International Association for the Evaluation of Student Achievement.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction.* Alexandria, VA: Association for Supervision and Curriculum Development.

ANASTASIA PANAYIOTOU

Maxwell, S., Kelley, K., & Rausch, J. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537-563

McCutcheon, G. (1981). On the interpretation of classroom observations. *Educational Researcher, 10*(5), 5-10.

McNamara, G., & O'Hara, J. (2005). Internal review and self-evaluation—the chosen route to school improvement in Ireland?. *Studies in Educational Evaluation*, *31*(4), 267-282.

Merrett, F., & Wheldall, K. (1993). How do teachers learn to manage classroom behaviour? A study of teachers' opinions about their initial training with special reference to classroom behaviour management. *Educational Studies*, *19*(1), 91-106.

Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and learning*, *1*(1), 85-97.

Milich, R., & Fitzgerald, G. (1985). Validation of inattention/overactivity and aggression ratings with classroom observations. *Journal of Consulting and Clinical Psychology*, *53*(1), 139.

Ministry of Education and Culture, Republic of Cyprus (2014a). *Annual report*. Nicosia, Cyprus.

Ministry of Education and Culture, Republic of Cyprus (2014b). *Incorporating technologies of information and communication in the learning and managing process* (in Greek). Nicosia, Cyprus.

Moerbeek, M. (2008). Powerful and cost-efficient designs for longitudinal intervention studies with two treatment groups. *Journal of Educational and Behavioral Statistics*, *33*(1), 41-61.

Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis*, *14*(4), 307-332.

Mortimore, P. (1991). The nature and findings of research on school effectiveness in the primary sector, in S. Riddell and S. Brown (eds) *School Effectiveness Research: its Messages for School Improvement,* 9-20, Edinburgh: HMSO.

Mortimore, P. (1992). Issues in school effectiveness. In D. Reynolds & P. Cuttance (Eds.), *School effectiveness research, policy and practice* (pp. 154–163). London, UK: Cassell.

ANASTASIA PANAYIOTOU

Mortimore, P. (1993). School effectiveness and the management of effective learning and teaching. *School effectiveness and school improvement, 4*(4), 290-310.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). School matters: The junior years. Somerset: Open Books (Reprinted in 1995, London: Paul Chapman).

Mullis, I. V. S., Martin, M. O., Gonzalez E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). TIMSS 1999-*International mathematics report.* Boston: International Association for the Evaluation of Student Achievement.

Mullis, I. V. S., Martin, M. O., Gonzalez E. J., & Chrostowski, S. (2004). TIMSS 2003 *International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades* (TIMSS and PIRLS International Study Center, 2003).

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O`Sullivan, C. Y., Arora, A., & Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks*: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Muijs, D., Ainscow, M., Chapman, C., & West, M. (2011). Achieving Excellence and Equity: Reflections on the Development of Practices in One Local District Over 10 Years. In *Collaboration and Networking in Education* (pp. 85-101). Springer Netherlands.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art–teacher effectiveness and professional learning. *School Effectiveness and School Improvement, 25*(2), 231-256.

Muijs, D., & Reynolds, D. (2011). *Effective teaching. Evidence and practice.* London: Sage.

Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: Some preliminary findings from the evaluation of the mathematics enhancement programme. *School Effectiveness and School Improvement, 11*, 273–303.

Muijs, D., & Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics. Educational *Research and Evaluation, 9*, 289–314.

Murillo, F. J. & Rincon, M. L. H. (2002). The Ibero American Network for Research on School Effectiveness and School Improvement: AWay to Increase Educational Quality and Equity. *School Effectiveness and School Improvement*, *13*(1), 123-132.

ANASTASIA PANAYIOTOU

Muthén, L. K., & Muthén, B. O. (2008). Mplus (Version 5.1). Los Angeles, CA: Muthén & Muthén.

Muthen, B. O., and Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (ed.), *Sociological Methodology*, pp. 267–316. Washington, DC: American Sociological Association.

Nevo, D. (2001). School evaluation: internal or external?. *Studies in Educational Evaluation*, *27*(2), 95-106.

Nordenbo, S. E., Holm, A., Elstad, E., Scheerens, J., Soegaard Larsen, M., Uljens, M., Hauge, T.E. (2009). *Research mapping of input, process and learning in primary and lower secondary schools: Technical report.* Copenhagen: DPU, Aarhus University.

Nutall D., Goldstein, H., Prosser, R. and Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research, 13,* 769-776.

OECD (2015a). *Call for tenders: Longitudinal study of social and emotional skills in cities.* Accessed on October 10, 2015 from: http://www.oecd.org/callsfortenders.

OECD (2015b). *Call for tenders: Implementation of the first cycle of the International Survey of Staff in Early Childhood Education and Care*. Accessed on October 10, 2015 from: http://www.oecd.org/callsfortenders.

Opdenakker, M.C. & Van Damme, J. (2006). Differences between secondary schools: A study about school context, group composition, school practice, and school effects with special attention to public and Catholic schools and types of schools. *School Effectiveness and School Improvement*, *17*(1), 87-117.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist*, *38*(1), 1-4.

Panayiotou, A., Kyriakides, L., & Christoforidou, M. (2015). The Long-term Effect of the Dynamic Approach to Teacher Professional Development on Improving Teaching. *Paper presented to the American Educational Research Association (AERA) 2015 Conference.* Chicago, Illinois, April 16 – 20, 2015.

Paris S.G., & Paris, A.H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist. 36*(2), 89–101.

Peaker, G. (1975). *An empirical study of education in twenty-one countries: A technical report.* New York: Wiley.

ANASTASIA PANAYIOTOU

Pereyra, M. A., Kotthoff, H. G., & Cowen, R. (2011). *PISA under examination* (pp. 1-14). SensePublishers.

Peterson, K., & Stevens, D. (1988). Student reports for school teacher evaluation. *Journal of Personnel Evaluation in Education*, *2*(1), 19-31.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421‑425.

Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. Cambridge University Press.

Plewis, I., & Hurry, J. (1998). A multilevel perspective on the design and analysis of intervention studies. *Educational Research and Evaluation, 4*(1), 13-26.

Plucker, J. A. (2015, August). Advanced Academic Performance: Exploring country-level differences in the pursuit of educational excellence. *Policy Brief*, No. 7, Amsterdam, IEA, http://www.iea.nl/policy_briefs.html

Postlethwaite, T. N. (1995) International empirical research in comparative education: an example of the studies of the International association for the evaluation of Educational Achievement (IEA). *Journal fur Internationale Bildungsforschung, 1* (1), 1–19.

Postlethwaite, T. N., & Ross, K. N. (1992). *Effective Schools in Reading: Implications for Educational Planners. An Exploratory Study*. Den Haag, the Netherlands: IEA

Price, J. H., & Murnan, J. (2004). Research limitations and the necessity of reporting them. *American Journal of Health Education*, *35*(2), 66-67.

Purkey, S. C, & Smith, M. S. (1983). Effective schools: A review. *Elementary School Journal*, *83*(4), 427-452.

Purves, A. C. (1973). *Literature education in ten countries:I nternationasl tudies in evaluation II.* New York: Wiley.

Rabash., J, Browne., W, Goldstein., H, Yang., M, Plewis., I, Healy., M, et al. (2000). A user's guide to MLwiN. London: University of London.

Rabe-Hesketh, S., & Skrondal, A. (2008). Multilevel and longitudinal modeling using Stata. STATA press.

Ralph, J. H., & Fennessey, J. (1983). Science or reform: Some questions about the effective schools model. *Phi Delta Kappan*, 689-694.

ANASTASIA PANAYIOTOU

Raudenbush, S.W., & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*, 1‑17.

Raudenbush, S., & Bryk, T. (1991). Hierarchical linear modeling. Thousand Oaks, CA: Sage.

Raudenbush, S., & Bryk, T. (2002). Hierarchical linear modeling (2nd ed.). Thousand Oaks, CA: Sage.

Reezigt, G.J., Guldemond, H. & Creemers, B.M.P. (1999). Empirical validity for a Comprehensive model on educational effectiveness. *School Effectiveness and School Improvement, 10*(2), 193-216.

Reynolds, D. (2000). School effectiveness: The international dimension. In C. Teddlie & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 232 – 256). London: Falmer.

Reynolds, D. (2006). World Class Schools: Some methodological and substantive findings and implications of the International School Effectiveness Research Project (ISERP). *Educational Research and Evaluation, 12*(6), 535-560.

Reynolds, D., Creemers, B.P.M., Stringfield, S. & Teddlie, C. (1998). Climbing an educational mountain: conducting an international school effectiveness research project in: G. Walford (Ed.) *Doing Research about Education* (Lewes, Falmer Press).

Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (Eds.). (2002). *World class schools.* London: RoutledgeFalmer.

Reynolds, D., Hopkins, D., & Stoll, L. (1993). Linking school effectiveness knowledge and school improvement practice: Towards a synergy. *School Effectiveness and School Improvement, 4*, 37‑58.

Reynolds, D. & Packer, A. (1992). School effectiveness and school improvement in the 1990s, in D. Reynolds and P. Cuttance (eds). *School Effectiveness: Research, Policy and Practice*, 171-188, London: Cassell.

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., & Van Damme, J. (2011, January). Educational Effectiveness Research (EER): A state of the art review. *Paper presented at the annual meeting of the International Congress for School Effectiveness and Improvement*, Limassol, Cyprus.

ANASTASIA PANAYIOTOU

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197-230.

Reynolds, D., & Stoll, L. (1996). Merging school effectiveness and school improvement: The knowledge base. In D. Reynolds, R. Bollen, B. Creemers, D. Hopkins, L. Stoll, & N. Lagerweij (Eds.), *Making good schools: Linking school effectiveness and school improvement*, pp. 94‑112. London: Routledge.

Reynolds, D. & Teddlie, C. (2000) The future agenda for school effectiveness research, in: C. Teddlie & D. Reynolds (eds) *The international handbook of school effectiveness research* (London & New York, Falmer Press), 22–343.

Reynolds, D., Teddlie, C., Creemers, B.P.M., Cheng, Y.C., Dundas, B., Green, B. Epp, J.R., Hauge, T.E., Schaffer, E.C., & Stringfield, S. (1994). School Effectiveness research: A review of the international literature. In D. Reynolds, B.P.M. Creemers, P.S. Nesselrodt, E.C. Schaffer, S. Stringfield, & C. Teddlie (Eds.), *Advances in school effectiveness research and practice* (pp. 25-54). London: Pergamon.

Reynolds, A.J. & Walberg, H. J. (1992). A structural model of high school mathematics outcomes. *The Journal of Educational Research, 85*(3), 150-158.

Rhoads, C. H. (2011). The implications of "contamination" for experimental design in education. *Journal of Educational and Behavioral Statistics, 36*(1), 76-104.

Robinson, J.P., Rush, J.G. & Head, K. B. (1974). Criteria for an attitude scale. In G. M., Maranell (Ed.). *Scaling: A sourcebook for behavioral scientists*. Chicago: Aldine.

Robson, J. (1998). A profession in crisis: status, culture and identity in the further education college. *Journal of Vocational Education and Training, 50*(4), 585–607.

Rosenbaum, P. R. (1986). Dropping out of high school in the United States: An observational study. *Journal of Educational and Behavioral Statistics, 11*(3), 207-224.

Rosenshine, B., & Stevens, R. (1986). Teaching Functions. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 376–391). New York: Macmillan.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.).*The handbook of research synthesis* (pp.231-244). New York: Russell Sage Foundation.

ANASTASIA PANAYIOTOU

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* Cambridge University Press.

Rutter. M., Maughan, B., Mortimore, P. & Ouston, J. (with Smith, A.) (1979). *Fifteen thousand hours: secondary schools and their effects on children.* London: Open Books.

Sammons, P. (1999). *School effectiveness: Coming of age in the twenty-first century*. Lisse, The Netherlands: Swets & Zeitlinger.

Sammons, P. (2006). The Contribution of International Studies on Educational Effectiveness: Current and future directions. *Educational Research and Evaluation, 12*(6), 583-593.

Sammons, P. (2009). The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools. *School Effectiveness and School Improvement, 20*(1), 123–129.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: OFSTED.

Sammons, P., Nuttall, D. & Cuttance, P. (1993) Differential school effectiveness: results from a reanalysis of the Inner London Education Authority's Junior School Project data. *British Educational Research Journal, 19,* 381-405.

Savage, J. (2012). Improving Quality in Education: Dynamic Approaches to School Improvement. *Journal of Educational Administration and History, 44*(4), 396-398.

Scheerens, J. (1992). *Effective schooling: research, theory and practice*. London: Cassell.

Scheerens, J. (1993). Basic school effectiveness research: Items for a research agenda. *School Effectiveness and School Improvement, 4*(1), 17-36.

Scheerens, J. (1993). Effective schooling: Research, theory and practice. School *Effectiveness and School Improvement, 4*(3), 230-235.

Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement, 24*(1), 1-38.

Scheerens, J. (2014). School, teaching, and system effectiveness: some comments on three state-of-the-art reviews. *School Effectiveness and School Improvement, 25*(2), 282-290.

Scheerens, J., & Bosker, R.J. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon.

ANASTASIA PANAYIOTOU

Scheerens, J. & Creemers, B.P.M. (1989a) ‘Conceptualizing school effectiveness’, *International Journal of Educational Research, 13*, 691-706.

Scheerens, J. & Creemers, B. (1989b). Towards a more comprehensive conceptualization of school effectiveness. In B. Creemers, T. Peters, and D. Reynolds (Eds.) *School Effectiveness and School Improvement.* Lisse: Swets and Zeitlinger.

Scheerens, J., Witziers, B., & Steen, R. (2013). A meta-analysis of School Effectiveness studies. *Revista de educacion*, 361, 619-645.

Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62-87.

Schoenfeld, A.H. (1998). Toward a theory of teaching in context. *Issues in Education, 4*(1), 1‐94.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*, (pp. 39-83). Chicago, IL: Rand McNally.

Scriven, M. (1994). Using student ratings in teacher evaluation. *Evaluation Perspectives, 4*(1), 3-6.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of educational research, 77*(4), 454-499.

Shavelson, R. J., Webb, N. M. & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44*(6), 922-932.

Shepard, L. A. (1989). Why we need better assessment. *Educational Leadership, 46*(2), 4–8.

Shuell, T.J. (1996). Teaching and leaming in a classroom context. In D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726-764). New York: Macmillan.

Shui-fong, L., Yin-kum, L. & Shiu-kee, S. M. (2009). Classroom discourse analysis and educational outcomes in the era of education reform. *British Journal of Educational Psychology, 79*(4), 617-641.

Slater, R. O., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement, 3*(4), 247-257.

ANASTASIA PANAYIOTOU

Slavin, R. E., (1987). Mastery learning reconsidered. *Review of Educational Research, 57*(2), 175-213.

Slavin, R. E. (1996). *Education for all.* Exton, PA: Sweets & Zeitlinger Publishers.

Slavin, R. E. (2010). Experimental studies in education. In. B.P.M. Creemers, L. Kyriakides & P. Sammons (eds). *Methodological Advances in Educational Effectiveness Research* (pp. 77-101). London and New York: Routledge.

Snijders, T. (2011). Multilevel analysis. In M. Lovric (ed). *International Encyclopedia of Statistical Science* (pp. 879-882). Berlin Heidelberg: Springer.

Snijders, T. & Bosker, R. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling.* London: Sage.

Sprinthall, N., Reiman, A., & Thies-Sprinthall, L. (1996). Teacher professional development. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 666–703). New York: Macmillan.

Stankov, L., & Lee, J. (2014). Quest for the best non-cognitive predictor of academic achievement. *Educational Psychology, 34*(1), 1-8.

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement?. *Educational Psychology, 34*(1), 9-28.

Stenmark, J. K. (1992). *Mathematics assessment: Myths, models, good questions and practical suggestions*. Reston, VA: NCTM.

