University of Cyprus

DEPARTMENT OF EDUCATION

COMBINING GENERIC AND CONTENT-SPECIFIC PRACTICES IN EXPLORING TEACHING QUALITY IN PHYSICAL EDUCATION AND ITS IMPACT ON STUDENT LEARNING

DOCTOR OF PHILOSOPHY DISSERTATION

ERMIS S. KYRIAKIDES

2016

University of Cyprus

DEPARTMENT OF EDUCATION

COMBINING GENERIC AND CONTENT-SPECIFIC PRACTICES IN EXPLORING TEACHING QUALITY IN PHYSICAL EDUCATION AND ITS IMPACT ON STUDENT LEARNING

ERMIS S. KYRIAKIDES

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

November 2016

Validation Page

Doctoral Candidate: Ermis S. Kyriakides

Doctoral Thesis Title: Combining Generic and Content-Specific Practices in Exploring Teaching Quality in Physical Education and its Impact on Student Learning

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the ***Department of Education*** *and was approved on the 9th of November 2016 by the members of the* ***Examination Committee.***

Examination Committee

Research Supervisor: Niki Tsangaridou, Professor ………………………………

Committee Member: Leonidas Kyriakides, Professor ………………………………

Committee Member: Charalambos Y. Charalambous, Assistant Professor (Chair of the Examination Committee) ………………………………

Committee Member: Phillip Ward, Professor, Department of Human Sciences, Ohio State University

Committee Member: Melissa Parker, Lecturer and Course Director, Department of Physical Education and Sport Sciences, University of Limerick

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.

Ermis S. Kyriakides

………………………… [Signature]

Abstract (in Greek language)

Η ποιότητα των σχολικών προγραμμάτων Φυσικής Αγωγής (ΦΑ) εξαρτάται από πολλούς παράγοντες, με τη διδασκαλία να αποτελεί έναν από τους πιο σημαντικούς. Για αυτό το λόγο, κατά τη διάρκεια των τελευταίων 40 χρόνων, μια συστηματική ερευνητική προσπάθεια έχει λάβει χώρα με σκοπό να διερευνήσει τη σχέση ανάμεσα στην ποιότητα της διδασκαλίας της ΦΑ και τα μαθησιακά αποτελέσματα. Προσπαθώντας να κατανοήσουν καλύτερα ποιες πτυχές της αποτελεσματικής διδασκαλίας συνεισφέρουν στη μάθηση ψυχοκινητικών δεξιοτήτων, οι ερευνητές στο χώρο της ΦΑ έχουν κυρίως ακολουθήσει δύο παράλληλες πορείες. Λαμβάνοντας υπόψη τις ερευνητικές προσπάθειες που πραγματοποιήθηκαν σε άλλα κύρια γνωστικά αντικείμενα, μια ομάδα ερευνητών εστίασε σε γενικευμένες πρακτικές διδασκαλίας, δηλαδή πρακτικές που μπορούν να εφαρμοστούν σε όλα τα γνωστικά αντικείμενα. Παρόλα αυτά, αυτή η προσέγγιση φάνηκε να μην λαμβάνει υπόψη τις ιδιαιτερότητες και τις διαφορές που υπάρχουν ανάμεσα στους στόχους, το περιεχόμενο και το συγκείμενο της ΦΑ με αυτά των άλλων γνωστικών αντικειμένων. Κάτι τέτοιο, όμως, είχε αναγνωριστεί από μια άλλη ομάδα ερευνητών της ΦΑ, η οποία επικεντρώθηκε σε εξειδικευμένες πρακτικές διδασκαλίας, δηλαδή πρακτικές που έχουν ιδιαίτερη λειτουργία και εξειδικευμένη εφαρμογή όταν χρησιμοποιούνται στο μάθημα της ΦΑ. Παρ’ όλη τη σημαντική συνεισφορά της καθεμιάς από τις δύο προσεγγίσεις στο ερευνητικό πεδίο της αποτελεσματικής διδασκαλίας της ΦΑ, η παράλληλη πορεία αυτών των δύο προσεγγίσεων οδήγησε σε μια κάπως αποσπασματική εικόνα του τι αποτελεί ποιοτική διδασκαλία της ΦΑ.

Έχοντας ως στόχο να καλύψει αυτό το ερευνητικό κενό, η παρούσα έρευνα προσπαθεί να συνδυάσει αυτές τις δύο προσεγγίσεις για να διερευνήσει την ποιότητα της διδασκαλίας στη ΦΑ και τις επιδράσεις της στη μάθηση ψυχοκινητικών δεξιοτήτων. Αυτή η εξερεύνηση καθοδηγείται

από τρία κύρια ερευνητικά ερωτήματα. Το πρώτο ερώτημα αναφέρεται στη συνεισφορά μεμονωμένων γενικευμένων και εξειδικευμένων πρακτικών διδασκαλίας στη μάθηση ψυχοκινητικών δεξιοτήτων. Το δεύτερο αφορά στην προστιθέμενη αξία που προκύπτει μέσα από την εξερεύνηση της κοινής συνεισφοράς των γενικευμένων και εξειδικευμένων πρακτικών, ενώ το τρίτο ερώτημα εξετάζει ποιες γενικευμένες και εξειδικευμένες πρακτικές διδασκαλίας μπορούν να διακρίνουν τους εκπαιδευτικούς ανάλογα με το επίπεδο αποτελεσματικότητάς τους.

Για να απαντηθούν αυτά τα ερωτήματα έχει συλλεχθεί μία μεγάλη γκάμα από δεδομένα. Συγκεκριμένα, για να μετρηθεί η μάθηση στις ψυχοκινητικές δεξιότητες των μαθητών, πραγματοποιήθηκε μία αρχική και μία τελική μέτρηση της επίδοσης των 944 μαθητών 3^{ης}, 4^{ης} και 5^{ης} τάξης σε ένα ψυχοκινητικό δοκίμιο, το οποίο χορηγήθηκε στην αρχή και στο τέλος της σχολικής χρονιάς. Ακολούθως, πραγματοποιήθηκαν τρεις παρατηρήσεις διδασκαλίας (μία ανά τρίμηνο) για καθένα από τους 49 συμμετέχοντες εκπαιδευτικούς. Κάθε παρατήρηση πραγματοποιείτο από τρεις παρατηρητές: έναν που χρησιμοποιούσε το γενικευμένο εργαλείο του Δυναμικού Μοντέλου Εκπαιδευτικής Αποτελεσματικότητας και άλλους δύο που χρησιμοποιούσαν μια τροποποιημένη έκδοση του εξειδικευμένου εργαλείου «Σύστημα Δόμησης Έργων». Επιπρόσθετα, χορηγήθηκε στους μαθητές ένα ερωτηματολόγιο ως συμπληρωματικός τρόπος μέτρησης της ποιότητας διδασκαλίας, στο οποίο οι μαθητές/μαθήτριες καλούνταν να συμπληρώσουν τον βαθμό στον οποίο συγκεκριμένες γενικευμένες και εξειδικευμένες πρακτικές διδασκαλίας χρησιμοποιούνταν από τον/την εκπαιδευτικό τους στα μαθήματα της ΦΑ. Τα δεδομένα αναλύθηκαν με τη χρήση τεσσάρων διαφορετικών προχωρημένων στατιστικών μεθόδων. Οι μέθοδοι που αφορούν στα μοντέλα Item Response Theory και στις αναλύσεις Structural Equation Modeling πραγματοποιήθηκαν για να ελέγξουν την εγκυρότητα της δομής και τις ψυχομετρικές ιδιότητες του ψυχοκινητικού δοκιμίου και των εργαλείων παρατήρησης.

Ακολούθως, πραγματοποιήθηκαν πολύ-επίπεδες αναλύσεις για να εξερευνηθεί η μεμονωμένη αλλά και η από κοινού επίδραση των γενικευμένων και εξειδικευμένων πρακτικών διδασκαλίας στη μάθηση ψυχοκινητικών δεξιοτήτων, ενώ ακολούθησε η ανάλυση διάκρισης, η οποία είχε σκοπό να καθορίσει ποιες πρακτικές μπορούσαν να διακρίνουν τους εκπαιδευτικούς βάσει του επιπέδου αποτελεσματικότητάς τους.

Τα αποτελέσματα της έρευνας που αφορούσαν στο πρώτο ερευνητικό ερώτημα ενίσχυσαν από τη μια τα υφιστάμενα ερευνητικά αποτελέσματα που υπογραμμίζουν τη σημασία συγκεκριμένων γενικευμένων και εξειδικευμένων πρακτικών στη διδασκαλία ψυχοκινητικών δεξιοτήτων, ενώ από την άλλη ανέδειξαν τη συνεισφορά άλλων γενικευμένων και εξειδικευμένων πρακτικών οι οποίες δεν έχουν μέχρι στιγμής μελετηθεί σε ικανοποιητικό βαθμό. Συγκεκριμένα, οι γενικευμένες πρακτικές του προσανατολισμού, της διαχείρισης χρόνου και των τεχνικών ερώτησης, καθώς και οι εξειδικευμένες πρακτικές της επίδειξης μιας δεξιότητας και της συναφής και συγκεκριμένης ανατροφοδότησης βρέθηκαν να έχουν τις μεγαλύτερες επιδράσεις στην ψυχοκινητική μάθηση. Όσον αφορά στην κοινή συνεισφορά των δύο τύπων πρακτικών, τα αποτελέσματα έδειξαν ότι ερμηνεύθηκε ένα μεγαλύτερο ποσοστό διασποράς στο επίπεδο του δασκάλου όταν συνδυάστηκαν οι γενικευμένες και οι εξειδικευμένες πρακτικές σε σύγκριση με το ποσοστό που ερμηνεύθηκε όταν χρησιμοποιήθηκε ο κάθε τύπος πρακτικής από μόνος του. Επιπρόσθετα, μία γενικευμένη (τεχνικές ερώτησης) και μια εξειδικευμένη (ποιότητα εξάσκησης των μαθητών) πρακτική διδασκαλίας ήταν οι δύο κύριες πρακτικές που μπορούσαν να κατανέμουν τους εκπαιδευτικούς στις κατηγορίες των πιο-αποτελεσματικών ή μη-πιο-αποτελεσματικών εκπαιδευτικών. Ασφαλώς, όλα τα πιο πάνω αποτελέσματα έχουν αρκετές σημαντικές θεωρητικές, μεθοδολογικές και πρακτικές εφαρμογές, οι οποίες συζητιούνται μαζί με εισηγήσεις για μελλοντική έρευνα στο πεδίο της αποτελεσματικής διδασκαλίας της ΦΑ.

Abstract (in English language)

The quality of school Physical Education (PE) programs depends on several factors with key among them being actual teaching. Thus, during the last 40 years a persistent research effort has been undertaken to investigate the relationship between teaching quality in PE and student learning. Trying to better understand what aspects of effective teaching relate to student psychomotor learning, PE researchers have largely pursued two parallel perspectives. Following the lead of classroom research, a group of researchers focused on generic teaching practices, namely teaching behaviors that cut across different subject matters. However, this perspective seemed to ignore the particularities and differences of the PE aims, content and context from those of classroom disciplines, something that had been recognized by the other group of PE researchers, who have focused on content-specific teaching practices. These teaching practices are considered to have a particular functioning and specialized manifestation when occurring in the teaching of PE. Despite the significant input of each perspective to the research field of teaching effectiveness in PE, their parallel course led to a somewhat fragmented picture of what constitutes quality teaching in the field of PE.

Aiming to address this research gap, the present study attempts to bring together those two perspectives to explore teaching quality in PE and its effects on student psychomotor learning. This exploration is guided by three main research questions. The first pertains to the individual contribution of certain generic and content-specific teaching practices to student psychomotor learning. The second concerns the added value that emerges when exploring the joint contribution of generic and content-specific teaching practices, as opposed to considering each type of practices in isolation; and the third examines which generic and content-specific teaching practices can discriminate among teachers based on their level of effectiveness.

To answer these research questions, a large gamut of data were collected. In particular, to measure students' psychomotor growth, a pre- and post- student performance test was administered to the 3rd, 4th, and 5th grade students (N=944) that participated in the study, at the beginning and culmination of the school year. Then, three classroom observations (one per each trimester) were conducted for each of the 49 participating teachers. Each observation was carried out by three observers: one using the *Dynamic Model of Educational Effectiveness* generic instrument and the other two the *modified Task Structure System* content-specific instrument. In addition, a student survey was also used as a supplementary measurement approach of instructional quality, to capture the extent to which certain generic and content-specific teaching practices were used in everyday lessons. Data were analyzed by employing four different advanced statistical techniques. First, Item Response Theory (IRT) models and Structural Equation Modeling (SEM) analyses were run to test the construct validity and the psychometric properties of the student performance test and the observation forms. Then, multi-level model analyses were run to explore the individual and joint effects of generic and content-specific practices on student psychomotor learning, followed by the employment of a discriminant analysis, which aimed to determine which teaching practices could discriminate among teachers based on their level of effectiveness.

The findings that concern the first research question not only corroborated existing research findings that underline the importance of certain generic and content-specific practices in teaching psychomotor skills, but also highlighted the contribution of other under-explored generic and content-specific teaching practices. Specifically, the generic practices of orientation, time management, and questioning as well as the content-specific practices of skill demonstration and congruent and specific feedback were found to have the largest effects on student psychomotor learning. As far as the joint contribution of these two types of practices is concerned, findings showed that more teacher-level variance was explained when combining generic and content-specific practices, as compared to that explained when considering either type of practices in isolation. Furthermore, one generic (i.e., questioning) and one content-specific (i.e., student quality practice) teaching practices were the two strongest practices responsible for the allocation of teachers to the categories of most-effective or non-most effective (i.e., typical or least-effective). All the above findings have several important theoretical, methodological, and practical implications which are discussed along with suggestions for future research in the field of teaching effectiveness in PE.

Acknowledgements

The completion of my PhD dissertation has been a long and difficult journey, which could not have been completed without the valuable support and help from several individuals. Thus, to thank these people, the least I can do is to acknowledge their significant contribution in enriching, facilitating and making pleasant this journey.

First of all, I would like to thank my advisor Professor Niki Tsangaridou for all her assistance and valuable support throughout the completion of this study. I honestly appreciate her constant encouragement and her readiness to help me at any time through the whole research process. The discussions we had were always initiated and guided by her wisdom, helping me create a more comprehensive picture of what I was looking for. Before embarking on my PhD, I remember asking her whether it would be worth pursuing this journey… Now I am confident that it sure did!

I would also like to thank Professor Leonidas Kyriakides who has offered his immense knowledge and expertise, which helped me further my knowledge on the field of teaching effectiveness and see the whole process from the researcher's perspective. His experience and well-aimed comments and suggestions helped me overcome many complicated situations emerged during this journey.

In addition, I would like to express my sincere gratitude to the Assistant Professor Charalambos Charalambous, who has been a tireless and supportive guide throughout the entire journey. His enthusiasm as well as the many hours he dedicated to helping me refine my work are well appreciated. In addition, the stimulating and insightful discussions we had, in conjunction with his persistence to employ a constructivist approach to elicit the answer I was looking for, led to my intellectual growth and helped me grow as a researcher.

My appreciation also goes to the two external members of my examination committee, Professor Phillip Ward and Lecturer Melissa Parker, whose thoughtful comments and suggestions contributed to the improvement of my work. It was truly an honor for me to present and discuss my research work with those two scholars, whose contribution to the research field of physical education is well recognised.

I am also grateful to the Ministry of Education and Culture of Cyprus and the Centre of Educational Research and Evaluation for their permission to conduct the study. Moreover, my gratitude goes to the University of Cyprus which funded the project: "*Integrating Generic and Domain-Specific Factors in Exploring the Association between the Quality of Instruction and Student Learning*" part of which was this study.

Furthermore, I would like to personally thank each of the participating teachers who were willing to open their doors and allowed us to observe their teaching; a phenomenon that is not common, neither in Cyprus nor in other countries. Many thanks go to their students as well, for their time and eagerness to participate in this study. I also wish to thank each of the principals who consented to enter their schools, since without their permission this study would not have been feasible to run. Special thanks also go to all the members of the research team who worked professionally and helped me collect quality data.

It would be remiss if I did not thank my fellow student Lambros Stefanou for his constant and valuable help. The team spirit we have developed as well as the mutual support we have offered to each other helped us overcome any obstacle occurred in our parallel journey. I will miss the passionate and endless discussions about our research interests!

Last but not least, I would like to thank my family for always being there for me, encouraging me to complete this research. First, I am grateful to my adorable wife Niki

Charalambous for her limitless patience, support and love as well as her critical comments which contributed to the refinement of this study. This dissertation is highly dedicated to her, since she lived every moment of this journey as if it was hers! Certainly, I would also like to dedicate this dissertation to my newborn twins, Marilia and Raphael, who even though significantly decreased my sleep hours, yet they have given me incredible happiness and courage to keep on my journey! Finally, particular thanks go to my parents, my brother and sister who have always supported me in fulfilling my dreams.

Table of Contents

Abstract (in Greek language) ... iv
Abstract (in English language) ... vii
Acknowledgements ... x
List of Figures ... xv
List of Tables ... xvi
List of Appendices ... xviii
List of Abbreviations ... xix
CHAPTER 1: INTRODUCTION ... 1
Abstract ... 1
The Importance of School PE and Teaching ... 1
Statement of the Problem ... 3
Purpose of the Study ... 4
Research Questions ... 5
Significance of the Study ... 5
Limitations ... 6
Definition of Key Terms ... 8
CHAPTER 2: THEORETICAL PERSPECTIVES ... 10
Abstract ... 10
Defining and Measuring Teaching Quality ... 10
The Two Parallel Research Perspectives ... 12
Generic Conceptual Frameworks ... 12
Research Findings on the Contribution of Generic Teaching Practices to Student Psychomotor Learning ... 20
Content-Specific Conceptual Frameworks ... 37
Research Findings on the Contribution of Content-Specific Teaching Practices to Student Psychomotor Learning ... 43
Combining the Two Types of Teaching Practices ... 62
CHAPTER 3: METHODOLOGY ... 67
Abstract ... 67
Research Setting and Participants ... 67
Research Setting ... 67
Participants ... 70
Instrumentation ... 71
Student Performance Test ... 72
Observation Instruments ... 73
Student Survey ... 78
Data Collection ... 80
Observer Recruitment and Training ... 81
Scorer Recruitment and Training ... 83

Teacher Recruitment 85
Pre-Test Administration 86
Lesson Observations 86
Post-Test and Student Survey Administration 88
Issues of Validity and Reliability 88
Data Analysis 90
Item-Response-Theory (IRT) Analyses 90
Confirmatory Factor Analyses (CFA) 101
Multilevel Analyses 112
Discriminant Analysis 113
CHAPTER 4: RESULTS 115
Abstract 115
Descriptive Statistics 116
Variance Decomposition and the Contribution of Student- and Teacher-Background Variables to Student Psychomotor Learning 121
The Contribution of Generic Teaching Practices to Student Psychomotor Learning 123
The Contribution of Content-Specific Teaching Practices to Student Psychomotor Learning 133
The Joint Contribution of Generic and Content-Specific Teaching Practices to Student Psychomotor Learning 142
Determining Generic and Content-Specific Teaching Practices That Discriminate Among Teachers Based on Their Level of Effectiveness 147
Summary of Findings with Respect to the Three Research Questions 149
CHAPTER 5: DISCUSSION AND CONCLUSIONS 152
Abstract 152
Teachers and Teaching Do Matter For Student Learning 153
Student Background Characteristics That Influence Student Psychomotor Learning 154
Enhancing and Corroborating Existing Research Evidence 155
Toward an Integration of Generic and Content-Specific Teaching Practices 166
Practices That Distinguished Most-Effective Teachers From Non-Most Effective 168
Implications of the Study 170
Theoretical Advancements of Educational Effectiveness Models: Toward the Development of a Comprehensive Conceptual Framework and its Associated Observation Instrument 170
Practical Implications: Assessing and Improving Teaching Quality in PE 177
Methodological Implications: Advancing Future Research Works 180
Additional Suggestions for Future Research 182
Conclusion 184
Appendices 186
References 208

List of Figures

Figure

2.1 Indirect influence of teaching practices on psychomotor learning through quality of student practice 50

3.1 Design of the study (presented in chronological order) 80

3.2 The one-parameter IRT scale developed for students' psychomotor learning 92

3.3 The one-parameter IRT scale capturing teachers' use of generic teaching practices 96

3.4 The one-parameter IRT scale capturing teachers' use of content-specific teaching practices 99

5.1 A preliminary attempt of integrating generic and content-specific teaching practices 173

List of Tables

Table

2.1 Description of the Eight Classroom Factors and the Five Dimensions of the Dynamic Model of Educational Effectiveness 15

2.2 Outline of the Two out of Four Domains of the Framework for Teaching Model..... 17

2.3 Description of the Three Domains and Thirteen Teaching Components of the latest version of CLASS 19

2.4 Summary of the Studies Exploring the Contribution of Generic Teaching Practices to Psychomotor Learning 33

2.5 Description of the Five Task Dimensions of the OSCD-PE 39

2.6 Description of the Three Domains of the QMTPS 40

2.7 Description of the Five Main Dimensions of the Task Structure System 42

2.8 Summary of the Studies Exploring the Contribution of Content-Specific Teaching Practices to Psychomotor Learning 54

2.9 Description of the Four Dimensions of the AQTR..... 63

3.1 The Six Main Aims of the Elementary PE Curriculum in Cyprus 68

3.2 The 13 Psychomotor Skills Included in the Criterion-Reference Test 73

3.3 Description of the Demonstration Dimension of the mTSS Instrument 76

3.4 The Distribution of the Six Pinafores to Targeted Students 88

3.5 Item and Student Parameter Estimates for the Scale Developed to Capture Student Psychomotor Learning 91

3.6 Item and Teacher Parameter Estimates for the Scale Developed to Capture Teaching Practices With DMEE Low-Inference Instrument..... 95

3.7 Variables Used From the DMEE Low-Inference Classroom Observation Instrument in IRT Analysis 97

3.8 Item and Teacher Parameter Estimates for the Scale Developed to Capture Teaching Practices With mTSS Low-Inference Instrument 98

3.9 Variables Used From the mTSS Low-Inference Classroom Observation Instrument in IRT Analysis 100

3.10 Maximum Likelihood Estimates for an Eight-Factor Solution of the DMEE High-Inference Classroom Observation Data and Fit Statistics for the Model..... 104

3.11 Values of Selected Fit Statistics for Alternative mTSS First-Order Models..... 106

3.12 Maximum Likelihood Estimates for a Five-Factor Solution of the mTSS High-Inference Classroom Observation Data and Fit Statistics for the Model..... 107

3.13 Values of Selected Fit Statistics for Alternative First-Order Models of Student Survey 110
3.14 Maximum Likelihood Estimates for a Four-Factor Solution of the Student Survey Data and Fit Statistics for the Model 111
4.1 Descriptive Statistics for the Five Factors and Their Constituent Aspects/Dimensions of the DMEE Low-Inference Form..... 117
4.2 Descriptive Statistics for the Seven Factors and Their Constituent Aspects/Dimensions of the mTSS Low-Inference Form 119
4.3 Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes--Generic Teaching Practices Captured With DMEE and mTSS Low-Inference Forms 124
4.4 Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes--Generic Teaching Practices Captured With DMEE and mTSS High-Inference Forms and Student Ratings 126
4.5 Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes—Content-Specific Teaching Practices Captured With mTSS Low-Inference Form 135
4.6 Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes—Content-Specific Teaching Practices Captured With mTSS High-Inference Form and Student Ratings 137
4.7 Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes--Combination of Generic and Content-Specific Teaching Practices Captured With Classroom Observation Forms and Student Ratings..... 143
4.8 Classification Results of Teacher Effectiveness..... 148
4.9 Structure Matrix Based on Analysis of Teacher Effectiveness 149

List of Appendices

Appendix

A Psychomotor Test Organizing Diagram and Examples of Task Description, Assessment Scoring Rubric and Assessment Score Sheet 187

B The Dynamic Model of Educational Effectiveness Observational Instruments (Low- and High-Inference Forms) 192

C The Modified Task Structure System Observational Instruments (Low- and High-Inference Forms) 199

D Student Survey 203

List of Abbreviations

ALT	Academic Learning Time
AQTR	Assessing Quality Teaching Rubrics
CFA	Confirmatory Factor Analysis
CLASS	Classroom Assessment Scoring System
DMEE	Dynamic Model of Educational Effectiveness
FfT	Framework for Teaching
IRT	Item Response Theory
MET	Measures of Effective Teaching
mTSS	modified Task Structure System
OSCD-PE	Observation System for Content Development-Physical Education
OTR-PE	Opportunity To Respond-Physical Education
PE	Physical Education
PELOT	Physical Education Lesson Observation Tool
PLATO	Protocol for Language Arts Teaching Observation
QMTPS	Qualitative Measures of Teaching Performance Scale
SEM	Structural Equation Modeling
TSS	Task Structure System

CHAPTER 1: INTRODUCTION

Abstract

The quality of school PE programs depends on several factors with key among them being actual teaching. Trying to better understand what aspects of effective teaching relate to student psychomotor learning, PE researchers have largely pursued two parallel perspectives: one that focused on generic teaching practices (i.e., teaching behaviors that cut across different subject matters) and another that emphasized content-specific teaching practices (i.e., teaching behaviors that are more pertinent to teaching PE). Although identifying the contribution of certain teaching practices to student learning in PE, the parallel work of these two research strands led to a somewhat fragmented picture of what constitutes quality teaching in the field of PE. Aiming to address this research gap, the present study attempts to bring together those two perspectives by combining generic and content-specific teaching practices to explore the quality of teaching PE and, in turn, its effects on student psychomotor learning. This exploration is guided by the three main research questions, which pertain to the individual and joint contribution of generic and content-specific teaching practices to student learning as well as the extent to which certain practices can discriminate among teachers based on their level of effectiveness. The reasons why addressing these questions are detailed as well as the limitations of this study, which should be considered when interpreting its findings.

The Importance of School PE and Teaching

Nowadays, the importance of engaging young children in quality Physical Education (PE) activities becomes even more crucial, as researchers and organizations are constantly recording increasing percentages of sedentary living, childhood obesity, and associated diseases (Daniels, 2006; Wang & Lobstein, 2006; WHO, 2012). Given that school remains the most likely and cost-effective place for all children to gain access to PE experiences (Kirk, 2005; McKenzie, 2007), many researchers turned their attention to the quality of school PE programs. Participation in quality school PE programs has been underlined as prophylactic, proactive, and effective way to combat inactivity and its negative consequences (Bailey, 2006; Bailey et al., 2009), but at the same time, it is considered as the main vehicle to lifelong physical activity (Green, 2012). Importantly, when starting at an early age, these quality PE experiences are even more vital, as they contribute to the development of fundamental psychomotor skills, which, in turn, serve as a primary mechanism that promotes engagement and adherence to a physically active lifestyle (Griggs, 2012; Jess, Dewar, & Fraser, 2004; MacNamara et al., 2011; Stodden et al., 2008; Trudeau, Laurencelle, Tremblay, Rajic, & Shephard, 1999).

Undoubtedly, the quality and effectiveness of school PE programs depends on several factors, including student-, classroom-, and school-level factors (Kyriakides & Tsangaridou, 2008). However, there seems to be a consensus among PE scholars that the most important school-related factor responsible for student learning concerns actual teaching (Siedentop & Tannehill, 2000). In other words, what teachers do in the classroom and how they interact with their students and the content is considered of great importance and has been shown to have considerable effects on student achievement not only in PE (Castelli & Rink, 2003; Rink, 2009; Silverman & Mercier, 2015), but also in various disciplines (Cohen, Raudenbush, & Ball, 2003; Hattie, 2009; Kyriakides, Creemers, Antoniou, & Demetriou, 2010; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Seidel & Shavelson, 2007; Wright, Horn, & Sanders, 1997). In her attempt to emphasize the significant impact the teacher can have on student learning, Roberts (2014) states:

> We can build beautiful facilities at our schools, we can receive large grants for equipment and monitoring of activity levels, we recruit and hire administrators and directors, but in the end, it is the work and success of the individual teacher at each school and in each gym that determines student outcomes and learning. (p. 27)

Consequently, the fact that the vast majority of research on teaching in PE has focused on teaching (Kulinna, Scrabis-Fletcher, Kodish, Phillips, & Silverman, 2009) or teacher effectiveness (Silverman & Skonie, 1997) cannot be considered a coincidence; that this research strand is increasingly growing also reveals that the field is still vibrant (Chatoupis & Vagenas, 2011).

Statement of the Problem

In their effort to better understand what aspects of effective teaching relate to student psychomotor learning, PE researchers have attended to various instructional aspects either in isolation or in conjunction. A close examination of this research work, suggests that PE researchers have largely pursued two distinct perspectives; a fact recognized even in the early 1980s (Graham & Heimerer, 1981). On one hand, guided by the principle that there are significant elements of effective teaching that are shared across domains, some PE scholars have followed the lead of classroom effectiveness research (e.g., Brophy & Good, 1986; Dunkin & Biddle, 1974; Rosenshine & Stevens, 1986) focusing on generic teaching practices--that is teaching behaviors that cut across different subject matters (e.g., maximizing Academic Learning Time (ALT), Siedentop, Tousignant, & Parker, 1982; establishing a positive classroom climate). These scholars supported that although the context within which the PE lesson takes place may be different, the essential teaching practices which needed to produce learning in PE are similar to those needed in the classroom (Rink & Hall, 2008; Siedentop & Tannehill, 2000). On the other hand, believing that each discipline has its own unique nature and specific structure that both influence actual teaching practices (Bransford, Brown, & Cocking, 2000; Schwab, 1978; Shulman, 1986), other researchers have claimed that PE may have unique elements of instruction. Hence, these scholars have focused on exploring content-specific teaching practices--that is teaching behaviors which are more pertinent to specific disciplines (e.g., for PE: demonstrating the desired motor skill performance, Magill, 2010)--supporting that the implementation of these practices within the discipline of PE can significantly contribute to student learning.

Despite the fact that the abovementioned research work has identified the contribution of some discrete teaching practices (either generic or content-specific) to student learning in PE, due to their parallel course these two research strands led to a somewhat fragmented picture of what constitutes quality teaching in the field of PE. As Rink (2013) aptly recognizes, this disjointed approach has led to the identification of "*necessary but not sufficient*" teaching practices for student learning. Differently put, although these lines of research underlined the important role of teachers as well as the significant effect that certain generic and content-specific teaching practices have on student learning in PE, yet, they did not investigate how the concurrent use of these practices during PE lessons can impact student psychomotor learning, something that would provide a more in-depth understanding of the teaching-learning process in PE.

Purpose of the Study

Aiming to address this research gap, the present study makes a step toward the construction of a more comprehensive picture of effective teaching in PE, capitalizing on both generic and content-specific teaching practices. Particularly, following the recent calls voiced by scholars from other fields who have pinpointed the importance of exploring in more depth the joint contribution of these two types of practices (e.g., Charalambous, Komitis, Papacharalambous, & Stefanou, 2014; Grossman & McDonald, 2008; Hamre et al., 2013) as well as the need for more evidence in why and how teachers matter (Konstantopoulos, 2012; Pianta & Hamre, 2009), this study aims to combine generic and content-specific teaching practices to explore the quality of teaching PE and, in turn, its effects on student psychomotor learning.

Research Questions

In this context, and taking into consideration that PE should be taught using effective teaching practices that have been shown to improve student learning (Rink, 2013; Siedentop, 2009; Tannehill & Lund, 2010), the present study seeks to answer the following research questions:

- To what extent do generic or content-specific teaching practices contribute toward explaining student psychomotor learning, when controlling for certain student background characteristics?
- What is the added value of exploring *both* generic and content-specific teaching practices as opposed to considering each type of practices in isolation?
- Which teaching practices (generic and/or content-specific) could discriminate among most-, typical and least-effective teachers, and which of them contribute most to teachers' separation in these three groups?