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.

Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and 'Getting it'for the white working class. *Research Papers in Education, 29*(2), 131-171.

Stringfield, S.C. and Slavin, R.E. (1992). A hierarchical longitudinal model for elementary school effects, in Creemers, B.P.M. and Reezigt, G.J. (Eds), *Evaluation of Effectiveness*. Groningen: ICO.

Stringfield, S., & Teddlie, C. (2011). School effectiveness research, 1932–2008, including a call for future research. In C. Day (Ed.), *International handbook of teacher and school development* (pp. 379–388). London: Routledge/Falmer.

ANASTASIA PANAYIOTOU

Sucharita, V. (2013). Improving Quality in Education – Dynamic Approaches to School Improvement. *Journal of Educational Planning and Administration, 27*(1), 128-130.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Sweller, J. (2011). Cognitive load theory. The psychology of learning and motivation: *Cognition in education*, *55*, 37-76.

Taba, H., Tylor, R., & Smith, B.O. (1998). *The predictive validity of student evaluations in the identification of meritorious teachers. Research Report*. Washington, DC: John Jones School District.

Tabachnick, B. G. & Fidell, L.S. (2011). Multivariate analysis of variance (MANOVA). In M. Lovric (ed). *International Encyclopedia of Statistical Science* (pp. 902-904). Berlin Heidelberg: Springer.

Teddlie, C., & Liu, S. (2008). Examining teacher effectiveness within differentially effective primary schools in the People's Republic of China. *School Effectiveness and School Improvement, 19*(4), 387-407.

Teddlie, C. & Reynolds, D., (2000). *The International Handbook of School Effectiveness Research.* London: Falmer Press.

Teddlie, C., Creemers, B.P.M., Kyriakides, L., Muijs, D. & Yu, F. (2006). The international system for Teacher Observation and Feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation*, *12*(6), 561-582.

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, 1-9.

Thomas, S. L., & Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in higher education*, *42*(5), 517-540.

Thrupp, M. (2001). Sociological and political concerns about school effectiveness research: Time for a new research agenda. *School Effectiveness and School Improvement*, *12*, 7 – 40.

Tomlinson, C. A. (2005). Grading and differentiation: paradox or good practice?. *Theory into practice*, *44*(3), 262-269.

ANASTASIA PANAYIOTOU

Townsend, T. (2007). 20 years of ICSEI: The impact of school effectiveness and school improvement on school reform. In *International handbook of school effectiveness and improvement*, ed. T. Townsend, 3–26. Dordrecht: Springer.

Travers, R. M. W. (1982). *Essentials Off Learning: The New Cognitive Learning for Students of Education.* MacMillan Publishing Company.

Tunstall, P., & Gsipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389-404.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E. J. (2015). *Model Comparison and the Principle.* The Oxford Handbook of Computational and Mathematical Psychology, 300.

Van de Grift, W. (1990). Educational leadership and academic achievement in secondary education. *School Effectiveness and School Improvement, 1*(1), 26 – 40.

Van Der Werf, G., Creemers, B.P.M. De Jong, R. & Klaver, E. (2000). Evaluation of school improvement through an educational effectiveness model: The case of Indonesia's PEQIP project. *Comparative Education Review, 44*(3), 329-355.

Van de Werfhorst, H. G., & Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual review of sociology, 36*, 407-428.

Vanhoof, J., & Van Petegem, P. (2007). Matching internal and external evaluation in an era of accountability and school development: Lessons from a Flemish perspective. *Studies in Educational Evaluation, 33*(2), 101-119.

Van Houtte, M., & Van Maele, D. (2011). The black box revelation: in search of conceptual clarity regarding climate and culture in school effectiveness research. *Oxford Review of Education, 37*(4), 505-524.

Vennemann, M. & Wendt, H. (2012). *The ADDITION Achievement results: Cross-sectional and longitudinal results of mathematics and science in Cyprus.* (ESF project: Establishing a knowledge base for quality in education: testing a dynamic theory for education 08-ECRP-012). Dortmund: IFS.

Von Glasersfeld, E. (1995). A constructivist approach to teaching. In L. P. Steffe & J. Gale (Eds.), *Constructivism in education* (pp. 3–15). Hillsdale, NJ: Lawrence Erlbaum Associates.

ANASTASIA PANAYIOTOU

Walberg, H.J. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*(8), 19-27.

Walberg H.J. (1986). Syntheses of research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 214–229). New York: Macmillan.

Watkins, D. (1994). Student evaluations of teaching effectiveness: A cross-cultural perspective. *Research in Higher Education*, 35, 251-266.

Weber, G. (1971). *Inner city children can be taught to read: four successful schools. Washington,* DC: Council for Basic Education.

Whitty, G., & Willmott, E. (1991). Competence-based teacher education: approaches and issues. *Cambridge Journal of Education, 21*(3), 309–318.

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles Policy and Practice, 11*(1), 49-65.

Williams, M. (1989). A developmental view of classroom observations. *ELT journal, 43*(2), 85-91.

Winne, P. H. (2005). A perspective on state-of-the-art research on self-regulated learning. *Instructional Science, 33*, 559–565.

Wubbels, Th., & Brekelmans, M. (1998). The teacher factor in the social climate of the classroom. In B.J. Fraser & K.G. Tobin (Eds.), *Intemational handbook of Science education, part one* (pp. 565-580). London: Kluwer Academic Publishers.

Wylie, C., & Lyon, C. (2009). *What schools and districts need to know to support teachers' use of formative assessment*. Teachers College Record.

Yang, Y. (2003). Dimensions of socio-economic status and their relationship to mathematics and science achievement at individual and collective levels. *Scandinavian Journal of Educational Research, 47*(1), 21 ‐ 41.

Yair, G. (1997). When classrooms matter: Implications of between-classroom variability for educational policy in Israel. *Assessment in Education, 4*(2), 225–248.

ANASTASIA PANAYIOTOU

APPENDIX A: Student Questionnaire

Dear student,

The [NAME OF INSTITUTION] carries out a lot of research including research on education.

We are conducting a study on students in Grade 4 and would like to know your opinion about the teaching of **Mathematics and Science** in your classroom.

The answers you give will not be shown to your teachers, anyone else in your school or your parents.

We are giving each student a special number so you do not need to write your name on the questionnaire.

Please answer **all** of the questions. To answer the questions, please circle a number on each line.

Please ask the interviewer if you do not understand what to do.

ANASTASIA PANAYIOTOU

PART A

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if this **never** happens in your class

2: if this **rarely** happens in your class

3: if this **sometimes** happens in your class

4: if this **often** happens in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q1.	In Mathematics and Science, we start the lesson with things that are easy to understand. As the lesson goes on what we cover is more difficult.	1	2	3	4	5
Q2.	The teacher gives us exercises at the beginning of the lesson to check what we have learnt from the previous lesson.	1	2	3	4	5
Q3.	At the beginning of the lesson, the teacher starts with what we covered in the previous lessons.	1	2	3	4	5
Q4.	My teacher helps us to understand how different activities (such as exercises, subject matter) during a lesson are related to each other.	1	2	3	4	5
Q5.	A few days before the test, my teacher gives us similar exercises to those that will be in the test.	1	2	3	4	5
Q6.	My teacher tells my parents how good I am compared to my classmates when they visit her/him (or in my school report).	1	2	3	4	5
Q7.	When the teacher is teaching, I always know what part of the lesson (beginning, middle, end) we are in.	1	2	3	4	5

ANASTASIA PANAYIOTOU

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if the situation described **never** happens in your class

2: if the situation described happens **rarely** in your class

3: if the situation described happens **sometimes** in your class

4: if the situation described happens **often** in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q8.	When doing an activity in Mathematics and Science I know why I am doing it.	1	2	3	4	5
Q9.	When we go over our homework, our teacher finds what we had problems with and helps us to overcome these difficulties.	1	2	3	4	5
Q10.	Our teacher has good ways of explaining how the new things we are learning are related to things we already know.	1	2	3	4	5
Q11.	At the end of each lesson, the teacher gives us exercises on what we have just learned.	1	2	3	4	5
Q12.	During the lesson our teacher often covers the same things that we have already learned or done exercises in.	1	2	3	4	5
Q13.	The teacher immediately comes to help me when I have problems doing an activity.	1	2	3	4	5
Q14.	The teacher gives more exercises to some pupils than the rest of the class.	1	2	3	4	5
Q15.	The teacher gives some pupils different exercises to do than the rest of the class.	1	2	3	4	5

ANASTASIA PANAYIOTOU

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if the situation described **never** happens in your class

2: if the situation described happens **rarely** in your class

3: if the situation described happens **sometimes** in your class

4: if the situation described happens **often** in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q16.	The teacher gives all pupils the chance to take part in the lesson.	1	2	3	4	5
Q17.	Our teacher encourages us to work together with our classmates during Mathematics and Science lessons.	1	2	3	4	5
Q18.	Some pupils in my classroom work together when our teacher asks us but some pupils do not.	1	2	3	4	5
Q19.	Our teacher makes us feel that we can ask him/her for help or advice if we need it.	1	2	3	4	5
Q20.	Our teacher encourages us to ask questions if there is something that we do not understand during the lesson.	1	2	3	4	5
Q21.	During the lesson, our teacher encourages and tells us that we are doing good work (i.e. she/he says to us "well done").	1	2	3	4	5
Q22.	When we are working in teams, our teacher encourages competition between teams. (If you do not work in teams, please circle the number 1).	1	2	3	4	5

ANASTASIA PANAYIOTOU

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if the situation described **never** happens in your class

2: if the situation described happens **rarely** in your class

3: if the situation described happens **sometimes** in your class

4: if the situation described happens **often** in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q23.	In Mathematics and Science lessons, some of my classmates hide their work and answers so that none of the other pupils can see it.	1	2	3	4	5
Q24.	When a pupil gives a wrong answer the teacher helps her/him to understand her/his mistake and find the correct answer.	1	2	3	4	5
Q25.	When the teacher asks us a question about the lesson he/she asks us for the answer but does not ask us to explain how we worked out the answer.	1	2	3	4	5
Q26.	When one of the pupils in the class is having difficulties with the lesson, our teacher goes to help him/her straight away.	1	2	3	4	5
Q27.	There are some pupils in the classroom that tease some of their classmates during Mathematics and Science lessons.	1	2	3	4	5
Q28.	I know that if I break a class rule I will be punished.	1	2	3	4	5
Q29.	The teacher has to stop teaching the class because one of the pupils is being naughty	1	2	3	4	5

ANASTASIA PANAYIOTOU

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if the situation described **never** happens in your class

2: if the situation described happens **rarely** in your class

3: if the situation described happens **sometimes** in your class

4: if the situation described happens **often** in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q30.	When a pupil gives a wrong answer in Mathematics and Science class some of the other children in the class make fun of her/him.	1	2	3	4	5
Q31.	Our teacher keeps on teaching us even though it is break-time or the lesson is supposed to be over.	1	2	3	4	5
Q32.	When I finish a task before my classmates my teacher immediately gives me something else to do.	1	2	3	4	5
Q33.	When the teacher talks to a pupil after they have been naughty, sometimes after a while, that pupil will be naughty again.	1	2	3	4	5
Q34.	We spend time at the end of the lesson to go over what we have just learned.	1	2	3	4	5
Q35.	There are times we do not have the necessary materials for the lesson to take place (e.g., dienes, unifix, test tubes, thermometers, calculators, rulers)	1	2	3	4	5
Q36.	There are times when I do not have anything to do during a lesson.	1	2	3	4	5

ANASTASIA PANAYIOTOU

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if the situation described **never** happens in your class

2: if the situation described happens **rarely** in your class

3: if the situation described happens **sometimes** in your class

4: if the situation described happens **often** in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q37.	During a Mathematics or Science lesson, our teacher asks us to give our own opinion on a certain issue.	1	2	3	4	5
Q38.	Our teacher asks us questions at the beginning of the lesson to help us remember what we did in the previous lesson.	1	2	3	4	5
Q39.	Our teacher uses words that are hard to understand when he/she asks us a question.	1	2	3	4	5
Q40.	When we do not understand a question, our teacher says it in a different way so we can understand it.	1	2	3	4	5
Q41.	When a pupil gives a wrong answer our teacher gets another pupil to answer the question.	1	2	3	4	5
Q42.	When I give a wrong answer to a question the teacher helps me to understand my mistake and find the correct answer.	1	2	3	4	5
Q43.	Our teacher praises all pupils the same when we answer a question correctly.	1	2	3	4	5

ANASTASIA PANAYIOTOU

After each statement you read there are four numbers. Think carefully and put a circle around the number that most fits your opinion:

1: if the situation described **never** happens in your class

2: if the situation described happens **rarely** in your class

3: if the situation described happens **sometimes** in your class

4: if the situation described happens **often** in your class

		Never	Rarely	Sometimes	Often	Almost Always
Q44.	When we have problem solving exercises and tasks in Mathematics and Science lessons, our teacher helps us by showing us easy ways or tricks to solve the exercises or tasks.	1	2	3	4	5
Q45.	Our teacher lets us use our own easy ways or tricks to solve the exercises or tasks we have in Mathematics and Science.	1	2	3	4	5
Q46.	In Mathematics and Science lessons, our teacher teaches us ways or tricks that can be used in different lessons.	1	2	3	4	5
Q47.	Our teacher encourages us to find ways or tricks to solve the exercises or work s/he gives us.	1	2	3	4	5
Q48.	I am there when my teacher talks to my parents for my progress.	1	2	3	4	5
Q49.	When we are having a test I finish up within the time given to us.	1	2	3	4	5

ANASTASIA PANAYIOTOU

PART B

In this part there are some statements. For each statement circle the answer that shows what usually happens in your class during Mathematics and Science lessons.

We have tests

A. Every week

B. Every two weeks

C. Every month

D. Every term

E. Never

The teacher gives corrected tests back to us

A. Within a week

B. Within two weeks

C. Within three weeks

D. In a month or even longer

E. S/he never returns them.

The teacher explains to us what s/he expects us to learn from the Mathematics and Science lessons. This happens:

A. in every lesson

B. in most of the lessons

C. only sometimes

D. very rarely

E. never.

ANASTASIA PANAYIOTOU

When no student raises his/her hand to answer a question, the teacher usually (please choose one answer)

A. answers the question and moves to something else

B. repeats the question using the same words

C. restates the question using simpler words

D. asks an easier question

E. gives us hints or clues to help us answer the question.

Further below, write down any comments you want to make about the questionnaire and about the teaching of mathematics and science in your classroom.

....

....

....

....

....

....

....

....

....

....

....

....

....

....

Thank you for your cooperation

ANASTASIA PANAYIOTOU

APPENDIX B: Student Questionnaire Specification Table

	Dimensions				
Teacher Factors	*Frequency*	*Focus*	*Stage*	*Quality*	*Differentiation*
Orientation				8	
Structuring	3	10	2, 34, 38	1, 4, 7	
Application			11, 12	26	13, 14, 15, 32
Management of Time	31, 35, 36	*Not applicable (N/A)*			
Questioning	25, 39			24, 37, 40, 41, 42	43
Modeling	44, 47			45, 46	
Classroom as a learning Environment / Teacher – Student Interaction	16, 17			19, 20, 21, 22	
Classroom as a learning Environment / Dealing with Misbehaviour	29, 18	28		23, 27, 33 , 30	
Assessment	50, 51			5, 6, 9, 48, 49	

ANASTASIA PANAYIOTOU

APPENDIX C: Teacher Questionnaire

Dear Teacher,

The University of Cyprus is carrying out research on how children get on in Grade 4 of primary school. This research is being carried out in a number of schools around the country and in other European countries. We would appreciate it if you could find the time to complete this questionnaire, it should take no more than 35 minutes. Your views are very important, as they will help develop policies to assist students, parents and school staff. All the information you give will be strictly **confidential** and will be used for research purposes only.

This study aims in investigating teachers' opinions on their school's policy. This study mainly examines the policy developed by your school with respect to the following aspects of teaching:

A. Making good use of teaching time

As far as the use of the teaching time is concerned issues related to management of time, student absenteeism, teacher absenteeism, homework assignment, school time-table scheduling and teaching time spent on extra-curriculum activities (e.g., practice time for school events) are examined.

B. Provision of learning opportunities:

The school policy is examined in relation to the achievement of specific goals set by the school, use of visual material and technological equipment in teaching, dealing with students that have educational needs (e.g., gifted children, children with learning difficulties, children with special interests) as well as the long-term planning of teaching by the teachers.

C. Quality of teaching:

The school policy is examined in relation to the following factors concerned with the teacher behaviour in the classroom: Student evaluation, structuring, orientation of students in achieving specific goals, application exercises, posing and using questions in teaching, use of learning strategies, time management, and classroom as a learning environment.

Your views about the policy on the broader learning environment of your school are also examined.

Four aspects of the **School Learning Environment (SLE)** are taken into account:

- policy on student behavior outside the classroom
- teacher collaboration
- relations with parents and the wider school community
- use of educational resources

Thank you very much for your help.

ANASTASIA PANAYIOTOU

PART A: ABOUT YOU

Put a √ in the appropriate box or fill where necessary:

Q1. Are you male or female?

Male……….□ 1 Female……….□ 2

Q2. What is your teaching position in this school?

Teacher……….□ 1 Deputy Head Teacher/Principal ……….□ 2 Head Teacher/Principal ……….□ 3

Q3. How many years have you been teaching at primary school level? (Please count this school year and exclude career breaks)

(a) in this school……………….._______years
(b) in other primary schools……_______years
(c) Total…………………………_______years

PART B: THE FORMATION OF SCHOOL POLICY AND THE LEARNING ENVIRONMENT OF THE SCHOOL

Part B refers to statements concerned with practices that may occur in your school. Please circle a number from 1-4 on each line to show the extent to which you agree with the statements describing what happens in your school. After reading carefully each statement circle the number:

1: if you **strongly disagree** with the statement

2: if you **disagree** with the statement

3: if you **agree** with the statement

4: if you **strongly agree** with the statement

ANASTASIA PANAYIOTOU

		Strongly Disagree	Disagree	Agree	Strongly Agree
Q4.	At staff meetings in our school we discuss and take decisions on issues concerned with:				
	a. Making good use of teaching time	1	2	3	4
	b. Provision of extra learning opportunities in addition to those offered by the formal curriculum (e.g., extra-curricular activities, festivals, fairs, school trips, clubs)	1	2	3	4
	c. Methods to effectively teach students (e.g., structuring lessons, questioning, application, student assessment etc.)	1	2	3	4
	d. Teacher's role during break time	1	2	3	4
	e. Developing trust between teachers and children	1	2	3	4
Q5.	My school keeps systematic records concerned with:				
	a. Student absenteeism	1	2	3	4
	b. Teacher absenteeism	1	2	3	4
	c. Special educational needs of students	1	2	3	4
	d. Long-term planning by teachers	1	2	3	4
	e. Organization of trips, visits and other extra-curricular activities not included in the formal curriculum	1	2	3	4
	f. Problems that arise among students during break time	1	2	3	4
	g. The use of educational tools for teaching supplied by the school (e.g., maps, software etc.)	1	2	3	4
Q6.	My school participates in programmes (e.g., Comenius, action research projects, collaboration with other schools, pilot initiatives) that aim at:				
	a. Making good use of teaching time	1	2	3	4
	b. Providing learning opportunities beyond the ones offered by the formal curriculum	1	2	3	4
	c. Improving the quality of teaching	1	2	3	4
Q7.	When designing the school-timetable we take into account that enough time should be provided for students and/or teachers to move between classrooms	1	2	3	4

ANASTASIA PANAYIOTOU

		Strongly Disagree	Disagree	Agree	Strongly Agree
Q8.	I feel that I am positively influenced by the staff meetings in relation to the following:				
	a. Management of teaching time	1	2	3	4
	b. Dealing with student absenteeism	1	2	3	4
	c. Homework	1	2	3	4
	d. Making good use of teaching time that is spent on activities not included in the formal curriculum (e.g. rehearsals)	1	2	3	4
	e. Use of visual aids and technological equipment in teaching (e.g. overhead projector, computer)	1	2	3	4
	f. Dealing with students that have special educational needs (e.g., gifted children, children with learning disabilities, children with special interests)	1	2	3	4
	g. Long-term planning of teaching	1	2	3	4
	h. Interaction with students during break time	1	2	3	4
	i. Student evaluation	1	2	3	4
	j. Structuring of lessons	1	2	3	4
	k. Student orientation (i.e., helping students to understand why a unit is taught)	1	2	3	4
	l. Using exercises to help students apply their learning (i.e., giving them tasks which apply the concepts taught to a situation in everyday life)	1	2	3	4
	m. Asking questions and making good use of them	1	2	3	4
	n. Strategies for learning	1	2	3	4
	o. The learning environment of the classroom (e.g., promoting interaction among students)	1	2	3	4
Q9.	My school takes into consideration the professional needs of each teacher and does not expect each teacher to implement the school policy for teaching in the same way (the school policy being what is decided at the school level regarding quantity and quality of education, providing learning opportunities ...)	1	2	3	4

ANASTASIA PANAYIOTOU

		Strongly Disagree	Disagree	Agree	Strongly Agree
Q10.	We take into account research findings (e.g., recently published articles in scientific journals, results of research studies) in developing the school policy on teaching	1	2	3	4
Q11.	We take into account research findings when we form a school policy concerned with:				
	a. parental involvement	1	2	3	4
	b. teacher collaboration	1	2	3	4
	c. use of resources for teaching provided by the	1	2	3	4
Q12.	Incentives are provided and/or support is given to teachers to implement the school policy on teaching (e.g., reward teachers who spend extra time with students who were absent from school in order to explain to them the concepts taught during their absenteeism)	1	2	3	4
Q13.	My school encourages teachers to cooperate with the parents of children who struggle academically	1	2	3	4
Q14.	The teachers in my school cooperate with each other by exchanging ideas and material when teaching specific units or series of lessons.	1	2	3	4
Q15.	Discussions at staff meetings help me to improve my practice in:				
	a. Making good use of teaching time	1	2	3	4
	b. Providing learning opportunities to students beyond the ones offered by the formal curriculum	1	2	3	4
	c. My teaching behaviour in the classroom	1	2	3	4
	d. My role during break time	1	2	3	4
	e. Using different educational tools for teaching provided by the school	1	2	3	4
	f. Involving parents in the learning process	1	2	3	4
Q16.	In my school, teachers observe each other teaching as a way to discuss and share opinions on effective teaching	1	2	3	4
Q17.	The teachers in my school participate in educational school-based seminars (e.g., workshops) which deal with **specific** issues that the school faces	1	2	3	4
Q18.	My school has formed a **specific** policy for student behaviour during break time	1	2	3	4

ANASTASIA PANAYIOTOU

		Strongly Disagree	Disagree	Agree	Strongly Agree
Q19.	In my school we share the opinion that break time is an opportunity for teachers to approach and interact with children that face problems which may affect their learning	1	2	3	4
Q20.	In my school, we have taken the decision to organize fun activities during break time that may help students to achieve specific learning goals (e.g., games, dance, sports)	1	2	3	4
Q21.	In parent-teacher meetings organized by the school, the way in which parents can help in dealing with the following issues are discussed:				
	a. Student absenteeism	1	2	3	4
	b. Homework	1	2	3	4
	c. Addressing children's educational needs (e.g., gifted children, children with learning difficulties, children with special interests)	1	2	3	4
	d. Parents providing learning opportunities in the school through activities organized on their own initiative (e.g., educational visits, educational games)	1	2	3	4
Q22.	There is material on notice-boards in the school relevant to:				
	a. Good use of teaching time	1	2	3	4
	b. Provision of learning opportunities beyond the ones provided by the formal curriculum	1	2	3	4
	c. Characteristics of effective teaching	1	2	3	4
	d. The use of different educational tools for teaching provided by the school	1	2	3	4
Q23.	At staff meetings, we usually make decisions on the ways in which parents can be involved in the learning process	1	2	3	4
Q24.	During break time, the teachers spend more time with students who face learning difficulties than with other students	1	2	3	4
Q25.	Parents are often invited to our school to observe teaching so that they are aware of the policy the classroom teacher adopts	1	2	3	4
Q26.	My school has a **clear** policy for parental involvement in the learning process	1	2	3	4
Q27.	In my school, there is an opportunity for different groups/people outside the school to become involved with and cooperate in the **learning**	1	2	3	4

ANASTASIA PANAYIOTOU

	process of (for example, a basketball player of a local team together with teachers teaches different basketball techniques)				

		Strongly Disagree	**Disagree**	**Agree**	**Strongly Agree**
Q28.	Discussions at staff meetings lead to an improvement in the way in which the school offers teachers opportunities for professional development and training	1	2	3	4
Q29.	My school invites specialists in to conduct in-service training for teachers (e.g., an expert that works on developing students' creativity or other types of in-service)	1	2	3	4
Q30.	The management team (principal and deputy heads) organizes in-service seminars for a specific group of teachers when they think it is needed (e.g., newly appointed teachers)	1	2	3	4

PART C: EVALUATION OF SCHOOL POLICY

This section is concerned with the evaluation of school policy. To answer questions in Part C, please circle a number from 1-4 on each line to show the extent to which you agree with each statement describing what happens in your school. After reading carefully each statement circle the number:

1: if you **strongly disagree** with the statement

2: if you **disagree** with the statement

3: if you **agree** with the statement

4: if you **strongly agree** with the statement

		Strongly Disagree	**Disagree**	**Agree**	**Strongly Agree**
Q31.	The principal and/or other members of the school staff observe the way the teaching policy is put into practice and presents the results of their observations to staff	1	2	3	4
Q32.	To evaluate the implementation of the school policy on teaching, we collect information from:				
	a. Teachers	1	2	3	4
	b. Students	1	2	3	4
	c. Parents	1	2	3	4

ANASTASIA PANAYIOTOU

Q33.	Teachers' **capacity** to implement the school policy on teaching (e.g. quantity of education, quality of education, provision of learning opportunities for students)is evaluated within the school	1	2	3	4
Q34.	Information collected during evaluation of the school policy on teaching is used for re-designing the policy or for taking new decisions	1	2	3	4
Q35.	The results of the evaluation of the school policy on teaching are used by the school principal for the summative evaluation of teachers (e.g. career development purposes)	1	2	3	4
Q36.	We evaluate the extent to which student discipline problems during break time are reduced as a result of the school policy	1	2	3	4
Q37.	Aspects of my school's policy on teaching which are considered problematic are evaluated further and/or in more detail	1	2	3	4
Q38.	The principal and/or school staff observe the implementation of the learning environment policy and present the results of their observations to staff	1	2	3	4
Q39.	Aspects of my school's policy concerned with the broader learning environment which are considered problematic are evaluated further and/or in more detail	1	2	3	4
Q40.	Our school identifies the professional development/further education needs of its teachers	1	2	3	4
Q41.	The evaluation of the school policy on the broader learning environment (e.g. further (school climate, students' behaviour outside the classroom, the cooperation and interaction between teachers, the support of teachers and students, the learning objectives ...) is carried out in a way that refers to a single aspect of the policy each time (i.e., evaluation focuses on student behaviour, relations with parents etc. separately)	1	2	3	4
Q42.	Information collected during the evaluation of the policy on the broader learning environment is used for re-designing the policy or for taking new decisions	1	2	3	4
Q43.	School policy evaluation results are useful to pinpoint areas in teaching for which we need support and/or further training	1	2	3	4

In the space provided below, please put down anything you consider important for the development and the evaluation of a school policy concerned with teaching and the learning environment of your school.

____

____

____

Thank you for your cooperation.

ANASTASIA PANAYIOTOU

APPENDIX D: Teacher Questionnaire Specification Table

School Factors	Items of the Teacher Questionnaire per school factor
A. School Policy on teaching	
Quantity of teaching	4A, 6A, 7, 8A, 8B, 8C, 15A, 22A
Provision of learning opportunities	4B, 5D, 5E, 6B, 8D, 8F, 8G, 8H, 15B, 19, 22B, 24
Quality of teaching	4C, 6C, 8I, 8J, 8K, 8L, 8M, 8N,8O, 9, 12,15C, 10, 22C
B. Policy on the school learning environment	
Student behavior outside the classroom	4D, 15D, 20, 18
Collaboration and interaction between teachers	4E, 11B, 16, 14
Partnership policy	11A, 13,15F, 21A, 21B, 21C, 21D, 23, 25, 26, 27
Provision of sufficient learning resources	8E, 11C, 15E, 22D
Relation with Community	17, 28, 29, 30
C. Evaluation of the school policy on teaching	
5A, 5B, 5C, 31, 32A, 32B, 32C, 33, 34, 37,35	
D. Evaluation of the learning environment	
5F, 5G,36, 38, 39, 40, 41, 42, 43	

ANASTASIA PANAYIOTOU

APPENDIX E: Items from the Belgian Schoolloopbanen in het BasisOnderwijs (SIBO) used in the Mathematics Achievement Tests

1 Which number on the number line is three units before 201?

Answer: ______

2 How many sides does an angle have?

A 1
B 2
C 3
D 4
E 5

3 If the day before yesterday was Monday, the day after tomorrow will be:

A Friday
B Thursday
C Wednesday
D Tuesday
E Saturday

4 How many glasses of 2 deciliter can I pour out in a bowl of 1 liter?

Answer: ______

5 At eight I leave for school and I'm there exactly at half past eight. How many minutes was I on the way?

Answer: ______

ANASTASIA PANAYIOTOU

6

Estimate the following sum
202 + 198 + 199

A 400

B 450

C 500

D 550

E 600

7

Subtract the smallest number of the series from the biggest number of the series:

302 600 589 105

A 302

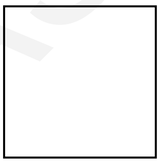
B 600

C 589

D 105

8

The side of this square is 2 cm.



How many cm is the perimeter of the square?

Answer: ______

ANASTASIA PANAYIOTOU

9 How many Sundays are in one month?

A 2 to 3

B 4 to 5

C 6 to 7

D 8 to 9

E 10 to 11

10 60 is the one third of which number?

Answer: ______

11 Grandmother was two months ill. About how many days was grandmother ill?

A 14

B 24

C 36

D 40

E 60

12 Which number follows to continue the series?

140 132 124

Answer: ______

13 A retailer bought 5 boxes of oranges. Each box can fit 19 to 20 kg of oranges. About how many kg of oranges did the retailer buy?