Significance of the Study

Combining generic and content-specific teaching practices to address the above research questions seems quite promising and productive for at least three reasons. First, teaching is a very complex phenomenon (Cohen, 2011; Danielson, 2007; Silverman & Ennis, 2003), and as such researchers do not have the luxury of being exclusive or studying instruction in a piecemeal way. Rather, to provide a more in-depth understanding of this multifaceted phenomenon, they need to be inclusive, considering different approaches. As Silverman, Woods, and Subramaniam (1999) argue, while it may be more demanding to conduct such research, this would offer a much deeper analysis of teaching and more authentic picture of the multidimensional nature of teaching PE. Second, as it is outlined in the next section, several studies in PE as well as recent meta-analyses in other content-areas (Hattie, 2009; Seidel & Shavelson, 2007) have shown these practices, largely in isolation, to contribute to student learning. Thus, one can speculate that combining practices identified to have a positive impact on student learning, might help us do even a better job in describing instructional quality and understanding how it affects student learning. Third, recent research findings in other content-areas (e.g., Kane & Staiger, 2012) suggest that the correlations between generic and content-specific constructs are not as high as those among the instruments incorporating the same type of practices. This implies that there might be a correlation among the two types of practices, as they measure the same construct (i.e., instructional quality), yet, one perspective cannot substitute the other. Therefore, it seems reasonable to assume that the combination of the two types of practices will corroborate the assumption about the added-value of exploring simultaneously both types of practices in exploring student learning. Taking all the above into consideration and given that, to the best of our knowledge, no other studies have been undertaken to synthesize generic and content-specific teaching practices to explore the association between the quality of instruction and student learning in PE, conducting such a study seems to be highly warranted. That this study is one of the few studies in PE with such a large teacher and student sample as well as one of few that used more sophisticated research design and advanced statistical analyses also reveals the significance of conducting this study.

Limitations

Apparently, this study could not be without limitations. To start, although a quality PE program aims at developing psychomotor, cognitive, affective, and social learning domains, the present study focused on the former, as psychomotor outcomes are considered to be the unique contribution of PE to students' education, since no other discipline emphasizes the development of those skills the way PE does (Rink, 2010). However, it is acknowledged that the study could have investigated the impact of generic and content-specific teaching practices on other learning outcomes as well (see Mercier & Doolittle, 2013), but due to budget constraints this was not feasible.

Second, to measure students' psychomotor development, a criterion-reference performance test was used, involving 13 fundamental psychomotor skills related to objectives set in the national curriculum. Despite the fact that such tests are widely used and have been developed and validated by international organizations to measure student achievement of the national standards (see NASPE, 2010), yet, they are criticized for measuring decontextualized psychomotor skills without capturing student learning in authentic contexts (Siedentop & Tannehill, 2000). However, once validated, these tests are deemed objective and are recognized as important indicators of achievement by educators, policymakers, and the public (Strong, 2011). In addition, due to budget constraints, students' performance in pre- and post-tests was coded in live conditions instead of being videotaped, something that would give the opportunity to scorers to rewind the tape and re-assess students' performance in cases where they were not entirely confident about the score they have assigned. Nevertheless, the fact that students' performance was coded in live-conditions does not impair the quality of the data, since this approach is deemed legitimate and reliable once the employed rubrics are clear and concise enough (Hushman, Hushman, & Carbonneau, 2015), something that was the case with this study.

Third, the study measured student performance at the beginning and end of the school year, and by doing so it only investigated the short-term effects of teaching on student psychomotor learning. Recognizing that students' success may be different between the acquisition and retention phase (Sidaway, Fairweather, Powell, & Hall, 1992), a retention test could have also been distributed at the start of the next academic year (when at least three months would occur between skill acquisition and test), something that would allow to test for the long-term effects of teaching. However, due to reasons not unrelated to time and budget constraints, a retention test was not administered to students.

Fourth, the study used live observations along with student ratings to measure teaching quality, both of which can avoid many of the biases of self-report data and can provide objective information (Strong, 2011). Certainly, classroom observations could have been videotaped to enable the observers re-examine/re-assess some events that happened very quickly during real teaching. In addition, it is acknowledged that other measurement approaches could also have been used for capturing teaching quality, such as post-lesson interviews with each teacher, to investigate teachers' interpretations about the employment of certain practices. Moreover, beyond the teaching practices, other aspects of teaching that have a significant impact on student learning in PE (e.g., the role of the motivational climate created by the PE teacher, see Standage, Duda, & Ntoumanis, 2006) could have also been taken into consideration. Yet, due to budget and time limitations, it was decided not to do all the above acts.

Definition of Key Terms

Before shifting to discussing the theoretical perspectives that underpine this research work, the key terms that are used throughout this study are defined.

Classroom Observation Instrument/System A tool that aids instructional analysis, as it helps capture whether a teacher employs certain teaching practices, as well as the quality of the employed practices (Wood et al., 2014).

Content-Knowledge The knowledge and skills: a) one needs to perform a task/skill (i.e., common content knowledge), and b) which inform how to teach a task/skill (i.e., specialized content knowledge). Importantly, the latter form of content-knowledge should not be confounded with pedagogical content knowledge, which also involves knowledge of students, pedagogy, and context (Ward, 2009).

Content-Specific Teaching Practices The teaching practices that are unique to or have a particular functioning and specialized manifestation when occurring in the teaching of specific subject matters (Hamre et al., 2013), without this implying that they cannot occur in the teaching of other disciplines as well.

Generic Teaching Practices The teaching practices that can be employed in teaching regardless of the subject matter that gets taught (Hamre et al., 2013; Rink, 1999), or the learning domain (e.g., psychomotor, cognitive, affective) that is targeted (Rink, 2010).

Student Learning A relatively permanent change in student behavior/skills which results from a process that targets to do so (e.g., teaching) (Rink, 2010). The present study explores student psychomotor leaning, namely the development of student psychomotor skills.

Conceptual Framework A model which encompasses variables that have been documented through empirical studies and theoretical research as significant for student learning (Danielson, 2007).

Teaching Quality Refers to both teacher characteristics (e.g., content-knowledge, teaching experience, in-service training) as well as teaching practices (Schacter & Thum, 2004; Singh & Sarkar, 2015). However, following Strong's (2011) recommendation, this study focuses only on teaching practices, since all other teacher characteristics that affect student learning are thought to be reflected in the practices employed by the teacher.

Teaching Practices Teacher actions (e.g., strategies, routines, moves) and interactions with students and the content, that aim to support student learning (Core Practice Consortium, 2014).

CHAPTER 2: THEORETICAL PERSPECTIVES

Abstract

Teaching quality plays a crucial role on what and how students learn, thus, during the last 40 years a persistent researcher effort has been undertaken to investigate the relationship between teaching quality in PE and student learning. However, this effort has mainly followed two distinct pathways. Following Carroll's and subsequent scholars' seminal works, some PE researchers have explored the effect of certain generic teaching practices on student psychomotor learning, believing that these practices could significantly contribute to student learning regardless of the subject matter that gets taught. Mostly by examining each generic teaching practice in isolation or in conjunction with few others, these researchers have underlined certain generic teaching practices as important for teaching psychomotor skills; yet, leaving unverified the contribution of some other generic practices to student psychomotor learning. Despite its significant input to the research field of PE, this pathway seems to ignore the particularities and differences of the PE aims, content and context from those of classroom disciplines, something that has been recognized by another group of PE researchers. This second group has focused on content-specific teaching practices, namely, teaching behaviors that have a particular functioning and specialized manifestation when occurring in the teaching of PE. Either by using comprehensive content-specific frameworks or by examining the effect of each PE-specific practice on student learning in isolation, these scholars have highlighted the significance of some discrete PE-specific practices; while not conclusive results were reported for some others. Recently few PE researchers have attempted to combine generic and content-specific teaching practices to develop more comprehensive observation instruments and to explore how these instruments may be associated with improved student achievement. However, these efforts are still at a nascent level and they have not explored the effects of teaching quality on student *learning*, something that the present study attempts to do.

Defining and Measuring Teaching Quality

There is consensus among all stakeholders that, although not being the sole contributor to student learning, teaching quality is a key determinant of what and how students learn (Cochran-Smith, 2003; Hattie, 2009), and one that deserves further scrutiny (Fenstermacher & Richardson, 2005). Regardless of the fact that there is not a unique definition of teaching quality, as it comprises several factors some of which are beyond the range of control of the teacher (Fenstermacher & Richardson, 2005), it could be argued that teaching quality refers to both teacher characteristics (e.g., content-knowledge, teaching experience, in-service training) as well as teacher actions and interactions with their students and the content during actual teaching in the classroom (Schacter & Thum, 2004; Singh & Sarkar, 2015). However, as Strong (2011) argues, when studying teaching quality, the main research emphasis should be on teacher actions and interactions with students and the content, since all other teacher characteristics that affect student learning are thought to be reflected in those actions and interactions.

Guided by this principle, researchers during the last 40 years have undertaken concerted efforts to more thoroughly examine teaching practices and understand how they might relate to student learning. A series of what are defined as process (i.e., teacher and student behavior) – product (i.e., student learning) studies have been conducted in the field of PE, and through the employment of different measurement approaches, these studies have identified several teaching practices to be related to student psychomotor learning (Rink, 2009).

Systematic observation is considered to be the foundation on which these research efforts have been built (Siedentop & Tannehill, 2000), as it offers a direct method to quantify the quality of teaching practice by measuring observable teacher actions (e.g., strategies, routines, moves) and interactions with students and the content (Core Practice Consortium, 2014; Joe, Tocci, Holtzman, & Williams, 2013). Once validated, observations can avoid many of the biases of self-report data (Strong, 2011) and can provide objective information that has strong internal validity (McKenzie & van der Mars, 2015). To stress the merit of this promising research method for measuring instruction, Goldhaber and Antony (2007) have wondered whether "it is even possible to judge teachers' effectiveness using measures other than direct observations of their teaching" (p. 134).

It is exactly for this reason that over the past four decades, PE scholars and researchers from other fields have developed several conceptual frameworks and/or associated classroom observation systems to explore the association between quality of teaching and student learning.

By conducting a closer examination to these attempts, one can notice that this research effort mainly followed two parallel perspectives, with some researchers attending, largely in isolation, to generic teaching practices, while others focusing on content-specific teaching practices.

The Two Parallel Research Perspectives

Generic Conceptual Frameworks

Generic teaching practices refer to basic behavioral dimensions of instruction that are relevant and observable *across* disciplines and whose frequent proficient enactment can lead to advancements in student learning **(**Hamre et al., 2013; Rink, 2003). In other words, these behaviors can be employed in teaching regardless of the subject matter that gets taught (Rink, 1999) or the learning domain (e.g., psychomotor, cognitive, affective) that is targeted (Rink, 2010), as their employment does not seem to be affected by the content or the kind of learning under consideration. For instance, take the teaching practice of increasing ALT for students (Siedentop et al., 1982). Regardless of the type of targeted learning objective, maximizing the time that students are appropriately engaged in subject-matter related tasks is expected to contribute to student learning. The rationale underpinning the attempt to identify common instructional practices that cut across different subject matters is nurtured by the belief that there are significant elements of effective teaching that are shared across domains (e.g., classroom and time management, orienting students toward learning goals, creating a pleasant classroom environment). These practices seem to mainly require general pedagogical knowledge (Rink, 2013), which is not informed by any content-specific considerations (Shulman, 1987).

Following this perspective, Carroll (1963) tried to develop a generic conceptual framework that could map effective teaching in all subject matters. His model of school learning is typically considered the starting point of modeling instructional effectiveness (Scheerens &

Bosker, 1997). Despite the fact that his model included "*quality of instruction*" as one of the five factors which were expected to explain variations in educational achievement, as Carroll (1989) later admits, this factor needed further elaboration. This expansion emerged later, through the development of several other models or principles of effective instruction (e.g., Brophy & Good, 1986; Creemers, 1994; Dunkin & Biddle, 1974; Rosenshine, 1983; Rosenshine & Stevens, 1986), which as many PE researchers admit (e.g., Rink, 2013; Silverman, 2011; Ward, 2013) have, in turn, influenced research in the field of PE. Despite their differences in structure, the majority of these models involved some common fundamental teaching practices (e.g., increasing student practice time, asking good questions, presenting the content in a clear manner, structuring the content, providing feedback), which influenced researchers in several fields and urged them to investigate the contribution of those practices to student learning in their own respective field.

During the last decade, a more systematic attempt has been undertaken to develop more comprehensive conceptual frameworks and classroom observation instruments incorporating generic teaching practices. This persistent work has led to the development of several generic frameworks, mainly used in "core" content-areas such as Mathematics and Language Arts (see Strong, 2011). Three notable examples of such theory-driven and evidence-based frameworks, which have been influential in the field of effective teaching, are briefly described below. First, frameworks that have been developed in Europe (*DMEE*) are presented; then, frameworks developed in the USA (*FfT* and *CLASS*) are considered. Despite the absence of studies utilizing these frameworks in exploring teaching quality in PE, each of the three selected frameworks (or modified versions of them) has theoretically or empirically been linked with PE teaching.

The first framework, the *Dynamic Model of Educational Effectiveness* (*DMEE*, Creemers & Kyriakides, 2008), is a multilevel model with factors contributing to student learning situated at four nested levels: student, classroom, school, and system. At the classroom level, which is the focus of this study, the model includes eight teaching factors, each of which can be measured along five different dimensions (see Table 2.1). The cross-tabulation of the eight teaching practices with the five dimensions yields forty different instructional aspects that are used to unpack the complex phenomenon of teaching. For example, the teaching practice of orientation is measured in terms of: a) the duration of orientation tasks that took place during a lesson (*frequency*), b) whether these orientation tasks referred to just a certain aim of the lesson, the entire lesson or even the unit/a series of lessons (*focus*), c) the juncture in the lesson (e.g., introduction, core, end of the lesson) during which this task occurred (*stage*), d) the quality of the orienting information (e.g., typical, related to lesson's aims, students specify the aims of the lesson) (*quality*), and f) the extent to which student background characteristics were taken into consideration when providing such information (*differentiation*).

Using the precursor of this multi-level model (i.e., Creemer's model, 1994) in PE, Kyriakides and Tsangaridou (2008) found that generic factors (e.g., time management) associated with student achievement in other content-areas (e.g., Mathematics and Language Arts) were also associated with achievement in psychomotor learning; leading to the conclusion that generic models of effectiveness could be used for evaluating the quality of PE teaching. Moreover, in other content-areas, teachers were found to be clustered into different stages according to the forty different instructional aspects of *DMEE*. Particularly, teachers clustered at lower stages were found to be less effective in promoting student learning compared to teachers situated at higher stages (Kyriakides, Archambault, & Janosz, 2013; Kyriakides, Creemers, & Antoniou, 2009).

Table 2.1

Description of the Eight Classroom Factors and the Five Dimensions of the Dynamic Model of Educational Effectiveness [a]

Teaching Practice	Dimension
Time management: is concerned with the organization and management of the classroom environment and learning time and thereby the maximizing of instructional time and student engagement.	*Frequency*: refers to how often and how long a teaching practice is present in a classroom.
Classroom learning environment: involves five elements: a) teacher-student interactions, b) student-student interactions, c) student treatment by the teacher, d) competition between students, and e) classroom disorder. Specifically, it refers to the establishment of on-task behavior through teacher-student or student-student interactions. It is also concerned with teachers' attempts to create an efficient and supportive environment for learning in the classroom (e.g., establishing rules and persuading students to respect and abide by the rules).	*Focus*: concerns the purpose(s) for which an activity takes place and the specificity of activities.
Structuring: concerns teacher's structuring of the tasks and materials to facilitate students' memorizing of information and learning (e.g., beginning a lesson with overview and/or review of objectives, outlining the content to be covered, signaling transitions between lesson parts, calling attention to main ideas, reviewing main ideas or providing summary reviews at the end of the lesson). It also examines whether instructional tasks' or lessons' difficulty level is gradually increased.	*Stage*: relates to the phase of the lesson (e.g., introduction, core, end of the lesson) that a teaching practice takes place, assuming that teaching practices need to take place over a long period of time to ensure their effects on student learning.
Application: refers to the opportunities provided to students to practice and apply the new information through seatwork or small-group tasks. It is also concerned with the number and the difficulty/complexity level of application tasks given to students and whether these tasks are used as starting points for the next step of teaching and learning.	
Questioning: comprises five elements: a) the type of questions (i.e., mixed of product and process questions) posed by the teacher, b) the wait time after questions are posed, c) the clarity of the questions, d) the appropriateness of the difficulty level of the questions, e) the way teachers deal with student responses to questions.	*Quality*: takes into account that the functioning of a teaching practice may vary, and thus, it captures the properties of a specific teaching practice, as these are discussed in relevant literature.
Teaching modeling/Learning strategies: is concerned with the extent to which teachers help or encourage students to use strategies and/or develop their own strategies that can help them solve different types of problems.	
Orientation: relates to providing the importance/objectives for which a specific task or lesson or series of lessons take(s) place and/or challenging students to identify the reason(s) for which the lesson involves a particular activity. Orientation process has the potential to make tasks and lessons meaningful to students and encourage their active participation in the learning process.	*Differentiation*: pertains to the extent to which teaching practices are adapted to the specific needs of (groups of) students.
Classroom assessment: refers to teacher behaviors that aim to collect information about student knowledge and skills, identify student needs and evaluate teacher's own practice. It is also concerned with the extent to which this information is used for formative rather than summative purposes, and whether parents are informed.	