A 39 kg

B 78 kg

C 97 kg

D 120 kg

E 192 kg

ANASTASIA PANAYIOTOU

APPENDIX F: Figures of the Within-country SEM Analyses for Examining the Construct Validity of the Student Questionnaire with Factor Parameter Estimates

Figures 1-6 provide a representation of the models that were best found to fit the data of each country (Model 1). Explanations for the first and second order factors that are shown in the diagram and specifically are also provided:

First Order Factors:

F1: Modeling

F2: Structuring – Quantitative Characteristics

F3: Structuring – Qualitative Characteristics

F4: Application

F6: Management of Time

F7: Classroom as a Learning Environment – Qualitative characteristics: Teacher - Student interaction

F8: Classroom as a Learning Environment – Quantitative Characteristics: Dealing with Misbehaviour

F9: Questioning – Quantitative Characteristics: Raising non-appropriate questions

F10: Assessment

F11: Questioning – Qualitative Characteristics

V1: Orientation

Second Order Factors:

SF1: Quality of Teaching

SF2: Quantity of Teaching (Management of Time, Misbehaviour & Questioning-Quantitative Characteristics)

ANASTASIA PANAYIOTOU

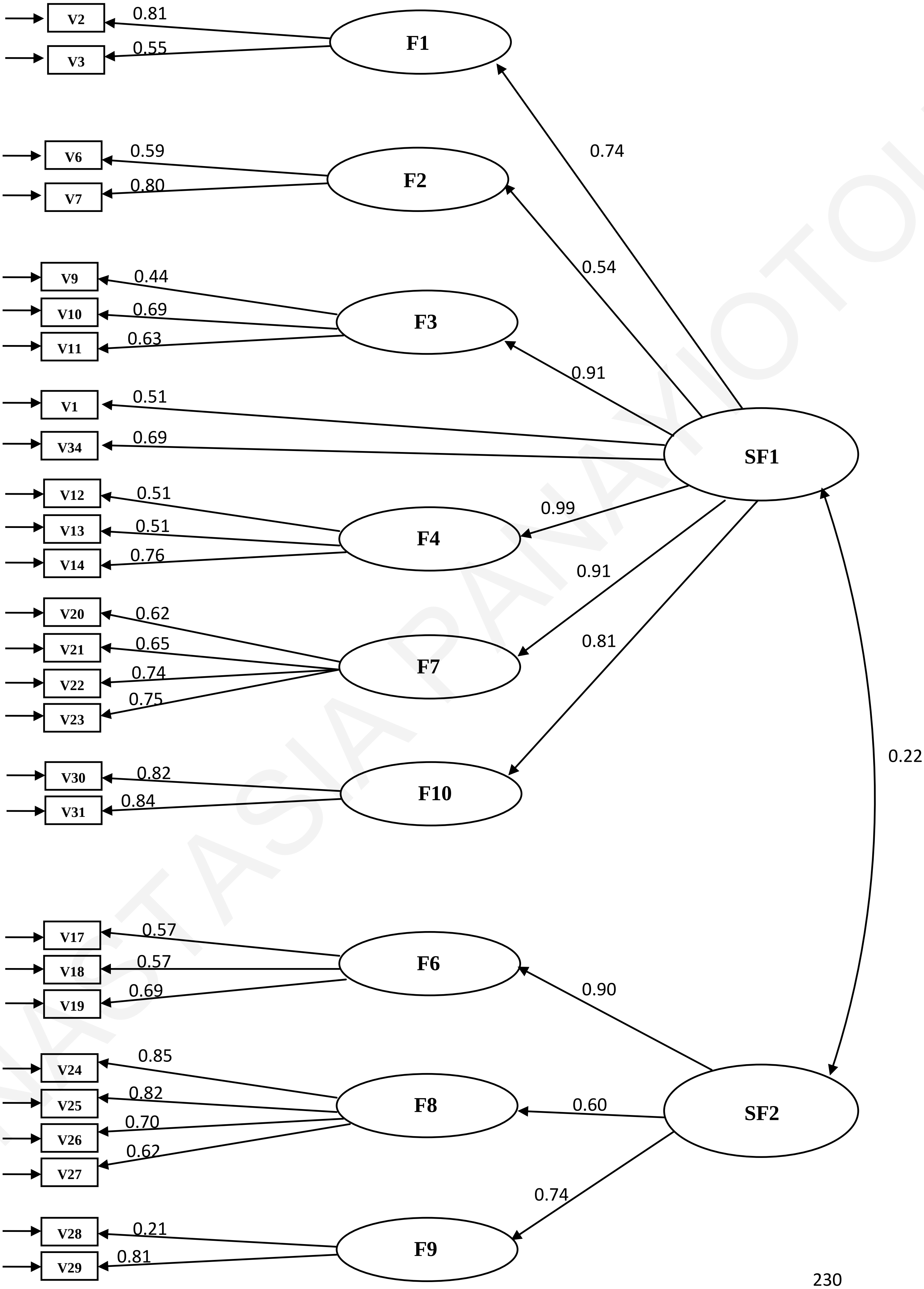


Figure 1 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates for Belgium

ANASTASIA PANAYIOTOU

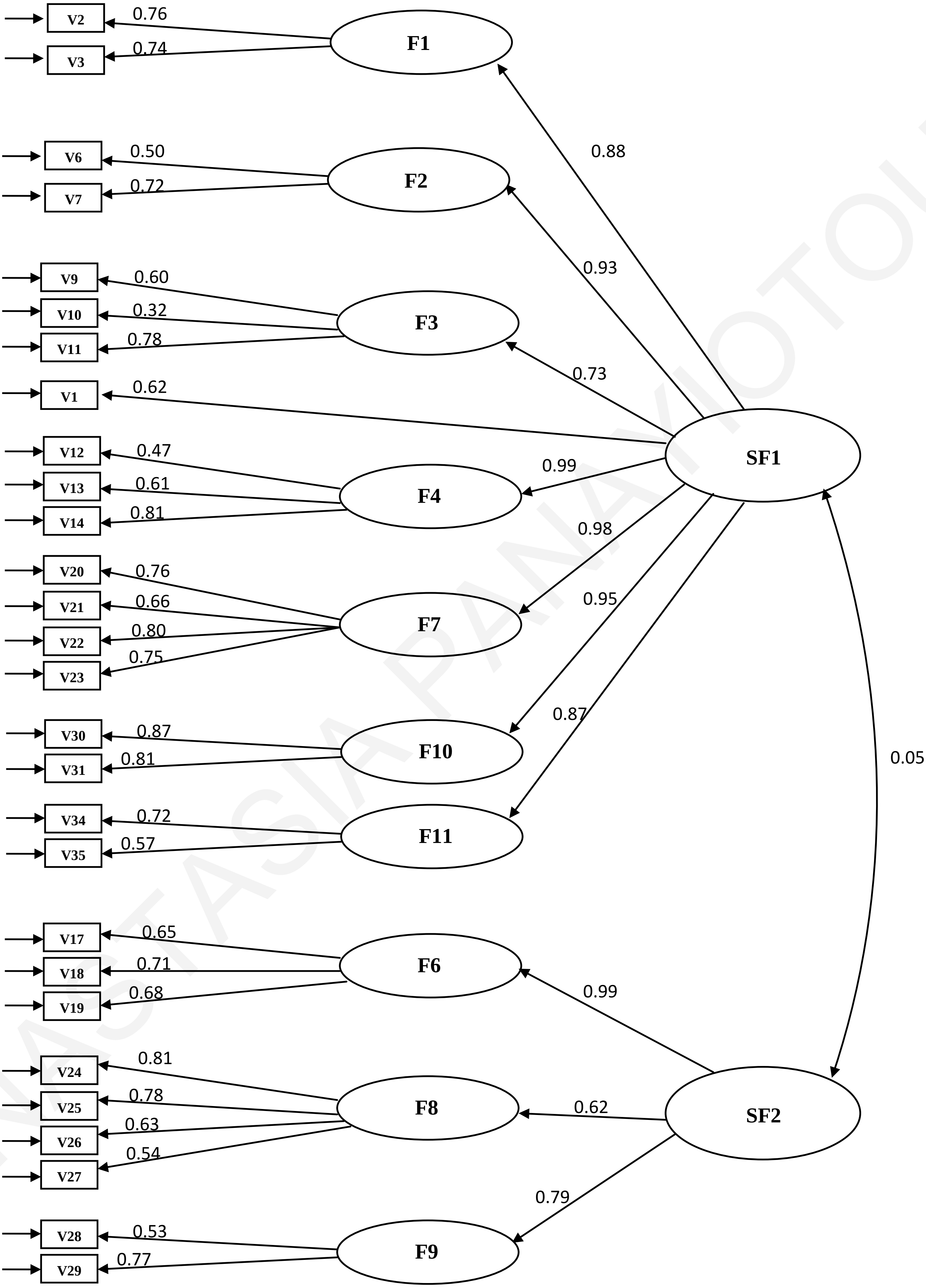


Figure 2 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates for Cyprus

ANASTASIA PANAYIOTOU

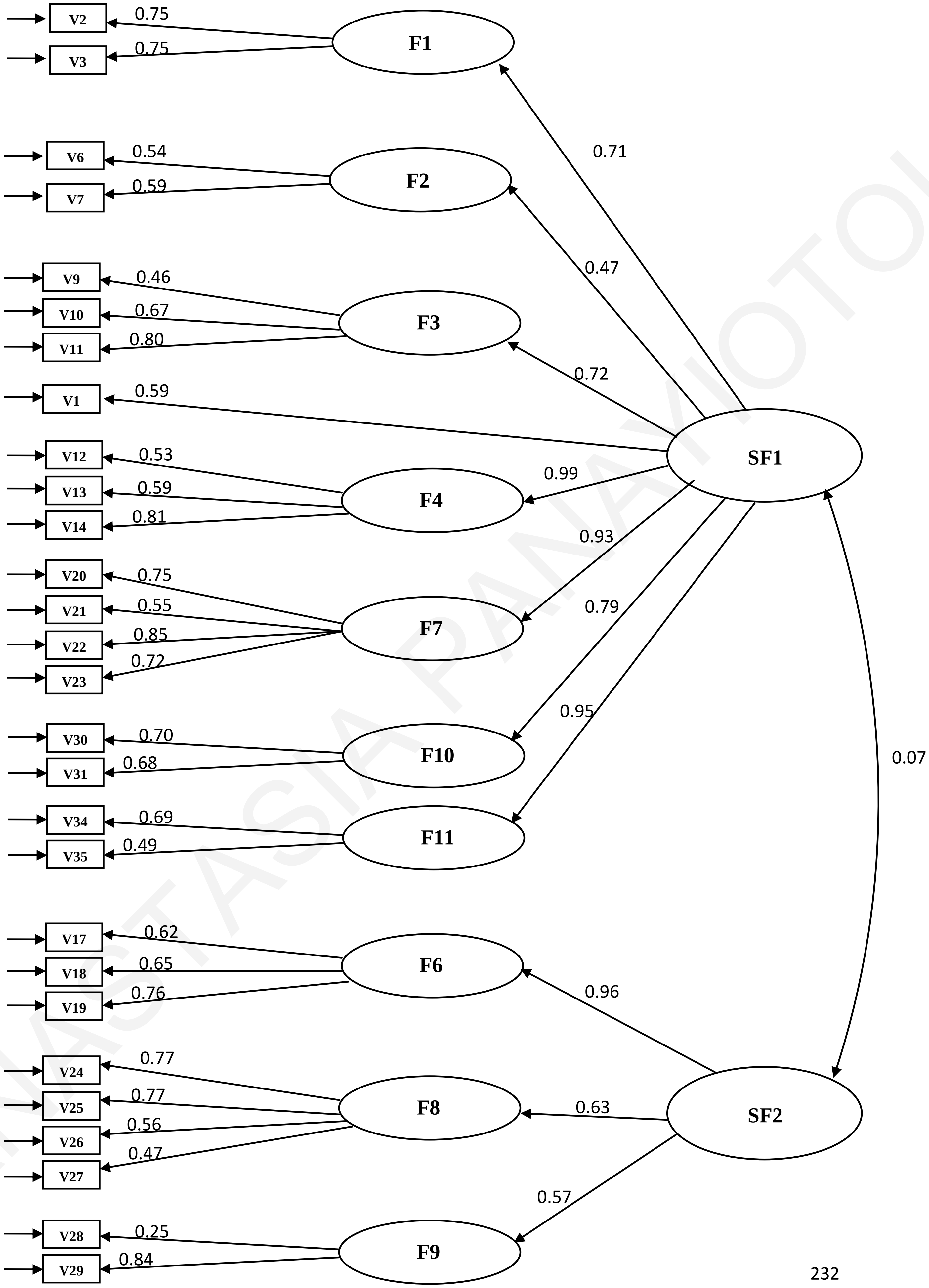


Figure 3 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates for Greece

ANASTASIA PANAYIOTOU

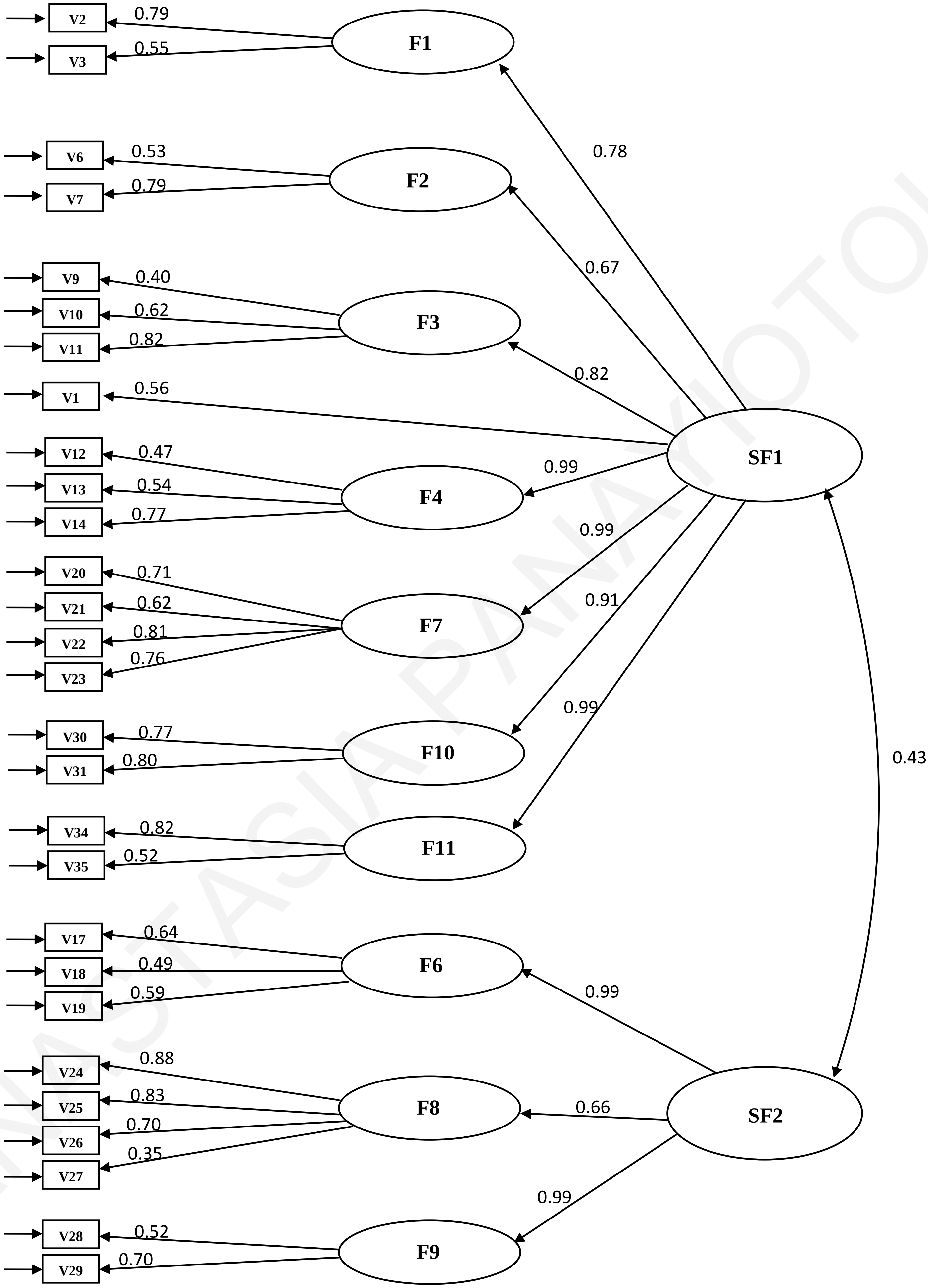


Figure 4 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates for Ireland

ANASTASIA PANAYIOTOU

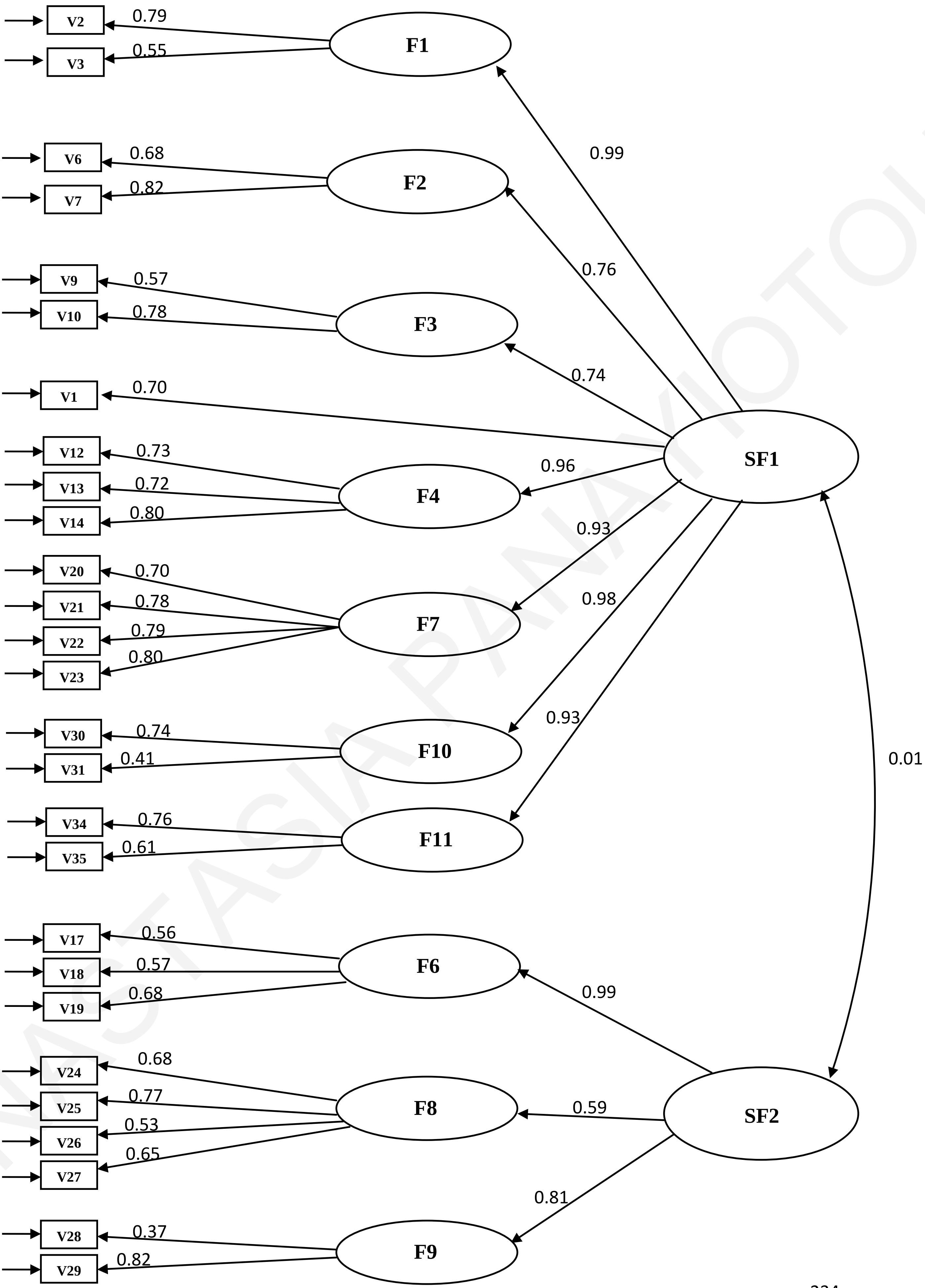


Figure 5 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates for Slovenia

ANASTASIA PANAYIOTOU

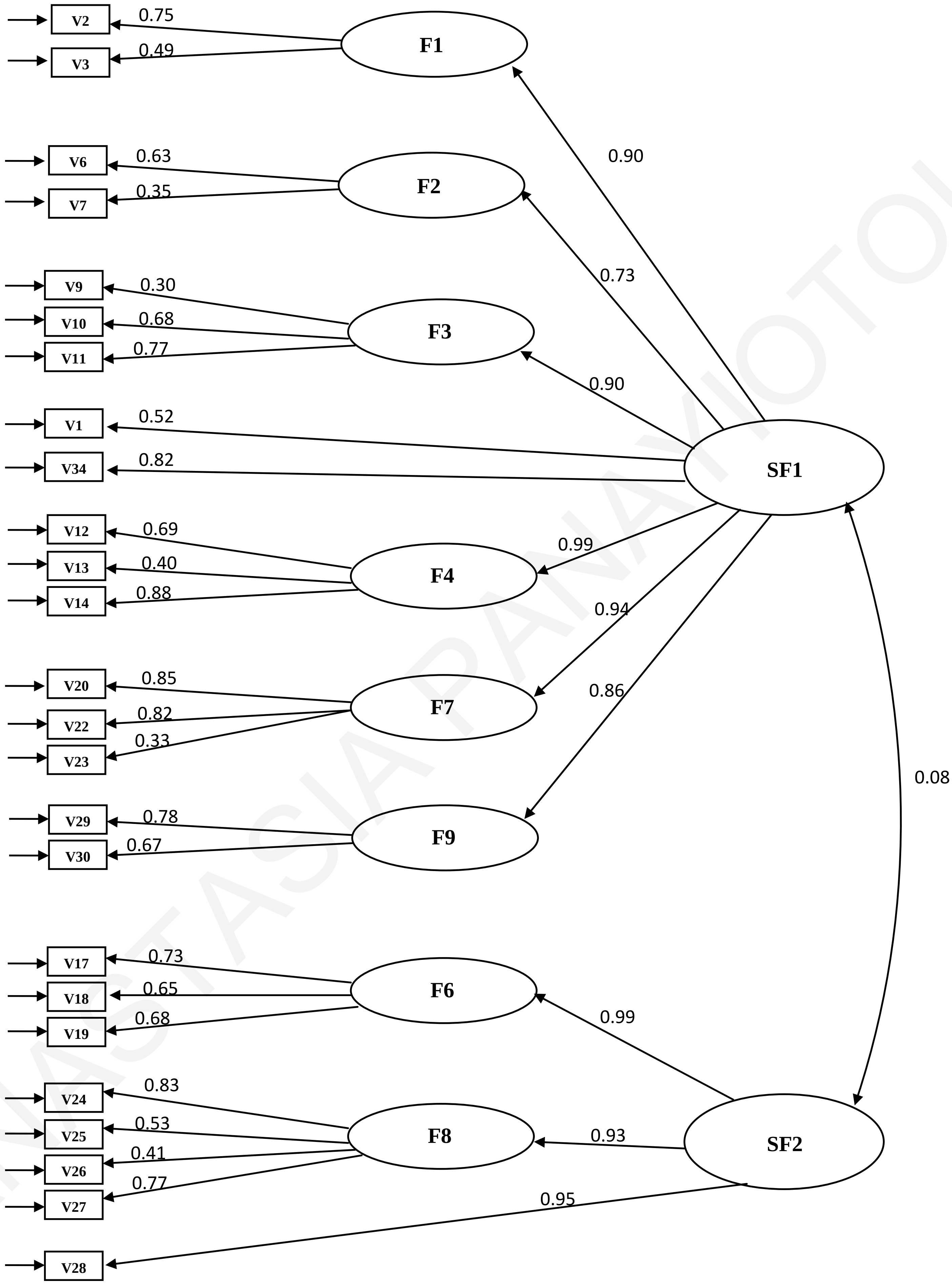


Figure 6 The second-order factor model of the student questionnaire measuring teacher factors with factor parameter estimates for Germany

Note: In Germany item 25 concerned with the frequency dimension of questioning (i.e., "When the teacher asks us a question about the lesson he/she asks us for the answer but does not ask us to explain how we worked out the answer") was not administered and therefore was excluded from the country analysis. Thus, from F9 (Quantitative aspects of questioning) only V28 (item 39) was included in the analysis. Also, items 9 (V33) from assessment and 17(V21) from classroom as a learning environment: teacher-student interaction, were excluded from the analysis in Germany as their loadings were found to be low.The explanation of the figure for Germany is presented below:

First Order Factors:

F1: Modeling
F2: Structuring – Quantitative Characteristics
F3: Structuring – Qualitative Characteristics
F4: Application
F6: Management of Time
F7: Classroom as a Learning Environment – Qualitative characteristics: Teacher - Student interaction
F8: Classroom as a Learning Environment – Quantitative Characteristics: Dealing with Misbehaviour
F9: Questioning – Qualitative Characteristics

V1: Orientation
V28: Questioning: Raising non-appropriate questions
V34: Assessment

Second Order Factors:
SF1: Quality of Teaching
SF2: Quantity of Teaching (Management of Time, Misbehaviour & raising non – appropriate questions)

ANASTASIA PANAYIOTOU

APPENDIX G: Figures of the Within-country SEM Analyses for Examining the Construct Validity of the Teacher Questionnaire with Factor Parameter Estimates

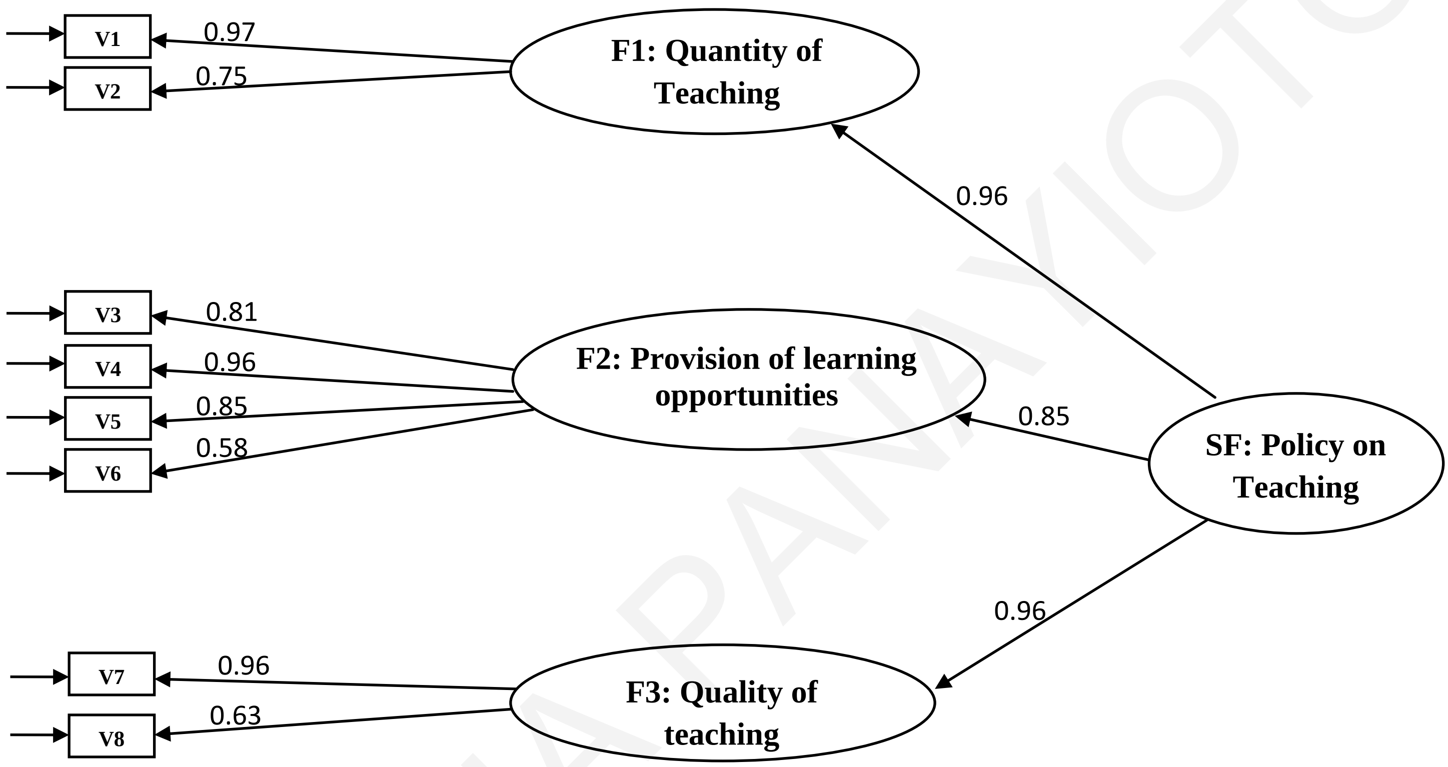


Figure 1.1 The second-order factor model of the teacher questionnaire measuring school factors on policy on teaching with factor parameter estimates for Belgium

ANASTASIA PANAYIOTOU

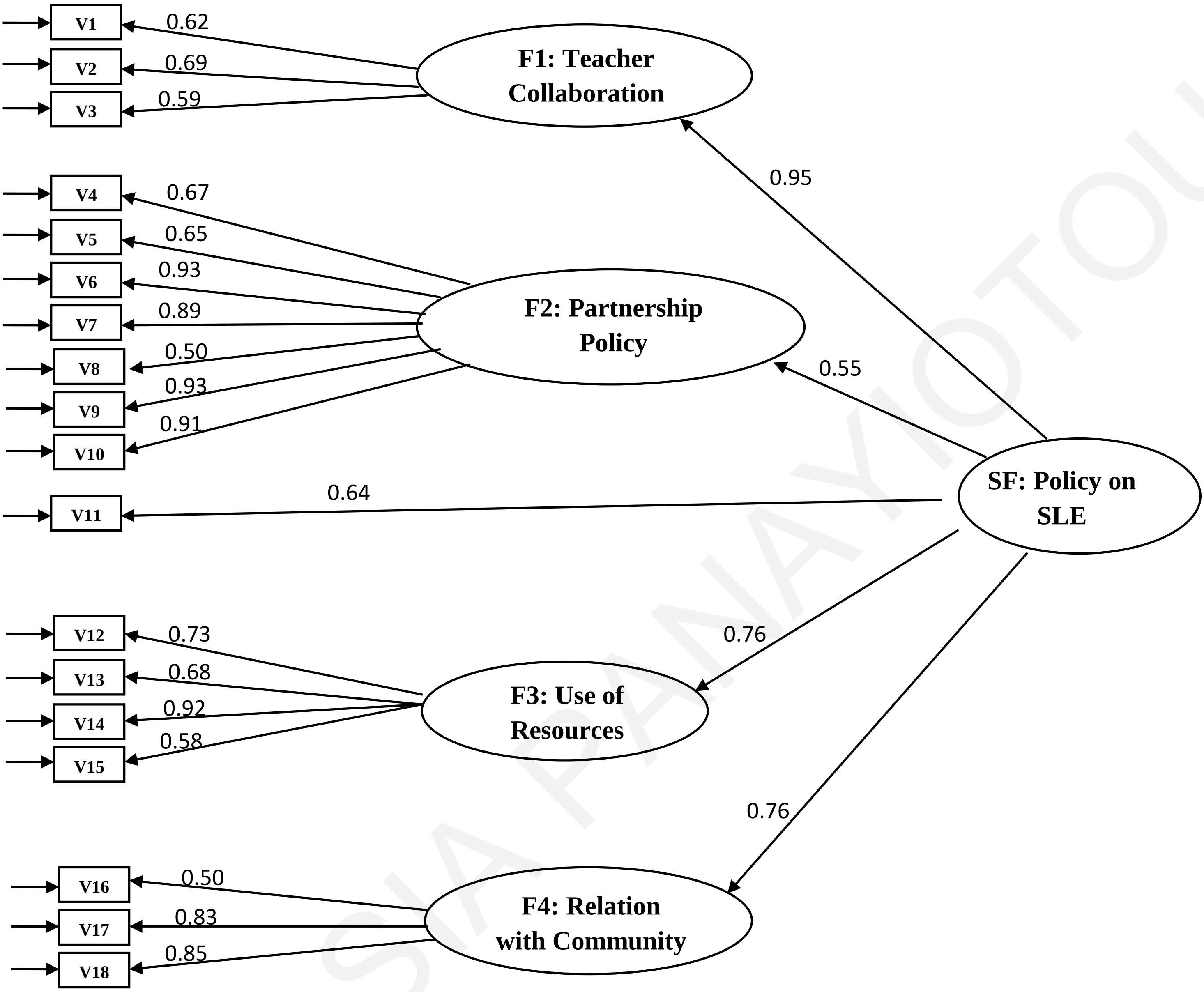


Figure 1.2 The second-order factor model of the teacher questionnaire measuring school factors on the SLE with factor parameter estimates for Belgium