Note. [a] Based on Creemers and Kyriakides (2008).

The second framework, the *Framework for Teaching* (*FfT*, Danielson Group, 2013), aims to define what teachers should know and be able to do in pursuing the work of teaching. By doing so, it divides teaching into four main domains (i.e., planning and preparation, the classroom environment, instruction, and professional responsibilities) with 22 components overall. Each component is explicitly described and involves a different number of elements, which serve to further define the component; yielding 76 total elements of teaching quality. For example, as can be seen from Table 2.2, when rating the “classroom environment” domain, the observer should assign a score (i.e., unsatisfactory, basic, proficient, distinguished) for each of the five components under consideration (i.e., creating an environment of respect and rapport, establishing a culture for learning, managing classroom procedures, managing student behavior, organizing physical space), having in mind the discrete elements that each component involves.

Despite not used in the field of PE, this instrument was part of recent scholarly discussions. Particularly, Rink (2013) underlined that the domains and components of the *FfT* can certainly be employed for evaluating PE lessons, cautioning, however, for the descriptors being used. As she warns, if *FfT* is to be used with any validity in PE, the descriptors and examples of this generic instrument need to be suitably adapted for PE lesson conditions. Such an attempt was undertaken by the stakeholders in Singapore who developed the *Physical Education Lesson Observation Tool* (*PELOT*), by adapting *FfT*’s descriptors to be suitable for PE lessons. However, no studies have been found employing this instrument to assess PE lessons. In other content-areas, a series of studies showed small to moderate correlations between scores obtained through *FfT* and student achievement gains (Strong, 2011).

Table 2.2

Outline of the Two out of Four Domains of the Framework for Teaching Model [a]

Domain	Component	Element
The Classroom Environment	- Creating an environment of respect and rapport	• Teacher interactions with students, including both words and actions • Student interactions with other students, including both words and actions
	- Establishing a culture for learning	• Importance of the content and of learning • Expectations for learning and achievement • Student pride in work
	- Managing classroom procedures	• Management of instructional groups • Management of transitions • Management of materials and supplies • Performance of classroom routines
	- Managing student behavior	• Expectations • Monitoring of student behavior • Response to student misbehavior
	- Organizing physical space	• Safety and accessibility • Arrangement of furniture and use of physical resources
Instruction	- Communicating with students	• Expectations for learning • Directions for activities • Explanations of content • Use of oral and written language
	- Using questioning and discussion techniques	• Quality of questions/prompts • Discussion techniques • Student participation
	- Engaging students in learning	• Activities and assignments • Grouping of students • Instructional materials and resources • Structure and pacing
	- Using assessment in instruction	• Assessment criteria • Monitoring of student learning • Feedback to students • Student self-assessment and monitoring of progress
	- Demonstrating flexibility and responsiveness	• Lesson adjustment • Response to students • Persistence

Note. [a] For space purposes this table presents only two domains of the *FfT*, which are directly observable during the teaching process. For more information about all four domains and 22 components see Danielson Group (2013).

Finally, the third framework, the *Classroom Assessment Scoring System* (*CLASS*, La Paro, Pianta, & Stuhlman, 2004), was originally designed to observe and assess emotional and instructional elements of quality in early childhood educational environments. In recent years, the framework has been expanded to tap instructional quality in elementary school and middle-grades. Particularly, the latest version of this model focuses on three domains of classroom characteristics (i.e., emotional climate, management, and instructional support), each having a different number of distinct components, creating 13 overall teaching components that are used to capture instructional quality (see Table 2.3). In addition, a fourth domain (i.e., student outcomes) that refers to student engagement is also scored separately from the three domains.

In the field of PE, this instrument was used by Ko (2008) to describe elementary physical educators' teaching practices and to provide insights into why teachers choose these practices in their lessons. However, despite the fact that the CLASS allowed for distinguishing teachers into two distinct profiles depending on the teaching practices they used (i.e., profile A involved teachers who scored high in both management and student outcome domains, while profile B included teachers who scored low in emotional climate, management and instructional support domains), Ko concluded that the instructional support domain seemed to have a gap in determining the quality of instruction in PE, calling for further investigation of the feasibility of this domain within PE lessons. Interestingly, Ko's conclusion seems to resonate with Rink's (2013) abovementioned concern about the descriptors included in generic instruments. Yet, as was the case with the other two generic instruments, CLASS dimensions were found to be significantly related with students' academic outcomes in other content-areas (Allen et al., 2013; Mashburn et al., 2008).

Table 2.3

Description of the Three Domains and Thirteen Teaching Components of the latest version of CLASS [a]

Domain	Teaching Component
Emotional Climate	- Positive climate: relates to teacher's enthusiasm, enjoyment and emotional connection with the students and the nature of students interactions.
	- Negative climate: refers to evidence of anger, hostility, or aggression that the teacher and/or students exhibit in the classroom.
	- Sensitivity: pertains to teacher actions showing the degree of his/her responsiveness to students' academic and emotional needs as well as the establishment of a secure environment within which students can be voluntarily engaged in.
	- Regard for student perspectives: concerns the extent to which the activities as well as teacher's interactions with students take into consideration student's interests, motivations and points of view.
	- Overcontrol: is concerned with teacher's flexibility related to students' interest and classroom schedules and the extent to which teacher encourages autonomous behavior in students.
Management	- Behavior management: concerns the degree to which the teacher monitors, prevents and redirects student behavior.
	- Productivity: relates to how well students understand and implement classroom routines as well as the degree to which teacher ensures maximum time in productivity learning activities.
	- Instructional learning formats: is concerned with the way teacher engages students in activities and facilitates activities to maximize learning opportunities.
	- Classroom chaos: is related with teacher's inability to manage children misbehavior, leading to chaotic environment.
Instructional Support	- Concept development: refers to the extent to which the teacher promotes higher-order thinking and problem solving.
	- Quality of feedback: is related with teacher's responses that aim at extending students' learning and understanding.
	- Language modeling: concerns the quality and quantity of teacher's use of language stimulation and language-facilitation techniques during interactions with students.
	- Richness of instructional methods: refers to the use of strategies that promote student's deeper and more complex thinking and understanding of material.

Note. [a] Based on the original CLASS from La Paro, Pianta, & Stuhlman (2004) as well as the latest version from Hamre et al. (2013).

Comparing the abovementioned frameworks, one can notice that despite the differences in their structure and the descriptors they use, all three frameworks include several common teaching practices (e.g., establishing a positive classroom climate to facilitate learning, time management, dealing with students' responses), whose contribution to student psychomotor learning has been verified through several past and recent research works, as described below.

Research Findings on the Contribution of Generic Teaching Practices to Student Psychomotor Learning

Despite the fact that research on teaching in PE has followed the lead of classroom research in general (Gusthart & Sprigings, 1989; Metzler, 2014; Rink, 2010), comprehensive generic observational frameworks, such those described above, were scarcely used by PE scholars to examine the relationship between instruction and psychomotor learning. However, researchers in the field of PE have investigated the effects of discrete generic teaching practices on student psychomotor learning through several process-product studies. The findings of these studies presented below and summarized in Table 2.4 are organized around the eight factors of the *DMEE*, the most up-to-date multilevel model of educational effectiveness (Scheerens, 2013), and the one selected to be used in this study.

Time management. As previously mentioned, time management includes practices such as maximizing student time on task, minimizing transition time and finishing the lesson on time. In other content-areas, this factor has been found to be among the most strongly related to student learning outcomes (Muijs et al., 2014; Muijs & Reynolds, 2003). In the field of PE, studies investigating time management mainly focused on teachers' work to maximize the academic learning time (ALT). Particularly, capitalizing on the ALT construct used in the classroom setting, Siedentop et al. (1982) designed the corresponding instrument for measuring time management in PE (ALT-PE). As van der Mars (2006) reports in his review chapter on time and learning in PE, the ALT-PE or variations thereof used in several studies with methodological differences, showed a consistent positive relationship between amounts spent by students on ALT and their psychomotor learning. These findings were corroborated by more recent studies showing that more effective teachers maximize time devoted to student motor-involvement at an appropriate level of difficulty with activities relevant to the goals of the their lessons, compared to their less effective counterparts (Kyriakides & Tsangaridou, 2008).

In addition, other studies exploring time management in PE focused on the aspect of minimizing transition and waiting time through the use of effective routines. As early as 80s, in analyzing the structure of tasks in PE lessons, Tousignant and Siedentop (1983) added a task that served the function of transitioning students from point to point. Subsequent studies, examining transitions and student waiting in lines in PE lessons, found that each of transition and waiting time accounts for 15-25% of class time (Rasmussen, Scrabis-Fletcher, & Silverman, 2014; Siedentop & Tannehill, 2000). Unsurprisingly, then, researchers deemed the use of effective transitional and organizational routines as imperative for effective teaching of PE lessons. Transitional and organizational routines consist of predefined teacher actions, which are established by the teacher, to effectively manage waiting time or the time spent on transitions among managerial or instructional tasks (e.g., student grouping strategies, distribution of equipment) (Rink, 2010). Such routines were examined through a series of studies, showing that effective teachers establish and use transitional routines from the beginning of the school year (Fink & Siedentop, 1989), and that less effective teachers have greater wait time than more effective (Constantinides, Montalvo, & Silverman, 2013; Hickson & Fishburne, 2004).

The classroom as a learning environment. Rendering classroom an efficient and supportive learning environment is another generic teaching practice that teacher effectiveness research has consistently found to distinguish between more and less effective teachers (Muijs & Reynolds, 2001). This factor comprises two types of instructional aspects. First, it involves teaching practices pertain to classroom management, such as treating all students fairly or minimizing disruptive behavior (Creemers & Kyriakides, 2008). Second, it encompasses teacher-student (e.g., adapting instructional tasks, building interpersonal rapport, providing the appropriate amount of feedback, motivating students) and student-student interactions, which are important components of classroom climate (Li, 2015). Although not directly related to the content to be taught, establishing and maintaining an orderly and effective learning environment, creates the necessary conditions for effective teaching and learning (Rasmussen et al., 2014; Rink, 2013; Siedentop & Tannehill, 2000).

Research findings showed that many of the characteristics of effective management in the classroom hold true for the PE environment as well, and they have been identified as critical components of effective teaching. For instance, communication and establishment of clear rules, routines, and expectations for behavior from the first days of the school year, have been found to ensure smooth operation of the classroom and minimal disruptive behaviors throughout the school year (Fink & Siedentop, 1989; Hickson & Fishburne, 2004; Jones, 1992; O' Sullivan & Dyson, 1994; Siedentop, 2002). These routines can save valuable teaching time, as in some cases administration/class management time was found to be higher than 20% of total class time (Bevans et al., 2010; Kelder et al., 2003; Simons-Morton, Taylor, Snider, & Huang, 1993). Similarly, researchers have found that teachers' behaviors intended to hold students accountable were positively related with increased levels of students' physical activity (Schuldheisz & van der Mars, 2001), and ALT-PE (Hastie, 1994).

As far as the teacher-student interactions are concerned, these have been systematically measured through the use of the Cheffers' Adaptation of Flanders' Interaction Analysis System (CAFIAS) (Cheffers & Mancini, 1989), an observational instrument that allows researchers to measure both verbal and nonverbal teacher and student interactions. Several studies employing this tool found that teachers mainly used direct methods of teaching, and they were not actively engaging students in the learning process (e.g. Anderson & Barrette, 1978; Cheffers & Mancini, 1989). In addition, research in PE showed that students' perceived motivational climate can influence their conception of ability, which in turn positively impacts on student's effort (Li & Lee, 2004). In other words, the quality of teacher-student and student-student interactions can contribute to better engagement and learning among students in PE (Li, 2015).

However, as far as the element of feedback is concerned, research findings in PE have been inconsistent, not always supporting feedback as an essential element of teaching that promotes psychomotor learning (Lee, Keh, & Magill, 1993). A plethora of research studies have investigated this relationship, with some reporting that more effective teachers typically provide more feedback than less effective teachers (e.g., Phillips & Carlisle, 1983), while others showing a non-significant direct relationship (e.g., Silverman, Tyson, & Krampitz, 1992). What complicates matters more is that what appears to be the right amount of feedback to promote learning for one student might inhibit learning for another (Lee et al., 1993). This inconsistency in research findings may be due to two main reasons. First, feedback is a broad and multidimensional concept; hence, deciding what it is about feedback that influences student psychomotor learning is hard. Second, these different results are partly due to different methodological approaches pursued to study this relationship (Silverman, 1994). For instance, early studies focused on experimental teaching units, usually with novice teachers, which lasted for very short periods of time, often not long enough to obtain a reliable sample of behavior, while several other studies did not control for student skill level and number of practice trials (Lee et al., 1993), both of which have been found to have a significant impact on student psychomotor learning.

As a consequence, to effectively and reliably study if and how teacher feedback contributes to student psychomotor learning, a clear and explicit description of the specific components under investigation should be made by researchers, who should ensure that the pursued research methods are appropriate. Having in mind the abovementioned results, and thinking that the effectiveness of feedback may be subject-matter specific for PE (Silverman et al., 1992), some researchers focused on certain aspects of feedback (e.g., congruency and specificity), something that is discussed when considering content-specific teaching practices.

Structuring of tasks and materials. Another generic teaching practice concerns the way in which teachers structure the tasks and materials to support student learning. This teaching practice refers to two elements of teacher behavior: a) facilitating memorizing of information (e.g., by repeating main ideas), promoting recognition of the relationships between parts and understanding how these parts create an integrated whole (e.g., by providing summary reviews at the end of the lesson or connecting previous lessons to the lesson of the day); and b) gradually increasing the difficulty level of the assigned tasks during the lesson or series of lessons (Creemers & Kyriakides, 2008). In the field of PE teaching, scholars have emphasized the importance of including both elements of this teaching practice in every PE lessons. However, although in classroom effectiveness research both elements were found to be important for student learning (Brophy & Good, 1986; Hattie, 2009; Rosenshine & Stevens, 1986; Scheerens & Bosker, 1997), in PE only the latter aspect has been adequately supported through research findings.

Particularly, sequencing practice in progressive levels of difficulty has been shown to significantly enhance psychomotor learning and motivation, especially for low-skill students, given that task difficulty was appropriate for the learner (French et al., 1991; Hebert, Landin, &

Solmon, 2000). In both studies under consideration, scholars concluded that students need to practice simpler variations before practicing a final, complex task, as this helps them develop the necessary prerequisite skills before practicing the more difficult task.

On the other hand, as far as the former element of the structuring factor is concerned (i.e., structuring comments) further research is needed to better understand its effect on student psychomotor learning, since the review of studies examining the contribution of this element yielded only two research works: the early study of Silverman, Tyson, and Morford (1988) and the recent study of Rasmussen et al. (2014). Specifically, Silverman et al. (1988) found a relationship among teacher's communication of what was going to occur during the lesson or provision of reviews of what occurred and student psychomotor achievement. Similarly, Rasmussen et al. (2014) found a positive and significant relationship between the introduction to a lesson and the amount of appropriate and total practice trials students received. Trying to understand the above relationships, Rasmussen et al. (2014) claimed that the structuring comments framed subsequent instruction, helped students focus on the impending tasks, and prepared them to achieve success. In addition, these researchers found that teachers spent an average of about 12% of lesson time in closure activities and in reviewing the day's events and objectives.

Nevertheless, despite the scarce empirical evidence that confirm the impact of the above element on student psychomotor learning, scholars deemed structuring practices to be particularly important in PE. As they argue this teaching practice can help students remember the main learning cues/critical aspects of the emphasized skill and understand how the skill can be later used in a game situation or in other contexts, thus making practice more meaningful. As Siedentop and Tannehill (2000) and Metzler (2011) argue, closure should be an integral part of every PE lesson, as it provides an opportunity to the teacher to not only bring together the parts of the lesson, but also make sure students understand the key ideas of the lesson, and to assess their feelings about the lesson.

Application. Turning to application, *DMEE* defines this teaching factor as the opportunities provided to students to practice and apply the emphasized skills. Moreover, it explores whether these practice opportunities have a progressively difficulty level. However, this latter part of the definition seems to overlap with the structuring element of gradually increasing the difficulty level of lesson activities (discussed above), thus the emphasis here will be paid only on the former element.

Through the examination of this instructional aspect in PE, researchers confirmed the results of classroom effectiveness research. Specifically, more effective teachers were found to provide more and a larger variety of instructional tasks during their PE lessons than the least effective (Constantinides et al., 2013; Rasmussen et al., 2014; Silverman, Subramaniam, & Woods, 1998). As Silverman and Mercier (2015) explain, when teachers change instructional tasks more frequently, students get more on-task and appropriate practice, which equals to increased total practice trials, which in turn lead to student psychomotor development.

Teacher questioning. Teacher questioning is a generic teaching practice that only recently has become an integral part of the effective PE teacher's repertoire (Metzler, 2011). Despite that effective questioning is one of the most widely studied aspects of teaching in classroom effectiveness research (Muijs et al., 2014), it has received inadequate research interest in PE, apparently because PE researchers considered this teaching practice more pertinent to teaching cognitive rather than psychomotor outcomes. Questioning refers to posing different types of questions (i.e., process and product) in a clear manner and at appropriate difficulty levels and giving students time to respond. In addition it involves the identification and remediation of student misconceptions and deal with student responses (Creemers & Kyriakides, 2008).

Yet, beyond some research efforts that provided some descriptive statistics regarding the use of questioning techniques (e.g., Thorburn & Seatter, 2015; Zeng, Leung, Liu, & Hipscher, 2009), PE research has only examined the impact of some of those elements when it comes to tactical awareness teaching—that is the ability to select the appropriate responses to solve tactical problems that arise during a game. Within this research field, the practice of questioning has received great research attention and it has been underlined as a critical teaching skill (Gubacs-Collins, 2007; McNeil, Fry, Wright, Tan, & Rossi, 2008; Pearson & Webb, 2008), without however being correlated with student psychomotor learning. As Metzler (2011) argues, when teachers employ different types of questions and offer adequate wait time for every student to come up with an answer, they can promote learning not only in the psychomotor domain, but in cognitive domain as well. However, as far as the development of psychomotor skills is concerned, the robustness of questioning practice needs empirical validation.

Teaching modeling/Learning strategies. Modeling, as defined in the *DMEE*, concerns the degree to which teachers encourage students to use strategies and/or develop their own strategies to solve different problematic situations (Creemers & Kyriakides, 2008). This teaching practice seems to be more consonant with constructivist theories, teaching higher-order thinking skills and problem-solving. Hence, the emphasis is on the development of student skills that will help them organize their own learning (i.e., self-regulation) and on the achievement of the new goals of education such as the development of meta-cognitive skills (Panayiotou et al., 2014).

As was the case with questioning, PE researchers have investigated the impact of the abovementioned elements of modeling on student psychomotor development mainly via the Teaching Games for Understanding approach (Rovegno & Dolly, 2006). This approach engages students in games that are constructed so that a tactical problem would emerge for the students, who are then encouraged by the teacher to solve the problematic situation. Teaching students via this approach found to have a positive impact not only on students' psychomotor development (e.g., Alison & Thorpe, 1997; Harrison et al., 2004; Turner & Martinek, 1999), but also on students' tactical awareness (e.g., Chatzipanteli, Digelidis, Karatzoglidis, & Dean, 2014; Nevett, Rovegno, & Babiarz, 2001; Nevett, Rovegno, Babiarz, & McCaughtry, 2001), and it was found to be as effective as the skill teaching approach (Harrison et al., 2004).

There are, however, and a few notable research works which explored the impact of modeling practice via metacognitive prompting on students' psychomotor development. Both Lidor's (2004) and Chatzipanteli and Digelidis's (2011) studies concluded that students' performance was positively and significantly influenced when teachers provided students with learning strategies. Similarly, in their review article on modeling and demonstration, Hodges and Franks (2002) argued that in the early stage of acquisition, it may be more successful to teach participants via an implicit and discovery learning method instead of using explicit learning strategies.

Orienting students toward learning goals. Influenced once again by the findings of classroom effectiveness research, PE scholars have postulated the generic practice of orientation to be an effective teaching principle when delivering PE lessons. Having in mind that classroom effectiveness research emphasized the role of teacher in explaining the reason and making explicit the importance of engaging students in certain activities as well as in providing opportunities to students to identify the merit of engaging in lesson activities (Brophy & Good, 1986; Creemers & Kyriakides, 2008), PE scholars considered this teacher role suitable for PE lessons as well. As constructivist theories support, when the goals of the lesson are made clear to the learners, then, not only the instruction and student practice become highly focused on the emphasized skills, but also learning becomes meaningful (Siedentop & Tannehill, 2000). However, as a PE research community we still lack empirical evidence to support to what extent and in what particular ways this teaching practice contributes to student psychomotor learning.

Student assessment. Although the contribution of student assessment is considered crucial in other subject matters (De Jong, Westerhof, & Kruiter, 2004; Kyriakides, 2005), in PE it represents another teaching practice that was not adequately explored. As Rink (2010) mentions, student assessment can serve two purposes, the formative and the summative purpose. The formative purpose is considered more beneficial to student learning rather than the summative (i.e., simply attributing final grades to students), as it includes the use of appropriate techniques to collect data on student knowledge and skills and the use of these results to inform students about their progress as well as teacher's own teaching (Siedentop & Tannehill, 2000). Although assessment in PE is generally accepted by scholars as an important teaching practice for student psychomotor learning (e.g., Kniffin & Baert, 2015; Matanin & Tannehill, 1994; Nye, Dubay, Gilbert, & Wajciechowski, 2009), research on the impact of assessment (and especially its formative purpose) on student psychomotor learning is relatively limited (Hay, 2006). This cannot be dissociated, however, from the fact that assessment is hardly performed in PE (Wright & van der Mars, 2004), as teachers—especially elementary school teachers—find assessment the most difficult aspect of their role and the area in which they feel the least competent (Morgan & Hansen, 2007).

Despite the limited research outcomes on this issue, some research studies regard assessment as an important teaching practice for student psychomotor learning, without, however, measuring its contribution to student learning. For example, Mintah (2003) found that PE teachers perceived the use of assessment to positively enhance the self-concept, motivation, and skill achievement of their students, although reporting that they spent more time in planning to incorporate this element in their teaching. In another study, MacPhail and Halbert (2010) found that when assessment was used, teachers and students strongly claimed that student learning in PE was improved. Similar findings were reported by Ni Chroinin and Cosgrave (2013), who found that the teachers participating in their research believed that the inclusion of assessment in PE lessons impacted positively student learning. What is obvious from the above work, however, is that student learning was not measured; rather it was just based on self-reports. Thus, more empirical validation of this teaching practice is needed through studies that will use observational data to measure student assessment and pre- and post-tests to measure student learning.

Summary. Taking into account all the above research findings concerning the contribution of generic teaching practices to student psychomotor learning, four major observations can be made. First, comprehensive frameworks involving generic teaching practices have scarcely been used to examine the contribution of these practices to students' psychomotor learning. Although the practices incorporated in these frameworks seem to be appropriate for evaluating PE lessons, researchers seem to avoid using them in PE research, because the descriptors incorporated in these frameworks might have a different meaning in PE than in other disciplines (Metzler, 2014). As Strong (2011) admits, due to the fact that generic frameworks have been largely designed to measure the quality of "classroom" subjects (e.g., Mathematics or

Language Arts), they may suffer from being too general and insensitive to the particular demands of other disciplines and fail to asses strategies that are specific to a given content-area. For example, as Rink (2013) underlines, what is good grouping and management in, say, Mathematics, is not necessarily good grouping and management in PE; similarly, what is considered an effective teaching practice in English Language Arts, might not be as effective in PE. Why this might be the case has not only to do with the fact that definitions of teaching effectiveness can never be content-free (Ward, 2013), but also with the instructional particularities of PE context, which are further explained in the next section.

Second, as it is obvious from the above research findings, some generic teaching practices (e.g., time management, rendering classroom an efficient learning environment) seem to have a significant impact on student psychomotor learning. Third, the contribution of other generic teaching practices to student psychomotor learning needs to be verified by more research studies in the field of PE, which will corroborate existing research findings. For instance, although deemed as important teaching practices, orientation, student assessment and some elements of structuring have not been adequately investigated and correlated with student psychomotor development. Despite highlighted as important teaching practices by classroom effectiveness research, as several PE scholars warn (Gusthart & Sprigings, 1989; Silverman et al., 1988, 1992), the contribution of generic teaching practices obtained from other disciplines should be tested in the PE environment, since such transfer of results could be problematic, given the different objectives the PE targets and the particularities of PE context.

Fourth, what largely seems to be missing from the above research findings is a close consideration of the content of PE and the requirements that teaching PE imposes on teachers. This, in fact, was a gap that different scholars identified for generic teaching practices at least three decades ago (e.g., Shulman, 1986; Stodolsky, 1988), urging researchers to attend to the disciplinary demands of teaching certain subject matters. As Bransford et al. (2000) contend, the interplay between content knowledge and pedagogical knowledge contradicts the “one size fits all” models and the commonly held misconception that effective teaching consists of a set of general teaching practices which apply to all disciplines. This becomes particularly true, when considering that the knowledge, skills, abilities, and attitudes students are expected to develop in various subjects are quite different, requiring teachers to employ different kinds of instructional practices to achieve these various goals (Stodolsky, 1988; Stodolsky & Grossman, 1995).

Table 2.4

Summary of the Studies Exploring the Contribution of Generic Teaching Practices to Psychomotor Learning [a]

Generic Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Main Results
Time management	Academic Learning Time (ALT-PE)	Kyriakides & Tsangaridou, (2008)	Elementary: 4th grade T: 49 generalists St: 1142	Three 40-minute live classroom observations per teacher were conducted and analyzed using a revised version of the ALT-PE system.	Pre- and post-criterion-reference performance test (Locomotor, non-locomotor, and manipulative skills)	• Time devoted to student motor-involvement at an appropriate level of difficulty with activities relevant to the goals of lesson positively impacted student learning. • Time devoted to classroom management, transitions and waiting negatively impacted on student learning.
		van der Mars (2006)	Review study. [b]			• A consistent positive relationship was found between amounts spent by students on ALT-PE and their psychomotor learning.
The classroom as a learning environment	Teacher-Student, Student-Student Interactions (Providing feedback)	Hastie (1994)	Secondary: 10th grade T: 3 specialists St: 3 classes (class size averaged 26 students)	Ten 40-minute videotaped classroom observations per teacher were conducted and analyzed using the Anderson's Physical Education Teachers' Professional Functions observational system.	Students' Academic Learning Time captured with the second version of the ALT-PE system (Volleyball unit)	• Teacher feedback intended to influence student performance and teacher intervening interactions were associated with higher amounts of ALT-PE.
		Lee, Keh, & Magill (1993)	Review study. [b]			• The findings are inconsistent and do not always support feedback as an essential element of psychomotor learning. Teacher feedback has not always been related to increased achievement. However, more effective teachers typically provide more feedback.
		Schuldheisz & van der Mars (2001)	Secondary: 7th grade T: 1 specialist St: 8 "low active" target students	Experimental Design. Four 40/45-minute lessons. Event and interval recording of two supervisory conditions: a) passive supervision b) active supervision	Students' accumulated moderate to vigorous physical activity captu-red with live use of System for Observing Fitness Instruction Time (SOFIT) (Basketball unit)	• Encouragement, prompts, and feedback provided by the teacher directly affected target students' moderate to vigorous physical activity levels.

Table 2.4 Continued

Generic Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Main Results
Structuring of tasks and materials	Sequencing practice in progressive levels of difficulty	French et al., (1991)	Secondary: 9th grade T: - St: 53	Experimental Design. Four groups: a) *Final-test*: practiced the actual skills, b) *Progression*: practiced a fixed number of trials in a specific practice progression, c) *Criterion*: was required to obtain an 80% success rate at the first and second level of difficulty in the practice progression, d) *Control*: did not practice the emphasized skills.	Pre- and post-performance test (AAHPERD-Volleyball serve and set skills)	• The *Progression* and *Criterion* groups had higher post-test scores, suggesting that sequencing practice in progressive levels of difficulty enhances psychomotor learning when task difficulty is appropriate for the learner.
		Hebert, Landin, & Solmon (2000)	Tertiary education T: 4 proficient tennis players and experienced instructors St: 81	Experimental Design. Three groups: a) *Part-to-whole*: practiced four tasks of increasing complexity of the serving motion, b) *Extension*: practiced four tasks of increasing difficulty of achieving the outcome goal of the serve, c) *Criterion*: practiced only the final skill.	Success (outcome) and appropriateness (correct serving motion) of student practice trials (Tennis serving skill)	• Easy-to-difficult task progressions positively affected success and appropriateness of student practice trials. • Low-skilled students benefited from both part-to-whole and extension groups of practice.
	Using structuring comments	Rasmussen, Scrabis-Fletcher, & Silverman (2014)	Elementary: 3rd grade T: 10 (not specified) St: 90 (nine randomly selected students in each class)	Two 30-minute videotaped classroom observations (indoor lessons) per teacher were conducted and analyzed using duration and event-recording instruments for both teacher and student behaviors.	Appropriate/ Successful and inappropriate/unsuccessful practice trials (variety of activities)	• A positive significant relationship was found among the initiation-introduction component of the lesson and the amount of appropriate practice and total practice trials.
		Silverman, Tyson, & Morford (1988)	Secondary: 6th, 7th and 8th grade T: 7 specialists St: 202	Seven consecutive regularly scheduled videotaped classes were coded using a systematic observation instrument that determined the amount of time spent in various categories of class organization.	Pre- and post-performance test (AAHPERD-Volleyball forearm pass and serve skills)	• Teacher communication of what will occur during the lesson or reviews of what occurred were related to forearm pass achievement.
Application	Number and variety of instructional tasks provided to students	Constantinides, Montalvo, & Silverman (2013)	Elementary: 4th grade T: 20 (10 specialists and 10 generalists) St: 180 (nine randomly selected students in each class)	Two 40-minute videotaped indoor lessons per teacher were conducted and analyzed using two coding instruments capturing overall six categories.	Appropriate/ Successful and inappropriate/unsuccessful practice trials (indoor lessons-not specified)	• A significant positive relationship was found between the number of tasks and the number of student appropriate (and total) practice trials.

Table 2.4 Continued

Generic Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Main Results
Application (continued)	Number and variety of instructional tasks provided to students (continued)	Rasmussen, Scrabis-Fletcher, & Silverman (2014)	Elementary: 3rd grade T: 10 (not specified) St: 90 (nine randomly selected students in each class)	Two 30-minute videotaped classroom observations (indoor lessons) per teacher were conducted and analyzed using duration and event-recording instruments for both teacher and student behaviors.	Appropriate/ Successful and inappropriate/unsuccessful practice trials (variety of activities)	• A strong positive relationship was found between the number of tasks and the amount of student appropriate (and total) practice trials.
		Silverman, Subramaniam, & Woods (1998)	Secondary: 7th, 8th and 9th grade T: 8 specialists St: 72 (three target students at each skill level--i.e., low, medium, high--per teacher)	Two 30-minute videotaped classroom observations (indoor lessons) per teacher were conducted and analyzed using a task instrument.	Appropriate/ Successful and inappropriate/unsuccessful practice trials (variety of activities)	• The number of tasks correlated with low-skilled students' appropriate (and total) practice trials.
Teacher questioning	-	-	-	-	-	-
Teaching modeling/ Learning strategies	Using Teaching Games for Understanding approach (TGfU)	Alison & Thorpe (1997)	Secondary: 8th & 9th grade T: 2 specialists St: 96	Experimental Design. Two groups per teacher: a) skill-based teaching approach b) Teaching Games for Understanding approach	Pre- and post-performance test (AAHPERD-Basketball passing and shooting accuracy skills and Henry-Friedel Field Hockey test-shooting accuracy & speed skills)	• TGfU groups improved more than the skill-based groups in all areas of skill development.
		Harrison et al. (2004)	Tertiary education T: 3 (had previously playing volleyball experience) St: 182	Experimental Design. Two groups per teacher: a) skill-based teaching approach b) Teaching Games for Understanding approach	Pre-, midterm- and post-performance test (AAHPERD-Volleyball set-up, passing and serving skills and Stanley spike test)	• Neither teaching approach was superior. A significant improvement on skills tests was found for both groups.
		Turner & Martinek (1999)	Secondary: 6th & 7th grade T: 2 specialists St: 71	Experimental Design. Three groups (two groups per teacher and a control group): a) technique teaching approach b) Teaching Games for Understanding approach c) control group	Pre- and post-performance test (Henry-Friedel Field Hockey test-dribbling, ball control, tackling, evading an opponent and shooting skills)	• TGfU groups scored significantly higher on control and passing execution than the technique teaching approach and the control group during post-test game play. However, no significant differences in hockey skill development (accuracy & speed) were found among the treatment groups.