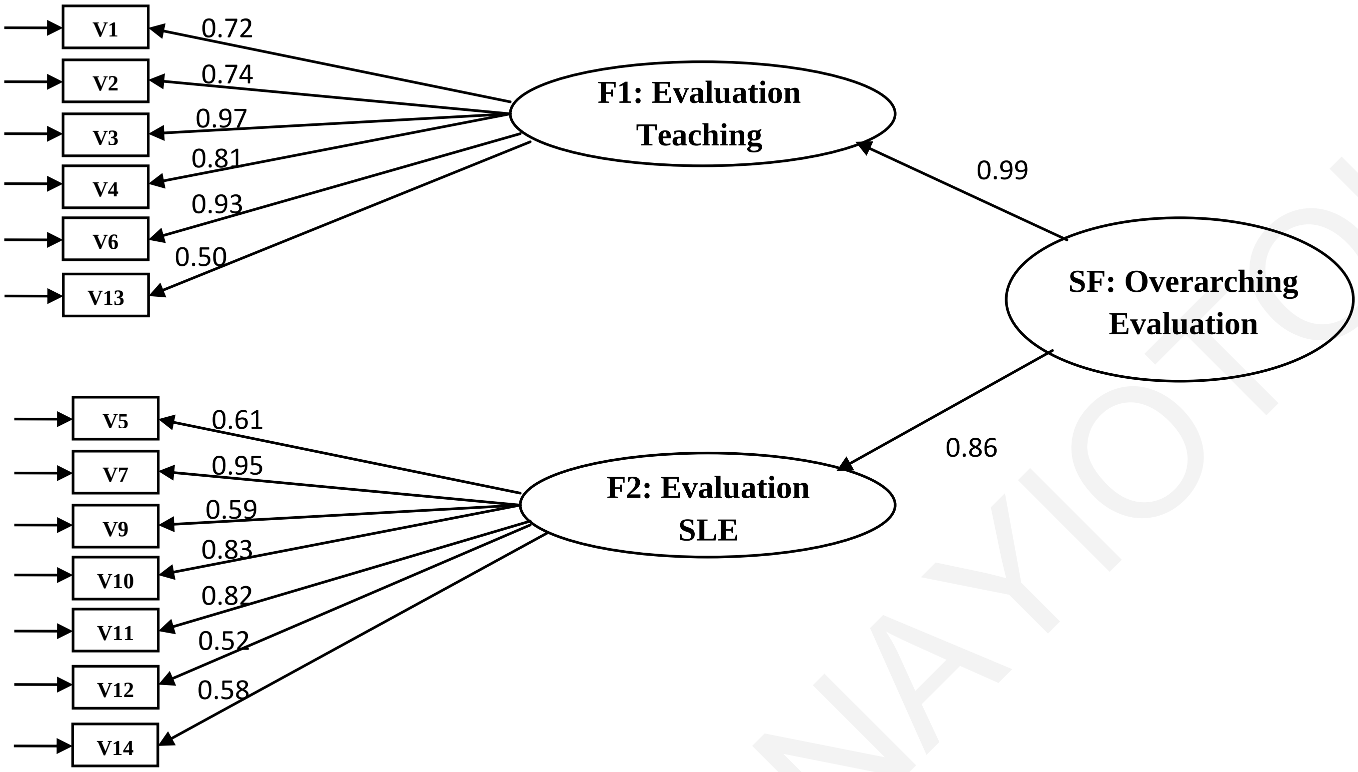


Figure 1.3 The second-order factor model of the teacher questionnaire measuring school factors on evaluation with factor parameter estimates for Belgium

ANASTASIA PANAYIOTOU

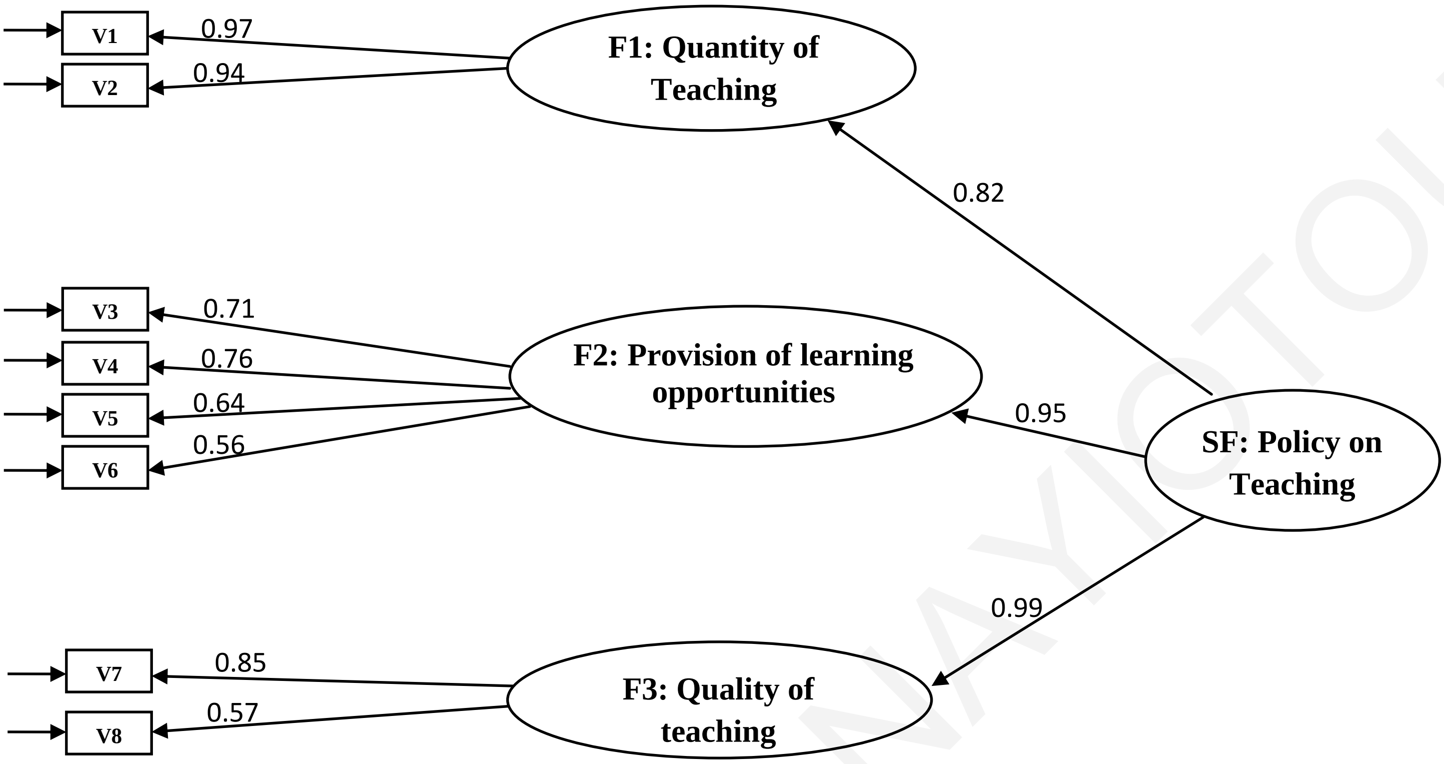


Figure 2.1 The second-order factor model of the teacher questionnaire measuring school factors on policy on teaching with factor parameter estimates for Cyprus

ANASTASIA PANAYIOTOU

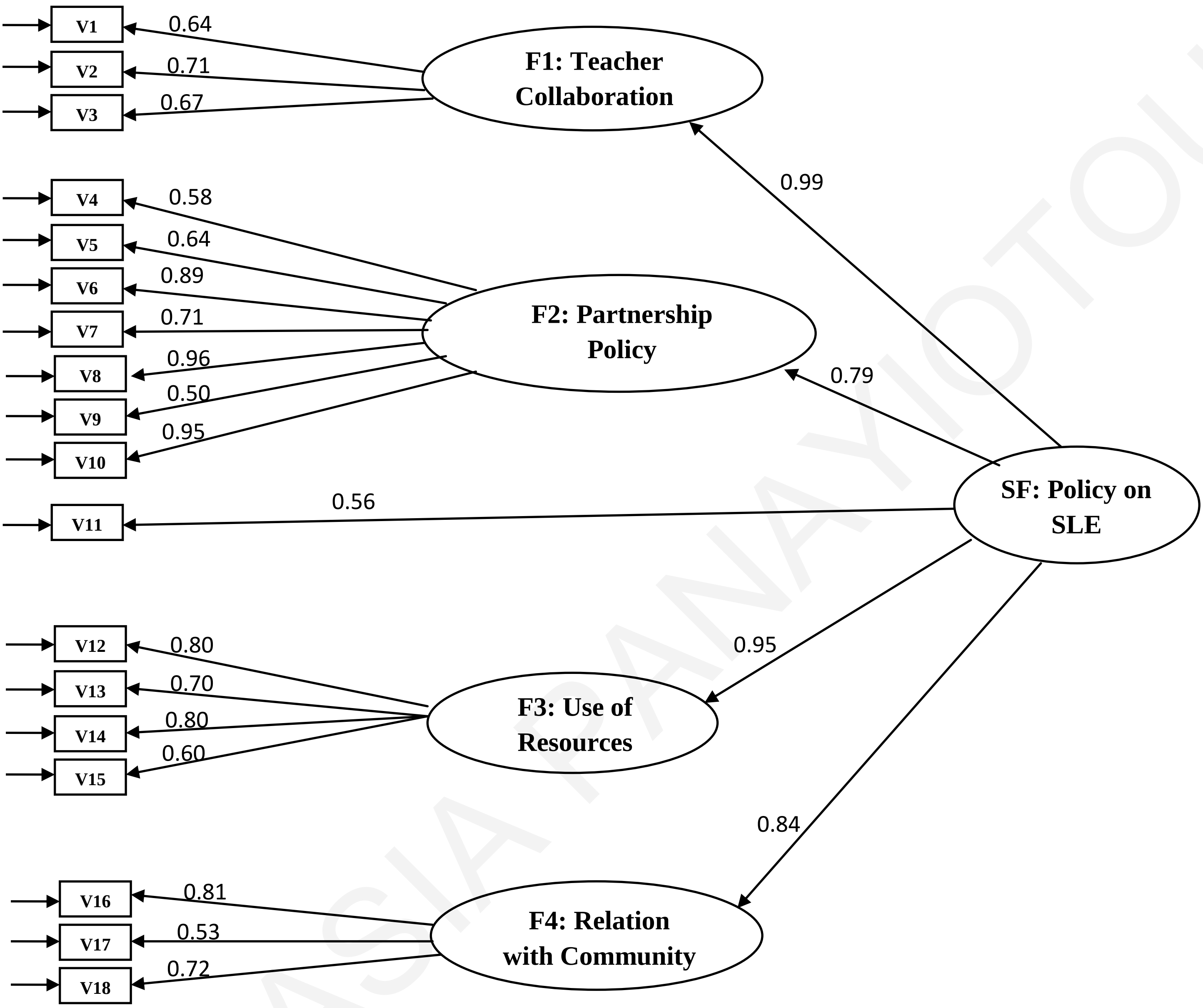


Figure 2.2 The second-order factor model of the teacher questionnaire measuring school factors on the SLE with factor parameter estimates for Cyprus

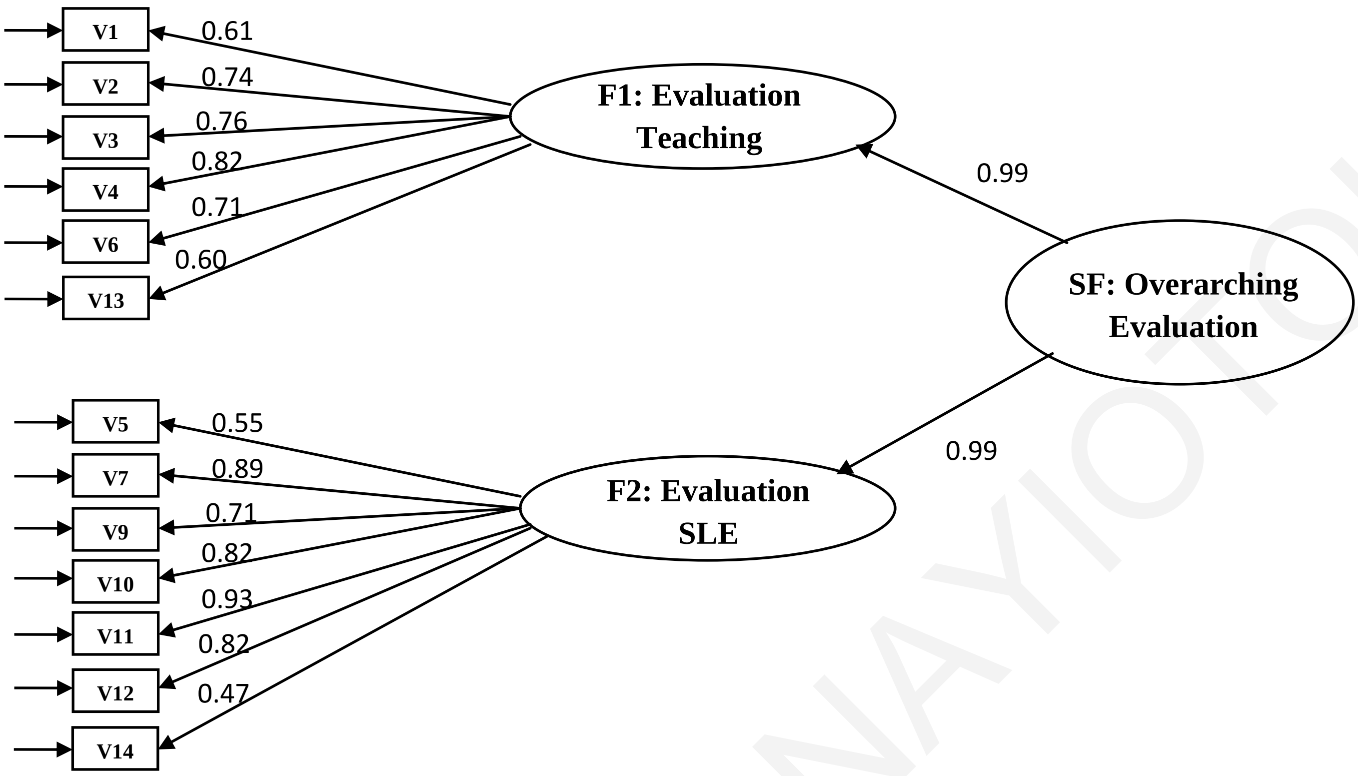


Figure 2.3 The second-order factor model of the teacher questionnaire measuring school factors on evaluation with factor parameter estimates for Cyprus