Table 2.4 Continued

Generic Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Main Results
Teaching modeling/Learning strategies (continued)	Meta-cognitive prompting	Chatzipa-nteli & Digelidis (2011)	Secondary: 7th grade T: 2 specialists St: 236	Experimental Design. One group. Two trials per student (each trial included five consecutive attempts). Before first trial, students were directed to pay attention to the shooting act-to feel the movements. During the second trial, each time students performed a service, they had to answer a written questionnaire designed for enhancing meta-cognitive activity.	Performance test (AAHPERD-Volleyball serving skill). Five consecutive attempts per student	• The findings indicated a statistically significant difference between the scores of the second trial and those of the first trial, with second trial's scores being higher. The findings imply that metacognitive question prompts guided students to assess themselves, think about their failed executions and modify their strategy and performance to execute more successfully the skill.
		Lidor (2004)	Secondary: 7th grade T: 1 specialist St: 56	Experimental Design. Four groups: a) *The 5-SA* group was instructed to apply five sequential steps: readying, imaging, focusing attention, executing, and evaluating, b) *The awareness group* was directed to pay attention to the shooting act, to feel the movements, and to use kinesthetic cues that were related to the shooting task and the performance environment, c) *The non-awareness group* was taught to preplan the shooting motion, to focus attention on the most related to the task specific cue and the environment, and to let the movement flow, d) *The control group* was not exposed to any strategy, but was provided with additional technical explanations on shooting and foul shooting.	Pre- and post-performance test (Basketball-free throw shooting skill) (two blocks of ten free throw shots during pre-test and three blocks of ten free throw shots during post-test)	• It was found that the 5-SA group and the non-awareness groups performed more accurately than the awareness and the control groups. The findings imply that learning strategies can facilitate accuracy of performance when applied during PE classes.
Orienting students toward learning goals	-	-	-	-	-	-
Student assessment	-	-	-	-	-	-

Notes. [a] The above review table is not exhaustive.
[b] Where review articles are presented, studies that preceded those reviews are not mentioned in the table.

Content-Specific Conceptual Frameworks

Acknowledging that PE differs from other subject matters in terms of the aims to be achieved, the content to be delivered, and the context within it takes place, another group of PE scholars turned their attention to teaching practices that are more pertinent to teaching the subject matter of PE (Metzler, 2014). As Smith (1983) argues, the aims to be achieved and the content of each subject matter mediate teachers' interactions with their students. Differently put, what is to be delivered and achieved largely determines the teaching practices that a teacher will employ. In the case of PE, the main focus is on the psychomotor development of the student (Rink, 2003). Although aiming at all four learning domains (i.e., psychomotor, cognitive, affective, and social), PE pays particular attention to student's psychomotor development, as learning occurs through movement, and through this movement, PE targets to develop the other three learning domains as well.

Beyond the aims and the content, PE also differs from other subject matters in terms of the context within it takes place, which, undoubtedly, provides more challenges for the teacher than a classroom environment does (Pickup, 2012). These instructional particularities of PE context relate to teachers' need of changing from one variation of a task to another, the large and open space within students are constantly moving, often times with equipment, and the possibility of injuries (Lindsay, 2014; Rink & Hall, 2008).

Influenced by all three abovementioned factors (i.e., aims, content, and context), research on teaching in PE has also turned its attention to teaching practices that are more pertinent to teaching the subject matter of PE (Metzler, 2014). These practices are regarded as *content*-specific, either because they are unique to or have a particular functioning and specialized manifestation when occurring in the teaching of specific subject matters (Hamre et al., 2013), without this implying that they cannot occur in the teaching of other disciplines as well. For instance, demonstrating how to perform a motor skill by providing selected learning cues in PE (Rink & Werner, 1989), using and connecting representations in Mathematics (Hill et al., 2008), capitalizing on texts in Language Arts (Grossman et al., 2010), selecting and adapting historical sources in History (Fogo, 2014), and engaging students in investigations in Science (Kloser, 2014) could be regarded content-specific teaching practices.

Contrary to generic teaching practices which are regarded as content-free, and if a teacher is not using them in one content-area, it is very likely that he/she is not using them in other content-areas (Ward, 2013), content-specific practices demand enhanced content knowledge to be effectively employed. In other words, taking into consideration the recent conceptualization of content-knowledge in PE (Ward, 2009), what seems to be needed from teachers to effectively implement such content-specific practices is both the knowledge and skills of appropriately performing a skill (i.e., common content-knowledge), and the knowledge and skills that represent how to teach that skill (i.e., specialized content-knowledge). As explained below, when discussing the contribution of each content-specific teaching practice to student psychomotor learning, these two forms of content-knowledge, individually or jointly, influence the successful employment of content-specific teaching practices in a direct or indirect manner, and in turn, advance student psychomotor learning (Iserbyt, Ward, & Li, 2015; Ward, Kim, Ko, & Li, 2015).

The outcome of PE scholars' attempts to identify such practices was the development of some influential classroom observation instruments encompassing content-specific teaching practices. Three notable examples of such PE-specific instruments are described below. First, frameworks that are considered precursors of the more comprehensive content-specific instruments in PE are presented (*OSCD-PE*); then, more comprehensive content-specific frameworks are considered (*QMTPS; TSS*).

The first framework concerns the *Observation System for Content Development-Physical Education* (*OSCD-PE)* (see Rink, 2010), originally developed by Rink in 1979 to explore the way teachers structure and present the content of their lessons. As can be seen from Table 2.5, according to the model, a teacher should apply the four main task dimensions (i.e., informing, refining, extending, applying/assessment) during a PE lesson, to ensure smooth transition across instructional tasks, and in turn, support student's gradual development of the emphasized psychomotor skill. Rink's seminal work on content development, in terms of the focus of the motor task, has largely influenced subsequent research works in the field of PE, as it is outlined in the next section.

Table 2.5

Description of the Five Task Dimensions of the OSCD-PE [a]

	Task Dimensions
Task Progression-PE Lesson Content Development	- *Informing:* relates to the task(s) that describe(s) a skill or a movement concept with no focus other than just to do it. It is usually presented first in a sequence of tasks. - *Refining:* is concerned with task(s) that focus(es) on helping students qualitative improve the mechanics of the lesson's main psychomotor skills. - *Extending:* refers to task(s) that alter(s) the conditions under which practice is performed, typically to make practice more complex or difficult. - *Applying/assessment:* refers to usually competitive task(s) that require(s) practice in situations similar to those in which emphasized skills will be used in games or other performance settings. - *Repeat*: concerns the simple repetition of the previous task with no changes.

Note. [a] Based on Rink (2010).

The second content-specific instrument concerns the *Qualitative Measures of Teaching Performance Scale (QMTPS*), developed by Rink and Werner (1989), and later validated by Gusthart, Kelly, and Rink (1997). This instrument was designed to observe selected characteristics of task presentation, student responses, and teacher feedback during PE lessons (see Table 2.6). Despite including a generic teaching practice under the category of task presentation (i.e., clarity of task presentation), this instrument mainly focused on content-specific instructional aspects (i.e., demonstration of the desired movement skill, use of appropriate learning cues to enhance demonstration, use feedback to direct student attention to the critical learning cues). In addition, this instrument captures the congruency of students' responses (i.e., whether students practice the skill as described by the teacher). As was the case with the *OSCD-PE* instrument, the contribution of *QMTPS*' teaching elements to student psychomotor development has also been scrutinized, something that is discussed in the next section.

Table 2.6

Description of the Three Domains of the QMTPS [a]

Domain	Instructional Aspects
Task Presentation	- *Clarity*: refers to whether teacher's verbal explanations/directions communicated a clear idea of what to do and how to do it. - *Demonstration*: relates to the visual information modeling the desired movement skill. - *Appropriate number of cues*: is concerned with the degree to which the number of learning cues presented was sufficient without overloading the student. - *Accuracy of cues*: concerns the degree to which the learning cues presented were technically correct. - *Qualitative cues provided*: refers to the verbal information provided to students on the process of the desired movement/performance.
Teacher Feedback	- *Specific congruent feedback*: captures the degree to which teacher feedback during student performance was congruent (matched) the focus of the task (i.e., the correct execution of the desired movement skill).
Students Responses	- *Appropriate to the focus*: refers to the degree to which student responses are on stated task.

Note. [a] Based on Werner & Rink (1989).

The third PE-specific instrument, the *Task Structure System (TSS)*, was originally developed by Siedentop and his colleagues in '90s, based on Doyle's (1986) model, according to which teaching is organized around two intertwined major structures: a) the learning structure which is served by the instructional function, and b) the order structure which is served by the managerial function. Drawing on the PE teaching effectiveness literature and incorporating and refining aspects from the two abovementioned content-specific instruments, Siedentop and his researcher team tried to develop a more comprehensive content-specific instrument to capture teaching quality in PE. In particular, *TSS* consists of five main dimensions (see Table 2.7): type of episode, task progression, task explicitness, teacher's accountability, and congruency and quality of students' responses, each consisting of a specific number of codes. As can be observed, the dimension of task progression includes another type of task (i.e., cognitive task) compared to *OSCD-PE*, and teacher's accountability expanded to capture teacher's overall supervision of students' practice during an instructional task compared to *QMTPS*'s dimension of teacher feedback. In addition, beyond the congruency of student responses, this instrument captures also the appropriateness of student trials, a dimension which is deemed as the ultimate PE teaching skill (Rink, 2003; Silverman, 2011), as most other instructional features affect psychomotor learning through their influence on student practice. This instrument was later used in several research works (e.g., Dyson, Linehan, & Hastie, 2010; Jones, 1992; Siedentop, Doutis, Tsangaridou, Ward, & Rauschenbach, 1994) to explore teacher's organization and presentation of PE content and students' responses to teacher's instruction.

Conducting a closer examination of the above content-specific observation instruments, one can notice that researchers considered certain teaching practices (e.g., task progression in terms of the focus of the motor task) as particularly important for teaching PE (i.e., *content-specific*), and they consistently incorporated them into their observational instruments. In addition, several other researchers, having in mind that these teaching practices could have a significant impact on student psychomotor development, investigated their individual contribution, in separated studies, using only parts of the abovementioned instruments. Thus, in what follows, the main research findings that emerged through these studies are outlined (and

Table 2.7

Description of the Five Main Dimensions of the Task Structure System [a]

Dimension	Codes
Type of episode: captures whether students are engaged in an instructional or non-instructional episode.	• *Management*: refers to issues that are not relevant to the instructional activities (e.g., roll taking, announcements, disciplinary issues). • *Transition*: pertains to teacher behaviors that aim to organize the class (e.g., grouping students, distributing equipment). • *Waiting:* relates to the time that students are waiting to perform a task (e.g., waiting in line for a turn, waiting for the next teacher direction). • *Warm up/Recovery activity:* captures the start-of-class warm up period and/or the end-of-class cool down/recovery period, during which students are engaged in such activities and intensity levels that *no skill improvement* is likely to occur. • *Instructional:* concerns activities/tasks that target the lesson's main psychomotor objectives.
Task progression: explores the way teachers develop the content of their lessons in terms of the focus of the motor task. It distinguishes instructional episodes into five task types.	• *Cognitive:* concerns the task(s) during which students are not engaged in any practice/physical activity, but their role is to attend to the teacher's presentation of information about the lesson's content/objectives; pose/answer questions; discuss how to solve a problem or make a decision; sum up the lesson's basic knowledge or skills; or attend to whole class feedback. • *Informing:* relates to the task(s) that describe(s) a skill or a movement concept with no focus other than just to do it. It is usually presented first in a sequence of tasks. • *Refining:* is concerned with task(s) that focus(es) on helping students qualitative improve the mechanics of the lesson's main psychomotor skills. • *Extending:* refers to task(s) that alter(s) the conditions under which practice is performed, typically to make practice more complex or difficult. • *Applying:* refers to usually competitive task(s) that require(s) practice in situations similar to those in which emphasized skills will be used in games or other performance settings.
Task explicitness: is concerned with the extent to which an instructional task has been explicitly described.	• *Implicit:* refers to a task description that is vague and lacks specificity, but implies a set of conditions or a criterion. • *Partially explicit:* relates to a task description that includes information about the task itself and the conditions under which to practice the task, but not a criterion for success or completion. • *Fully explicit:* pertains to a task description that includes information about the task itself, the conditions under which to practice the task, and some criterion/criteria by which success and/or completion can be judged.
Teacher's accountability: refers to teacher's overall supervision of students' practice during an instructional task.	• *No supervision:* refers to teacher behavior that relates to other chores than monitoring or supervising students' involvement. • *Monitoring:* relates to teacher behavior that is mostly watching students to practice without encouraging or providing them with feedback. • *Reinforcement:* pertains to teacher behavior that reinforces students' practice, by encouraging them to continue practicing and by making positive comments for their trials. • *Specific and congruent feedback:* refers to teacher feedback during practice that is specific and congruent with the appropriateness/correctness of practice (i.e., correct application of learning cues/critical elements of the desired movement skill) and the focus of the task.
Congruency and quality of students' responses: concerns the extent to which student trials are congruent with the teacher's description of task and appropriate.	• *Stated: relates* to student response that is congruent with teacher's description of task. • *Modified up:* refers to student deliberate modification of task to make it more difficult or challenging. • *Modified down:* refers to student deliberate modification of task to render it easier or less challenging. • *Off task:* relates to student behavior that is completely unrelated to the task description. • *Appropriate:* pertains to an acceptable form of practice in terms of the technical features of the performance and its success within the context. • *Inappropriate:* pertains to unacceptable form of practice in terms of the technical features of the performance and its success within the context.

Note. [a] Based on Siedentop, Doutis, Tsangaridou, Ward, & Rauschenbach (1994).

summarized in Table 2.8), either emerging from the use of the abovementioned content-specific instruments as developed or the use of parts of them.

Research Findings on the Contribution of Content-Specific Teaching Practices to Student Psychomotor Learning

Task progression. The way in which teachers develop and present the content during a PE lesson has been a key content-specific instructional aspect considered by PE scholars (Siedentop & Tannehill, 2000). As early as 1979, Rink created the *OSCD-PE* instrument to explore how teachers develop the content of their lessons in terms of the focus of the motor task (see Table 2.5). Task progression or content development refers to teacher's ability to plan and sequence the lesson's movement tasks in such a manner that would facilitate learning (Rink, 2010), and it is considered a content-specific teaching practice for PE, as it can be informed by teacher's (specialized) content-knowledge (Tsangaridou, 2006; Ward, 2009). This, in fact, was supported by a recent study (Ayvazo & Ward, 2011), which found notable differences in task adaptations (e.g., refining, extending) even among within teachers, when the latter were asked to teach two units: a strong unit and one in which they did not feel particularly competent. What differentiates task progression from structuring factor of *DMEE* is that the former does not only involve sequencing students' learning experiences from simpler to more complex or from easier to harder. Rather, it is a measure of the teacher's ability to combine the progression of practice with the quality of performance and the integration of application experiences (Rink, 2010) (see four main types of instructional tasks: *informing*, *refining*, *extending*, *and applying*; Table 2.7).