ANASTASIA PANAYIOTOU

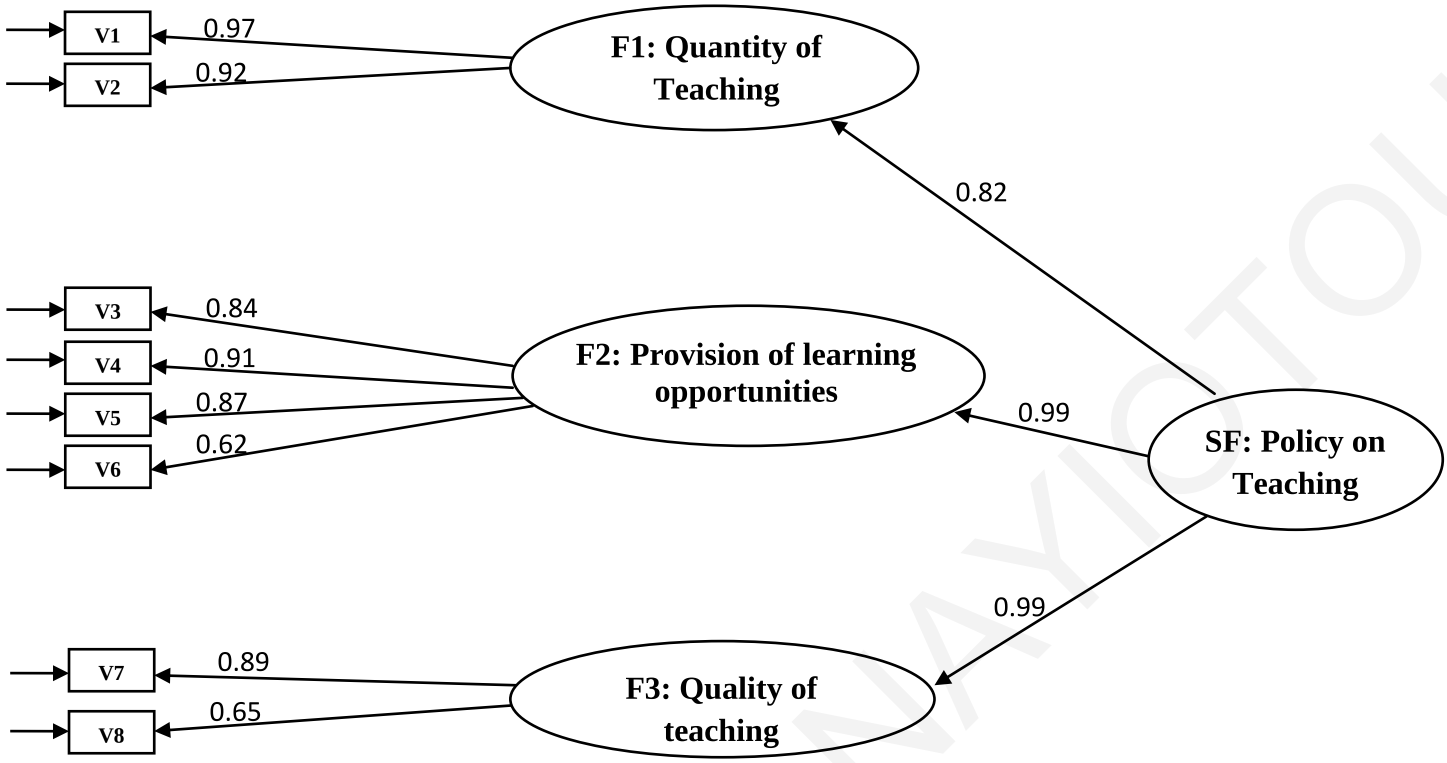


Figure 3.1 The second-order factor model of the teacher questionnaire measuring school factors on policy on teaching with factor parameter estimates for Germany

ANASTASIA PANAYIOTOU

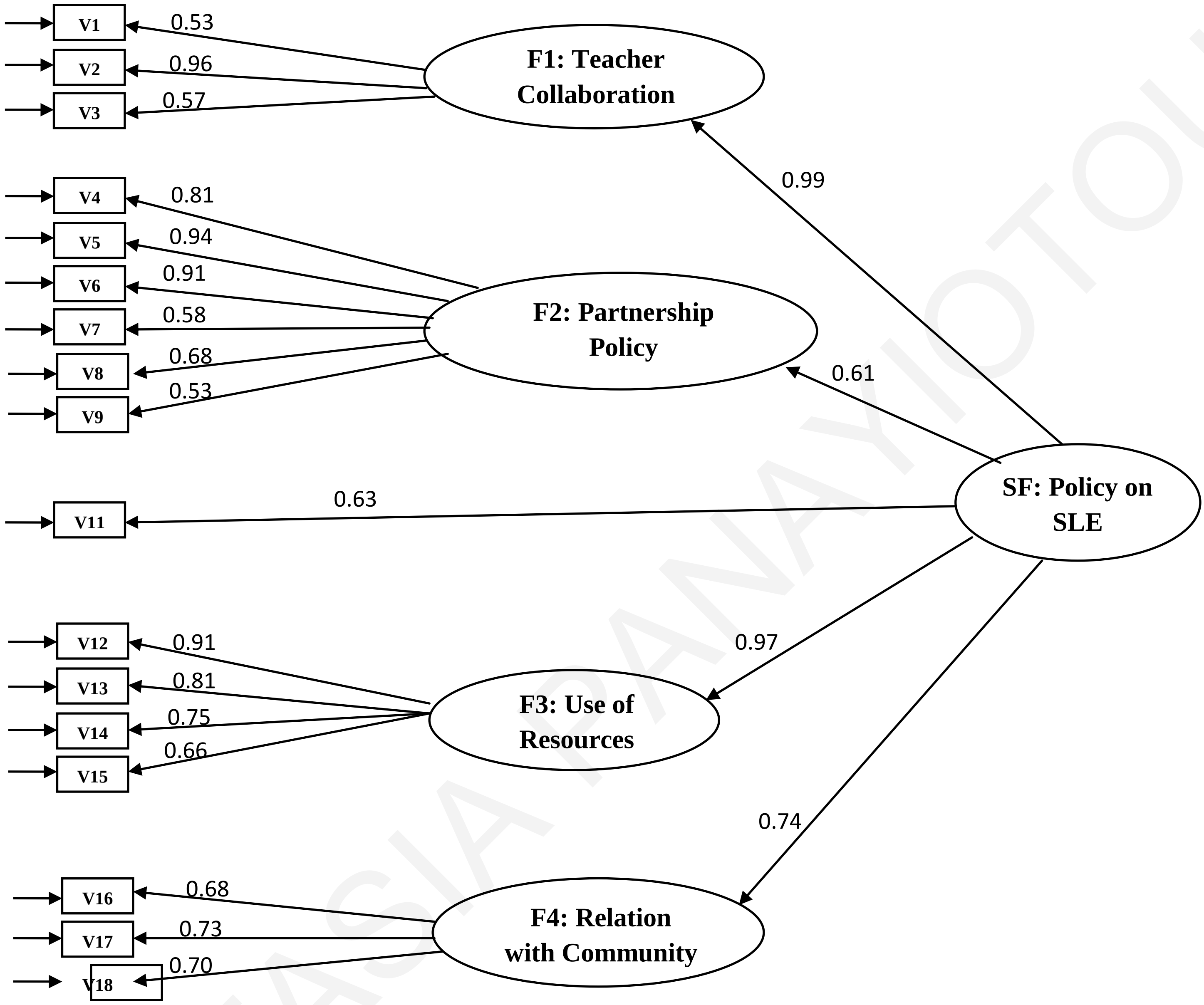


Figure 3.2 The second-order factor model of the teacher questionnaire measuring school factors on the SLE with factor parameter estimates for Germany

ANASTASIA PANAYIOTOU

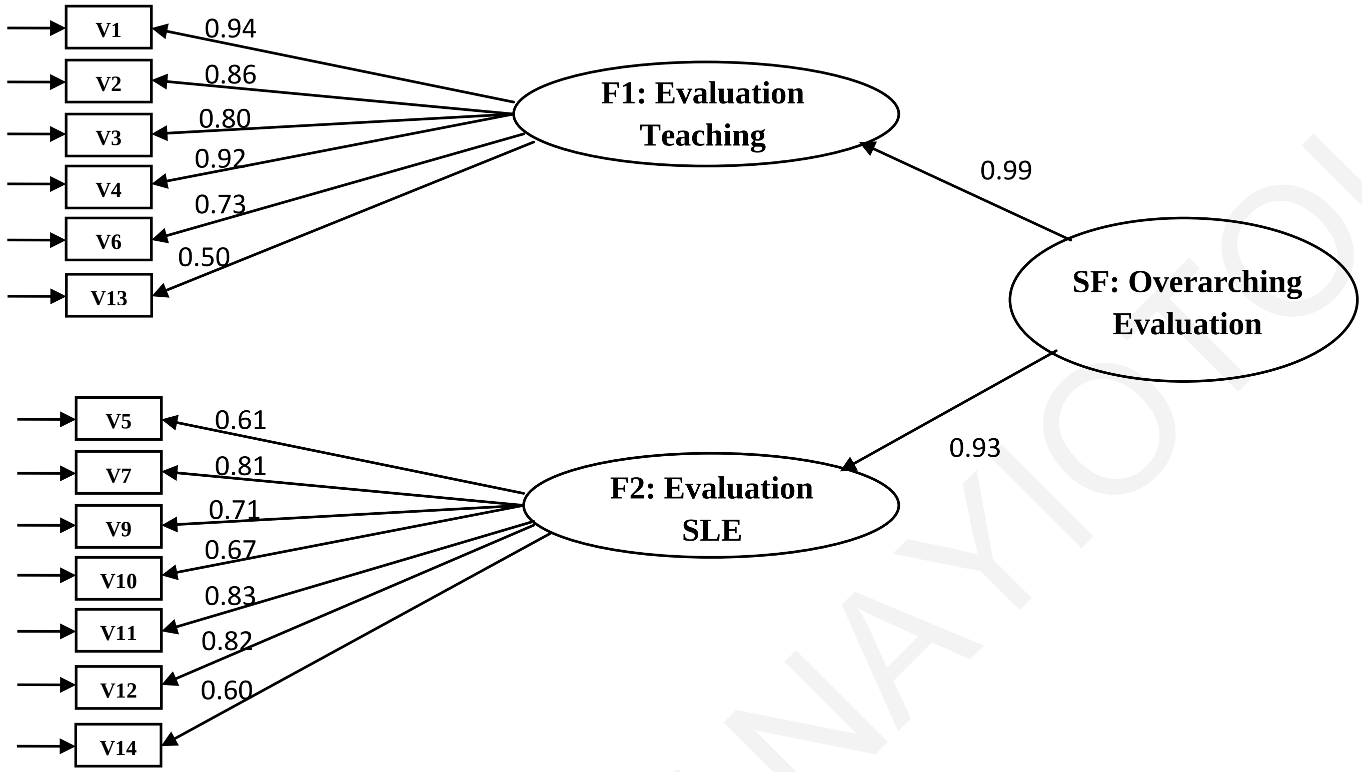


Figure 3.3 The second-order factor model of the teacher questionnaire measuring school factors on evaluation with factor parameter estimates for Germany

ANASTASIA PANAYIOTOU

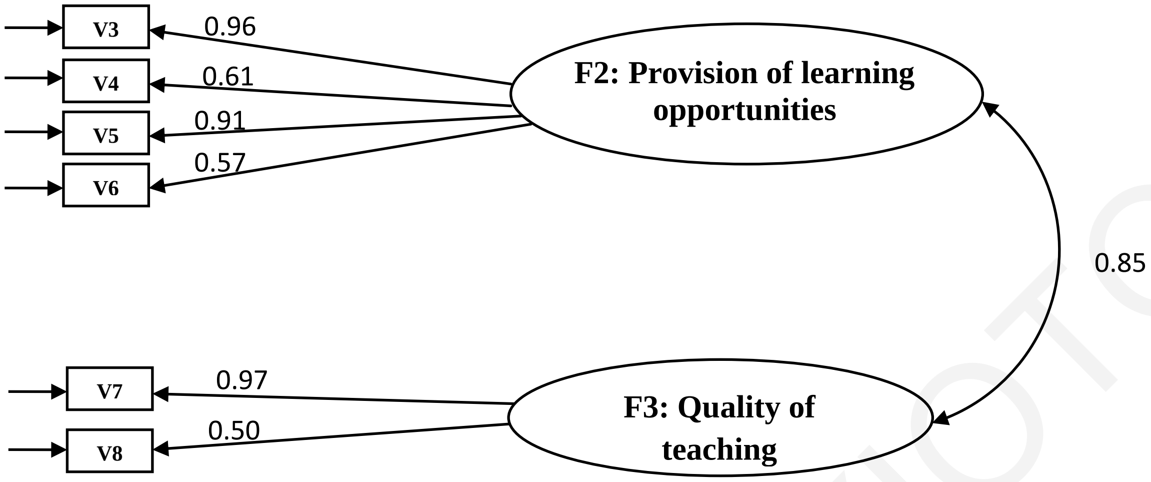


Figure 4.1 The second-order factor model of the teacher questionnaire measuring school factors on policy on teaching with factor parameter estimates for Ireland

ANASTASIA PANAYIOTOU

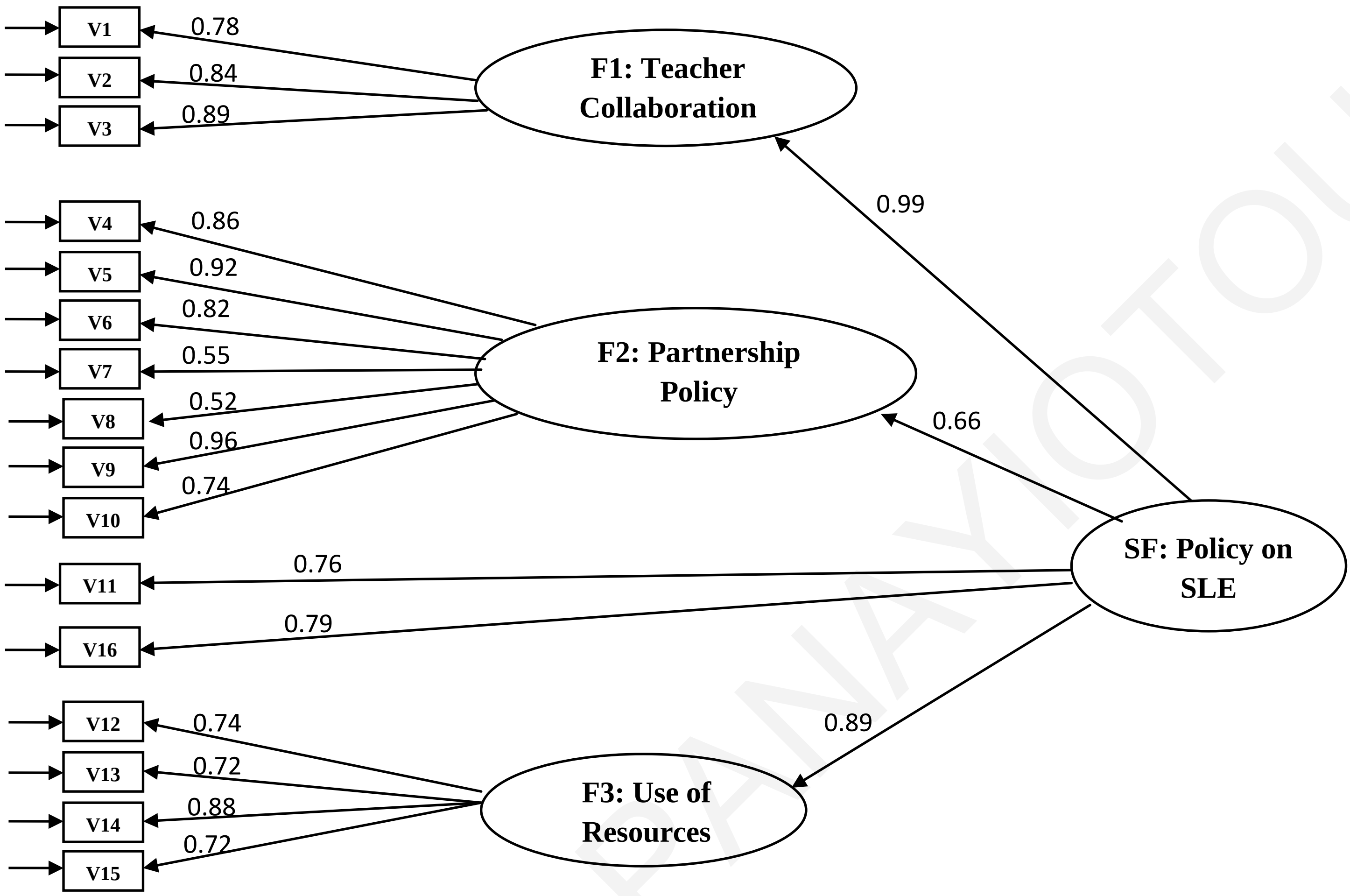


Figure 4.2 The second-order factor model of the teacher questionnaire measuring school factors on the SLE with factor parameter estimates for Ireland