Studies examining the task progression pattern during a lesson have reported interesting findings. Perhaps, one of the most significant and consistent findings was that, although an ideal sequencing of tasks could not be found (Rink, 2010), when refining tasks were employed, psychomotor learning was enhanced (Masser, 1987; Pellett & Harrison, 1995a; Rikard, 1992; Rink, French, Werner, Lynn, & Mays, 1991). This was not surprising for at least two reasons. First, refining tasks are considered the building blocks of successful skill development (Hastie & Siedentop, 1999), since they aim at improving the quality of a skill. Second, when students are engaged in refining tasks they have the opportunity to practice the skill from a different perspective, and in doing so, they do not become frustrated and they persevere with practice (Silverman, 2009). Adding to that point, Rink (2010) underlines that when students are engaged in a cycle of refining-extending instructional tasks there is a high probability for those students to satisfactorily develop the emphasized skill.

Alas, however, as Tannehill, van der Mars and MacPhail (2015) note, refining tasks are the most neglected type of instructional tasks. In several research works, teachers were found to rarely ask students to engage in refining tasks, with perhaps a few exceptions all sampling one teacher (e.g., Dyson et al., 2010; Dyson & Strachan, 2004; Hastie, 2000): in general, they initially present informing tasks, then add extensions, and finally provide applying tasks opportunities (Hastie & Siedentop, 1999; Jones, 1992; Rikard, 1991; Ward et al., 2015)—a sequence that can be partly attributed to teachers' insufficient content knowledge (Siedentop, 2002, 2009).

Nevertheless, more research work is needed to further and deeper examine how the application of this quartet of instructional tasks, could contribute to student's psychomotor development. Particularly, a key question that remains unaddressed is whether the teachers who use a larger variety (or all four) of those types of instructional tasks during their daily lesson, have better student outcomes than those who do not.

Demonstration of the desired movement skill. As was the case with task progression, demonstration also seems to be informed by teacher's content knowledge (mainly common content-knowledge), as it requires knowing how to correctly perform the technique (i.e., critical elements) of a psychomotor skill (Ward, 2009). Contrary to other disciplines, "in physical education settings words alone don't tell the story" (Veal & Anderson, 2011, p. 42). Demonstrating how to perform a motor skill gives students the opportunity to watch a model performing the desired movement, transforming, in that way, the abstract concept of the emphasized skill to something more concrete and tangible (Hunt, Wiseman, & Touzel, 2009; Valentini, 2004). Importantly, what differentiates this practice from teaching modeling and what makes it a content-specific one, is the essential emphasis on the skill's basic learning cues (few words or a phrase). Demonstration is considered to become even more powerful when it is simultaneously accompanied with some basic learning cues, which help students better understand the desired movement to be practiced (Landin, 1994; Metzler, 2011). Learning cues are pieces of information that compress the different parts of a movement skill into few words or a phrase, which, in turn, direct students' attention to what is relevant in a response (Belka, 2002; Magill, 2010; Metzler, 2011). For example, when teaching students how to dribble a basketball, the teacher can direct their attention to pushing the ball with their upper finger pads, bouncing the ball below their waist level, and keeping their heads up and their eyes forward (Krause, Meyer, & Meyer, 2008; Paye & Paye, 2013).

Prior studies have documented, in isolation and in conjunction, the role of effective demonstration and concise learning cues for student PE learning. For example, investigating only the contribution of visual demonstration, Horn, Williams, Hayes, Hodges, and Scott (2007) found that the participants who observed a video demonstration concerning throwing a ball toward a target, performed better than their counterparts who did not observe the video. On the other hand, examining only the contribution of learning cues to student learning, Masser (1993) found that learners improved their performance and maintained that improvement over a 3-month period when their teachers directed their attention to a critical learning cue of the emphasized movement.

Similar results were also found when investigating the joint contribution of demonstration and verbal cues. A recent study showed that demonstration of key features of manipulative skills along with presentation of related learning cues (as a part of the task presentation dimension) played one of the most significant contributing roles in increasing students' manipulative skill competency (Chen, Zhu, Mason, Hammond-Bennett, Colombo-Dougovito, 2016). A result that was corroborated by other research works as well, which found that students who received full demonstration accompanied with succinct learning cues performed better than their counterparts who received a partial demonstration or no demonstration at all (Janelle, Champenoy, Coombes, & Mousseau, 2003; Kwak, 2005; Werner & Rink, 1989). Moreover, in their meta-analysis, Ashford, Bennett, and Davids (2006) found that demonstration is effective for both the acquisition of the form and the performance outcome of the desired movement; yet, the researchers noted that demonstration seems to be particularly efficient for the former, a result that was replicated in another subsequent study (Ram, Riggs, Skaling, Landers, & McCullagh, 2007).

Specific and congruent feedback. A content-specific teaching practice, which is highly related with the practice of skill demonstration and it has received significant research attention, concerns the extent to which feedback provided during student practice is specific and congruent. Once again, this teaching practice is informed by teachers' content-knowledge, and specifically by both the common and specialized knowledge, since teacher not only needs to know the critical elements of the skills to provide this type of feedback, but also he/she should be able to discriminate and correct students' errors of technique (Ward, 2009). As mentioned in a previous section, finding that feedback did not consistently have a significant impact on student psychomotor learning (Lee et al., 1993), researchers in the field of PE opted to divide the multidimensional concept of feedback into smaller segments and examine what it is exactly about it that influences psychomotor learning. In other words, considering that each form of feedback serves a different purpose in the instructional setting (Rink, 2010), and that some forms are more powerful than others (Hattie, 2009), researchers selected to investigate those aspects of feedback that could be particularly important for teaching psychomotor skills.

Specifically, some scholars opted to investigate the specificity and congruency of the individualized feedback provided during student performance. This type of feedback targets the quality of student performance, by narrowing students' attention to the emphasized learning cues; consequently, it was deemed to be more significant than the general feedback (Lee et al., 1993). In essence, specific and congruent feedback gives information about student performance that is directly related to what the learners have been asked to focus on. In addition, it is important that this type of feedback be provided during student performance, as once the performance is completed, the teacher will not have permanent products of student motor performance, as it happens in other disciplines.

Beyond the rules of thumb that have been provided by scholars (i.e., specific and immediate feedback are more effective than general and delayed feedback, respectively, see Metzler, 2011), some research evidence exists supporting that this kind of feedback can have a positive impact on student psychomotor learning. Specifically, teacher specific and congruent feedback found to be supportive for both low- (Rikard, 1991) and high-skilled students (Pellett & Harrison, 1995b), helping them improve their immediate practice success (Oslin, Stroot, & Siedentop, 1997; Silverman, Kulinna, & Crull, 1995). Likewise, Silverman et al. (1992) found that descriptive feedback (i.e., description of an error without evaluation) and prescriptive feedback (i.e., description of how to make performance better) or combination thereof were related to student achievement; a relationship previously supported by students participated in De Knop's (1986) study, who evaluated specific feedback as a sound and desirable teaching practice to improve learning. In addition, despite not confirming the above consistent relationship, Rikard (1992) concluded that this type of feedback is important for skill acquisition when combined with other learning conditions (e.g., feedback following extending and applying tasks), whereas Sariscsany, Darst, and van der Mars (1995) found that specific feedback had a significant positive impact on on-task performance, but not on total or successful trials. Moving this research effort a step forward, Burzycka-Wilk (2010) examined the effect of visual congruent feedback provided to students through the use of gestures, finding that students who received congruent feedback demonstrated a more accurate technique. Nevertheless, due to limitations pertain to the above studies, these research findings are not conclusive and more relevant studies are needed to corroborate the relationship among teacher's specific and congruent feedback and student psychomotor learning.

Task explicitness. Another PE-specific teaching practice refers to the extent to which the teacher is explicit in presenting the instructional tasks. As opposed to other subject matters, in PE, students do not have a written description of what to do in the next task, but they usually listen to the teacher providing verbal directions. Thus, these directions have to be fully explicit to help students engage in appropriate practice. Unlike clarity in presenting a task, which concerns the extent to which teacher's verbal directions communicate a clear idea to students of what to do and how to do it, explicitness is mostly concerned with the extent to which teacher directions communicated to students a criterion by which successful performance can be judged. Thus, to be fully explicit, teacher directions must communicate to students three critical aspects: a) the content of the task (i.e., the skill to be practiced); b) the organization of the task (i.e., arrangements of time, space, people, and equipment); and c) the goal orientation of the task (i.e., the qualitative or desired aspect of the skill to be practiced) (Rink, 2010). In other words, when a task is less than fully explicit, it becomes ambiguous, since almost any answer can become acceptable. Accordingly, students may not be 100% focused on what the teacher wants them to achieve, since they might be more interested in the final result, rather than how the skill is performed (Metzler, 2011; Rink, 2010). As Ward et al. (2015) mention, the accurate description of tasks requires common content knowledge; thus this teaching practice was deemed content-specific.

However, research findings pertaining to the effect of this teaching practice on student psychomotor learning in real settings have been mixed, calling for further investigation. Despite the fact that some studies have documented task explicitness to be an effective teaching practice that produces higher rates of on-stated-task behavior (e.g., Silverman et al., 1995; Tousignant & Siedentop, 1983), other studies did not report such a clear relationship (e.g., Silverman et al., 1998). A basic point in these latter studies was that what eventually determines the work that students accomplish, is not the explicitness of task description, but the subsequent supervision and accountability of student practice (Hastie & Siedentop, 1999).

Increasing quality of student practice trials. The final content-specific teaching practice concerns teachers' ability to maximize quality practice opportunities by increasing the appropriate student practice trials. Undoubtedly, practice is essential to learning, but if students do not perform the assigned task as described by the teacher or if they perform it inappropriately/unsuccessfully because the task is too difficult, they will not learn (Silverman & Mercier, 2015). Therefore, the effectiveness of teachers' behavior inextricably depends on what the students are doing. Unsurprisingly, then, many PE scholars deem this teaching practice (i.e., increasing students' appropriate trials) as the ultimate teaching skill in PE (e.g., Ennis & Chen, 2011; Rink, 2003; Silverman, 2011), because, as they contend, most other teaching skills affect psychomotor learning through their influence on student practice (see Figure 2.1).

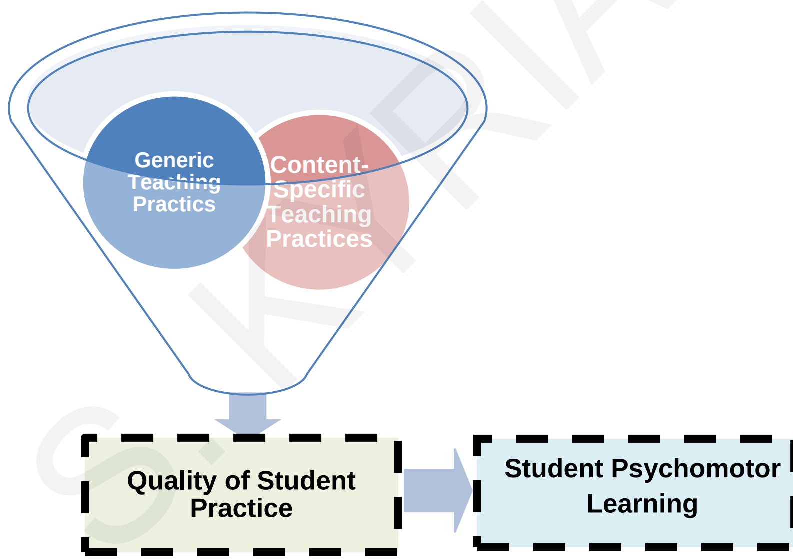


Figure 2.1. Indirect influence of teaching practices on psychomotor learning through quality of student practice.

However, as recent studies have shown (e.g., Iserbyt et al., 2015; Ward et al., 2015), increasing the quality of student practice (i.e., correct trials) is significantly influenced by teachers' common and specialized content-knowledge, thus, this teaching practice is deemed content-specific. Differently put, teachers' content knowledge informs the employment of certain content-specific teaching practices, which, in turn, impact on student quality practice. What teachers actually influence when they sequence the tasks in a progressive manner or demonstrate a movement skill or provide congruent and specific feedback (i.e., correcting the student errors) or explicitly present the instructional tasks, is the quality of students' practice, which will subsequently lead to student learning.

At a surface level, this teaching practice seems to be identical to teachers' work to maximize the ALT described above. However, student discrete practice trials were examined through a different, very influential research line, known as the "Opportunity To Respond-Physical Education" (OTR-PE), and which is deemed a more sensitive metric of student practice (van der Mars, 2006). In particular, some researchers in the field of PE claim that time on task could be a somewhat misleading variable influencing student achievement in PE, as its definition is not specific to particular kinds of content within PE (e.g., Rink 1999; Silverman, 1985). This is especially apparent in game-like situations, where students—particularly low-skill students—might appear to actively participate in the game, but their substantial engagement (e.g., practicing a skill like giving a pass, or even touching the ball) is minimal, resulting in a negative relationship with their achievement (Silverman et al., 1988). This fact was indeed highlighted by a recent study which did not find any significant differences in the quantity and quality of student trials, among students practicing in either huge lines (resulting in large amounts of waiting time) or in game-like activities involving the concurrent participation of the entire class (performing the same skills as in lines) (Hastie, Calderon, Palao, & Ortega, 2011).

As a consequence, during the last years, research on OTR-PE has increased, while research on time-based variables in PE has waned (van der Mars, 2006). Several researchers turned their attention to student practice trials and investigated the relationship among the quality (i.e., appropriate, inappropriate) of discrete practice trials and student psychomotor learning. The research findings were more than encouraging, since a strong and positive relationship among the two variables was found (Ashy, Lee, & Landin, 1988; Buck, Harrison, & Bryce, 1991; Silverman, 1985); it should, however, be mentioned that some researchers warned about a point of diminishing returns (e.g., Silverman, 1990), thus alluding to a potentially curvilinear than linear effect. Equally interesting was that some researchers found a significant negative relationship among inappropriate practice (e.g., that is too hard or too easy) for low-skill students and psychomotor learning (Silverman, 1993). This finding becomes even more important, when considering that low-skill students reported that when they succeeded in PE lessons, they felt motivated toward practicing and learning in PE (Portman, 1995; Walling & Martinek, 1995).

Summary. Before shifting to outline what the few studies attending to both generic and content-specific practices have shown, a couple of comments emerging from the above research findings are set forth. First, as was the case with the contribution of generic teaching practices, the above research findings show that some discrete content-specific teaching practices (e.g., demonstration of the desired movement skill, teachers' ability to increase the quality of student practice trials) seem to have a significant impact on student psychomotor learning, while other content-specific teaching practices (e.g., task progression, task explicitness) still need more empirical support through PE studies. Consequently, this does not only call for further investigation of some discrete content-specific teaching practices, but also for the investigation of the joint contribution of content-specific teaching practices to student-psychomotor learning.

However, and turning to the second comment, ignoring that there might be a generic set of effective teaching practices that can be employed in PE lessons seems to be equally problematic as overlooking the importance of employing practices that are more pertinent to teaching psychomotor skills. Differently put, considering only one type of practices could lead to "*necessary, but not sufficient conditions for learning*" (Rink, 2013, p. 408), because, as one can conclude from the abovementioned research findings, both types of practices seem to be important for psychomotor learning. Having this in mind and that each type of practices includes a somewhat different set of teaching skills (Kane & Staiger, 2012), what seems reasonable to assume, is that working at the intersection of generic and content-specific teaching practices can help better explore and understand teaching, and in turn, can lead to greater advancements in student learning. In other words, to discover the silver bullets which lead to better learning, more extensive and comprehensive research is needed.