ANASTASIA PANAYIOTOU

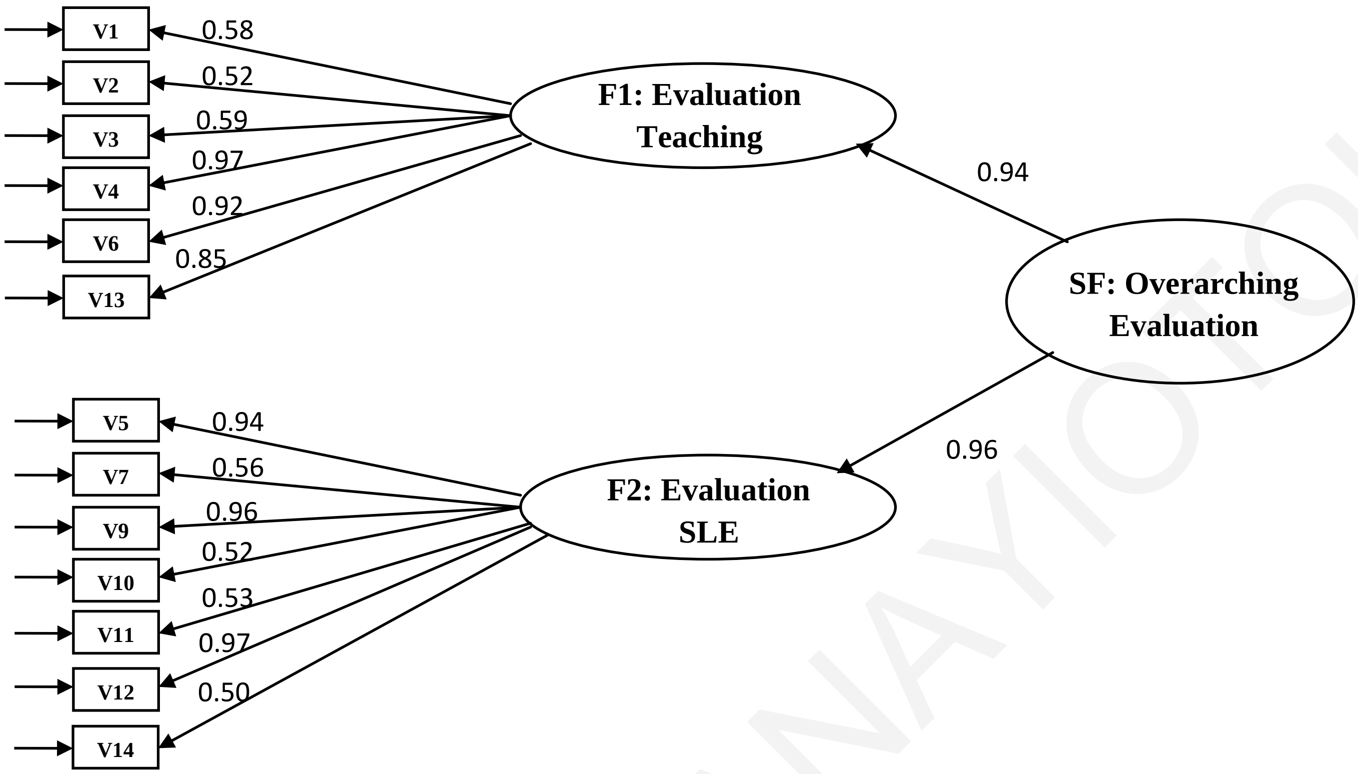


Figure 4.3 The second-order factor model of the teacher questionnaire measuring school factors on evaluation with factor parameter estimates for Ireland

ANASTASIA PANAYIOTOU

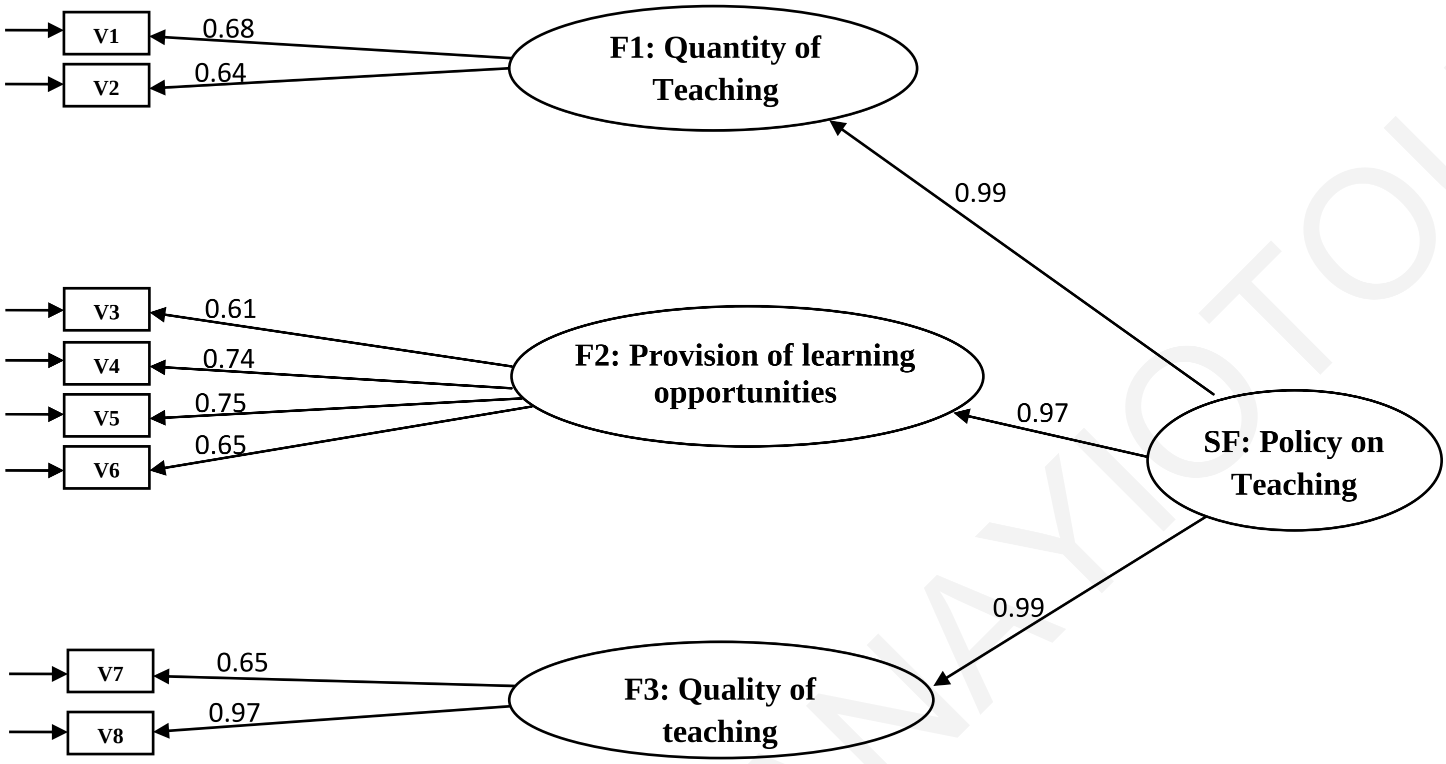


Figure 5.1 The second-order factor model of the teacher questionnaire measuring school factors on policy on teaching with factor parameter estimates for Slovenia

ANASTASIA PANAYIOTOU

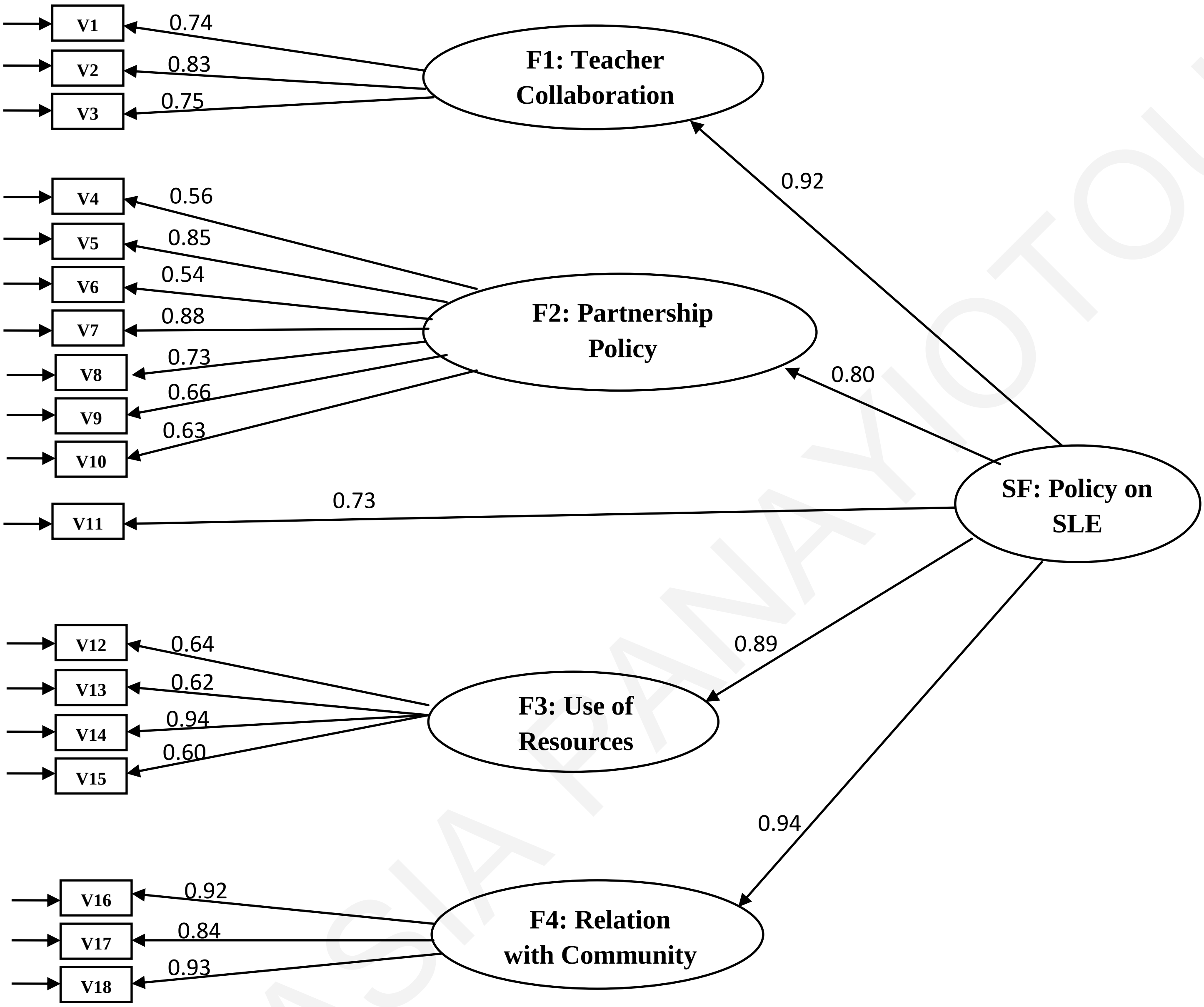


Figure 5.2 The second-order factor model of the teacher questionnaire measuring school factors on the SLE with factor parameter estimates for Slovenia

ANASTASIA PANAYIOTOU

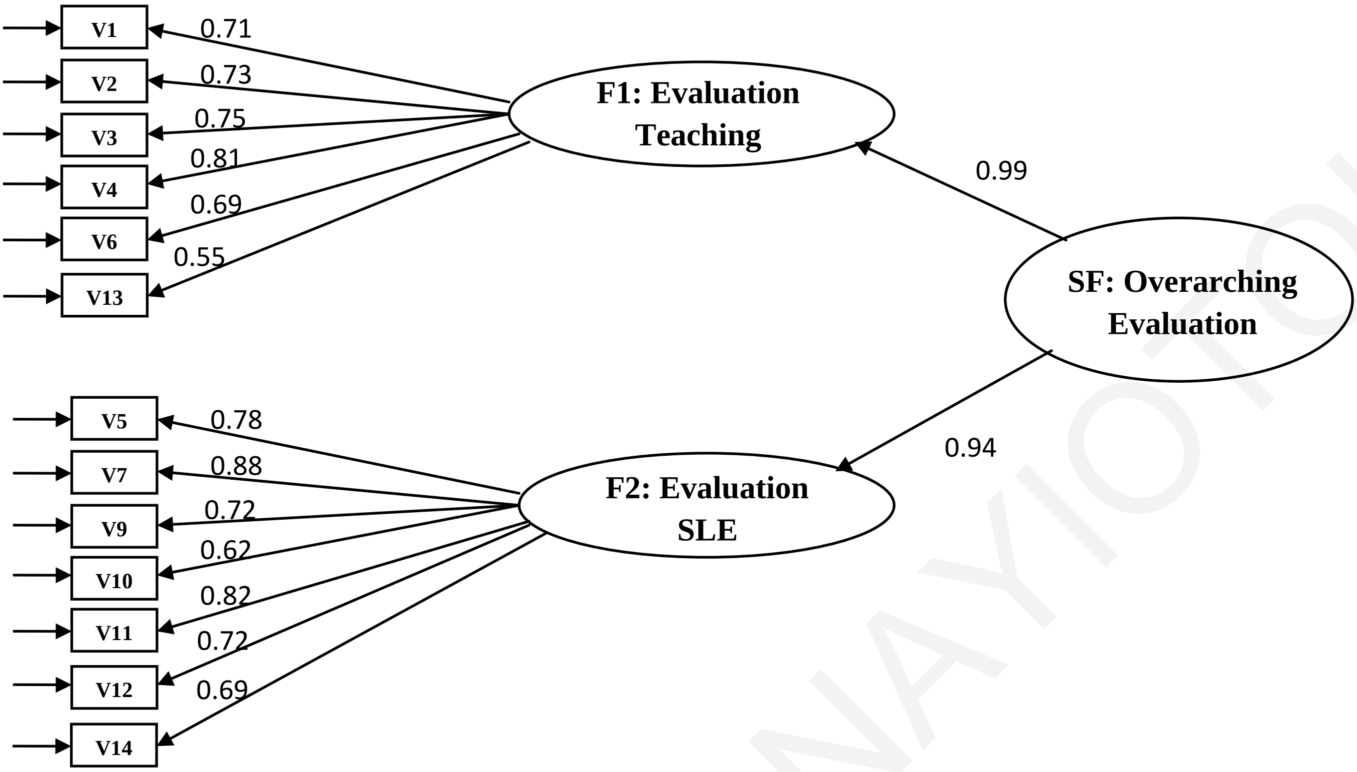


Figure 5.3 The second-order factor model of the teacher questionnaire measuring school factors on evaluation with factor parameter estimates for Slovenia.

ANASTASIA PANAYIOTOU