Table 2.8

Summary of the Studies Exploring the Contribution of Content-Specific Teaching Practices to Psychomotor Learning [a]

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Main Results
Task Progression	Engaging students in refining tasks	Masser (1987)	Elementary: K-6th grade T: 1 (the researcher) St: 529	Quasi-experimental Design. Three groups: a) *Control group*: no instruction or practice in the standing broad jump, but in the overhand throw, b) *"E"*: received instruction and practice in the standing broad jump, c) *"E+"*: received the verbal teacher behavior of specific refinement during instruction and practice in the standing broad jump. The Observation System for Content Development-Physical Education was used to code all verbal teacher behaviors in the lessons taught to the experimental groups.	Pre-, post-performance and retention test (after 7 months) (Standing broad jump skill)	• The teacher behavior of refinement had both immediate and long term positive effects on student achievement in performing the standing broad jump.
		Pellett & Harrison (1995a)	Secondary: 7th and 8th grade T: 1 specialist St: 200 female students	Quasi-experimental Design. Three groups: a) *Control group*: no instruction or practice in volleyball, but in aerobics, b) *Extension and Application*: for each task within a lesson received a verbal presentation of specific cues and a visual demonstration before practicing, c) *Extension, Refinement, and Application*: as was the case with the Extension and Application group, but, in addition, this group received refinement tasks as well.	Pre- and post-performance test (AAHPERD and game -like tests- Volleyball underhand serve, forearm pass and overhead set skills)	• Refinement tasks had a significant positive effect on students' daily practice success and overall achievement.
		Rikard (1992)	Elementary: 4th grade T:2 specialists St: 8 low- & high-skilled target students	Intervention encouraging teachers to develop refining tasks. Six-to-Seven 45-minute videotaped classroom observations (indoor lessons) per teacher were conducted and analyzed using author-developed coding sheets.	Pre- and post-performance test (A wall-volley striking test, striking skill)	• Results showed no significant increases for low-skilled students. In contrast, high-skilled students' practice improved when they received refining tasks, as compared to when they received extending and applying tasks.
		Rink, French, Werner, Lynn, & Mays (1991)	Secondary: 9th grade T: 2 specialists St: 76	Quasi-experimental Design. Five groups: a) *Control group*: no instruction or practice in the volleyball, but in other skills during their regular PE class, b) *Final-test Motivation*: practiced the final test each day and received a great deal of reinforcement for their effort, c) *Final-test Refinement*: practiced the final test each day and after every five trials of a skill, the teacher selected a performance cue as a group practice focus for the next set of trials, d) *Four-step Progression-Extension*: received a task presentation on each step of the progression only, e) *Four-step Progression-Combination*: as was the case with the Extension group, plus a focused refinement task after every five trials of practice. All groups had an equal number of trials (60 overall).	Pre- and post-performance test (AAHPERD-Volleyball overhead set and serve skills)	• The results supported the positive effect of providing students with a progression and the need for refinement tasks for parts of the progression.

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Main Results
		Ashford, Bennett, & Davids (2006)	Meta-Analysis. [b]			• Demonstration is effective for both the form and performance outcome acquisition of the desired skill. However, it seems to be particularly efficient for the former.
		Chen, Zhu, Mason, Hammond-Bennett, & Colombo-Dougovito (2016)	Elementary: 4th and 5th grade T: 9 (not specified) St: 2709-3420 (two-year study)	Seven 30-minute videotaped lessons per teacher were coded and analyzed using the Assessing Quality Teaching Rubrics.	Post-performance test (PE Metrics Assessment Rubrics- Soccer dribbling, passing and receiving, overhand throwing, and striking with a racket skills)	• Demonstration of key features of manipulative skills along with presentation of related learning cues (as a part of task presentation) played one of the most significant contributing roles in increasing students' skill competency.
Demonstration of the desired movement skill	Combination of visual and verbal (learning cues) demonstration	Horn, Williams, Hayes, Hodges, & Scott (2007)	Mean Age: 31.9 years T: - St: 16 male participants	Experimental Design. Two groups: a) *Model*: observed five consecutive demonstrations of the model and tried to replicate exactly the model's form in all acquisition trials. In addition, they observed one demonstration after each of the first five acquisition trials, b) *Control group*: practiced the skill based on the initial verbal instruction. All subjects had an equal number of trials (21 trials overall).	Pre-test and acquisition trials (Throwing a ball towards a target with maximal velocity using a back-handed reverse baseball pitch) (outcome and coordination)	• Participants in the model group immediately changed their motion pattern to resemble to the model's motion. The model group showed a large, significant increase in ball speed, while the control group showed no change.
		Janelle, Champenoy, Coombes, & Mousseau (2003)	Tertiary education T: - St: 60	Experimental Design. Six groups: a) *Control group*: received a written manuscript of soccer history that did not include any specific reference to the soccer pass, b) *Verbal instruction*: received only a verbal description of the task that outlined the critical steps to executing the skill, c) *Video+visual cues*: viewed a video model performing the task that included directional arrows indicating critical elements of the skill (visual cues), d) *Video+ verbal cues*: viewed a video model executing the skill that was accompanied by verbal cues highlighting the critical elements of the skill, e) *Video+ visual+verbal cues*: viewed a video model with both the visual and verbal cues that were used independently in other experimental conditions, and f) *Video-only*: viewed an unmodified video of the model performing the skill. All subjects had an equal number of trials (8 blocks of 10 trials).	Acquisition and Retention phase (Soccer accuracy pass skill) (outcome and form)	• The findings showed that the *video+visual+verbal* group collectively displayed less error and more appropriate form across acquisition and retention trial blocks compared with other groups.

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Results
		Kwak (2005)	Secondary: 7th and 8th grade T: 2 specialists St: 120	Experimental Design. Five groups: a) *Control group*: received no instruction, b) *Succinct verbal explanation with partial demonstration*: received short amount of verbal information on aspects of the skill with partial demonstration on each cue, c) *Full demonstration only*: received three times full demonstration of the skill, d) *Overload verbal explanation with partial demonstration*: received detailed and extensive information on aspects of the skill with partial demonstrations, and e) *Succinct verbal explanation with full demonstration, verbal/visual cue, and rehearsal*: received descriptive succinct verbal information with both full and partial demonstrations and summary cues and verbal/visual rehearsal strategies.	Post- performance test (Lacrosse throw skill) (accuracy & use of appropriate movement process characteristics)	• Results indicated that the *verbal explanation with full demonstrations, summary cues, and verbal/visual rehearsal* group achieved the best student performance (on both the accuracy and appropriate movement process characteristics).
Demonstration of the desired movement skill (continued)	Combination of visual and verbal (learning cues) demonstration (continued)	Masser (1993)	Elementary: 1st grade T: 1 (the researcher) St: 69 (Experim. 1) and 44 (Experim. 2)	Quasi-experimental Design. Experiment 1: Three groups: a) *Control group*: received neither instruction nor practice in performing the handstand, b) *Cue*: received instruction and practice in performing the handstand with particular emphasis on its critical aspects/learning cues, c) *No-cue*: received the same instruction and practice in performing the handstand as the Cue group, but without the emphasis on its critical elements. Experiment 2: Two groups: both groups received the same instructions, but different phrases were used for one of the critical elements (i.e., forehead on your knees vs make yourself into a tight ball). In both experiments teacher behaviors were coded and analyzed using the Observation System for Content Development- Physical Education.	Pre-, post-performance and retention test (after 3 months for handstand skill and after 2 months for forward roll skill)	• The cue group in the experiment 1 had a significant positive improvement in both the post- and retention tests. On the contrary, the no-cue group did not significantly improve the emphasized skill. In the experiment 2 the group that instructed with the phrase “forehead on your knees” had a significantly higher achievement in the retention test than the group instructed with the “make yourself into a tight ball” phrase.
		Ram, Riggs, Skaling, Landers, & McCullagh (2007)	Tertiary education Experim. 1 T: - St: 41female students Experim. 2 T: - St: 60 female students	Experimental Design. Four groups: a) *Modeling*: viewed a video performing the squat lift/balancing the stabilometer with ideal form, b) *Imagery*: listened to an audiotape guiding them through mental imagery of the squat lifting/stabilometer task, c) *Combination*: viewed the modeling video before the first trial, listen to the imagery videotape before the second trial, and so forth, alternating between modeling and imagery interventions, d) *Control group*: offered a newspaper to read during the rest periods between trials. All subjects had an equal number of trials (Exper.1: 14 lifts of four acquisition and four retention trials, Exper. 2: four blocks of five acquisition trials and one block of 10 retention trials)	Acquisition and Retention trials (Exper. 1. Free-weight lifting squat, Exper. 2. Balance a stabilometer)	• *Modeling* and *Combination* groups had a more appropriate form than *Imagery* and *Control* groups. Researchers conclude that even a single bout of modeling can have immediate beneficial effects on both the movement form (exper. 1 & 2) and the outcome (exper. 1).

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Results
Demonstration of the desired movement skill (continued)	Combination of visual and verbal (learning cues) demonstration (continued)	Werner & Rink (1989)	Elementary: 2nd grade T: 4 (not specified) St: 160	Two phases of six-lesson jumping and landing units per teacher (a baseline and re-teach) were coded and analyzed using the Observation System for Content Development-Physical Education and the Qualitative Measures of Teaching Performance Scale instruments.	Pre- and post-performance test (Jumping and landing skills)	• Although a statistical relationship was not established between skill demonstration and student achievement, the use of: a) visual demonstration coupled with verbal explanation, b) appropriate number of cues, and c) qualitative cues, was considered to improve effectiveness.
Congruent and specific feedback	Narrowing students' attention to the emphasized learning cues during or immediately after practice	Burzycka-Wilk (2010)	Elementary: age 9 T: 5 qualified swimming instructors St: 86	Experimental Design. Two groups: Both groups were taught the same lesson subjects/objectives during seven 45-minute lessons by qualified swimming instructors. The only difference among the groups was: a) *Control group*: errors were eliminated by means of verbal information, b) *Experimental group*: errors were eliminated by means of visual information communicated by a gesture.	Post- performance test (Swimming backstroke skill)	• Correcting student errors by employing visual information /gestures seems to be more effective than using only verbal information.
		DeKnop (1986)	Tertiary education T: 8 tennis specialists St: 48 first-year students	Five three-hour videotaped tennis lessons were coded and analyzed using author-developed teacher and student observation instruments.	Pre- and post-performance test (Bornemann-Mester Scale-Tennis forehand, backhand and service skills)	• More effective teachers devoted significantly greater amount of time for specific feedback. In addition, according to students, the amount of specific feedback was the most important element of good teaching.
		Oslin, Stroot, & Siedentop (1997)	Pre-School: 3.3-5.8 ages T: 1(one of the researchers) St: 7 female students	Experimental Design. Two groups: Each child received component-specific instruction/feedback following each trial based on the efficiency level demonstrated. a) *Force-Production Sequence*: the component-specific feedback provided to students was sequenced based on the components that contributed most to force production, b) *Forward-Chaining Sequence*: the component-specific feedback provided to students was sequenced based on the order of components occurrence. For both groups, the number of student trials and the distance between the children and the target varied.	Pre-performance test, acquisition trials and retention test (after 3 and 7 working days) (Overarm throw skill)	• Component-specific instruction was found to be effective for improving the efficiency of the skill. However, during the retention test, efficiency levels varied from child to child.
		Pellett & Harrison (1995b)	Secondary: 7th and 8th grade T:1 specialist St: 68 female students	Intervention encouraging teacher to provide appropriate specific, congruent and corrective feedback. Eleven videotaped lessons were coded and analyzed using an instrument that emphasized the type of instructional tasks (i.e., extending, refining, applying), the appropriateness of student practice trials and the type of teacher feedback.	Pre- and post- performance test (AAHP-ERD and game-like tests-Volleyball underhand serve, forearm pass and overhead set skills)	• Students significantly improved the emphasized skills. Specific, congruent and corrective feedback shown to be important for both low- and high- skilled students, as helped in improving their immediate practice success for the pass and set skills.

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Results
Congruent and specific feedback (continued)	Narrowing students' attention to the emphasized learning cues during or immediately after practice (continued)	Rikard (1991)	Elementary: 4th grade T:2 specialists St: 8 low- & high-skilled target students	Four-to-five 40-minute videotaped lessons were coded and analyzed using an instrument that emphasized the type of instructional tasks (i.e., informing, extending, refining, applying), the type of teacher feedback and the appropriateness of student practice trials.	Appropriateness of student trials	• After receiving teacher feedback, all four low-skilled students and three out of four high-skilled students increased their practice success, while the practice success of the fourth high-skilled student declined.
		Rikard (1992)	Elementary: 4th grade T:2 specialists St: 8 low- & high-skilled target students	Intervention encouraging teachers to use specific and corrective feedback to individual and groups. Six-to-Seven 45-minute videotaped classroom observations (indoor lessons) per teacher were conducted and analyzed using author-developed coding sheets.	Pre- and post-performance test (A wall-volley striking test, striking skill)	• Specific corrective feedback did not consistently improve subject's practice success; yet, it was found to be important when combined with other learning conditions (i.e., modest increases in success occurred for low- and high-skilled students when specific feedback followed refining tasks and significant increases in success occurred for high-skilled students when specific feedback followed extending and applying tasks).
		Sariscsany, Darst, & van der Mars (1995)	Secondary: 6th and 7th grade T: 3 specialists St: 3 inattentive target students	Experimental Design. Duration recording of three supervisory conditions: a) close supervision with specific skill feedback b) distant supervision with specific skill feedback c) distant supervision with no feedback.	On-task and appro-priate skill trials were coded using event or duration recording, depending on behave-ioral dimensions of the skills observed (Volleyball unit)	• Close and distant supervision with specific skill feedback produced higher levels of on-task performance for two of the three target students, than the no feedback condition.
		Silverman, Tyson, & Krampitz (1992)	Secondary: 6th, 7th and 8th grade T: 7 specialists St: 200	Seven 30-minute videotaped lessons were coded and analyzed using an author-developed instrument that emphasized the amount, type, form, focus, and quality of feedback received by individual/group of students.	Pre- and post-performance test (AAHPERD and Brumback tests-Volleyball serve and forearm pass skills)	• Total feedback by itself did not relate to student achievement. Feedback on the outcome of the skill attempt was the only category related to achievement for both skills. The combination of positive, auditory, and whole feedback was positively related to achievement for both skills

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Results
Congruent and specific feedback (continued)	Narrowing students' attention to the … (continued)	Silverman, Kulinna, & Crull (1995)	Secondary: 6th, 7th and 8th grade T: 7 specialists St: 202	Seven 30-minute videotaped lessons were coded and analyzed using an author-developed coding instrument that captured task structure information (e.g., focus, organization, and explicitness of the task, teacher accountability and student practice).	Pre-and post performa-nce test (AAHPERD and Brumback tests-Volleyball-underhand serve and forearm pass skills)	• Individual skill-related feedback was positively related to underhand serve achievement.
Task Explicitness	Providing fully explicit directions (informing students about the content, organization, and goal orientation of the task to be practiced)	Silverman, Kulinna, & Crull (1995)	Secondary: 6th, 7th and 8th grade T: 7 specialists St: 202	Seven 30-minute videotaped lessons were coded and analyzed using an author-developed coding instrument that captured task structure information (e.g., focus, organization, and explicitness of the task, teacher accountability and student practice).	Pre- and post-performance test (AAHPERD and Brumback tests-Volleyball-underhand serve and forearm pass skills)	• For the forearm pass, combinations of task explicitness that contained all the outcome, situation, and criteria-product were positively related to achievement. For the underhand serve, task explicitness that contained the situation, criteria-product and combination of outcome-situation were positively related to achievement.
		Silverman, Subramaniam, & Woods (1998)	Secondary: 7th, 8th and 9th grade T: 8 specialists St: 72 (three target students at each skill level--i.e., low, medium, high—per teacher)	Two 30-minute videotaped classroom observations (indoor lessons) per teacher were conducted and analyzed using a task instrument.	Appropriate/ Successful and inappropriate/unsuccessful practice trials (variety of activities)	• Situation-outcome and situation-criterion product variables had the greatest numbers of appropriate trials. However, authors conclude that task explicitness may not be as powerful as the organization variables (i.e., individual, reciprocal, small-group practice and lead-up game) of influencing appropriate practice.
		Tousignant & Siedentop (1983)	Secondary: 7th, 8th, 9th and 10th grade T: 3 specialists St: not specified	A total of 127 (full class period) live classroom observations were conducted employing a detailed narrative account of events combined of quantitative techniques of data collection.	Student behavior during practice	• Fully explicit tasks were usually associated with the highest rate of on-stated-task behavior. In addition, partially explicit tasks led to a high rate of on-stated-task behavior, whereas during implicit tasks almost any related response was accepted.

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Results
		Ashy, Lee, & Landin (1988)	Elementary: 4th grade T:8 preservice PE teachers St: 80 (eight groups of ten students)	Two 20-minute videotaped lessons were coded and analyzed using an event recording system capturing the total number and the appropriate practice trials (using correct technique) of each student.	Post- performance test after each of the two lessons (Soccer kick-up skill)	• A moderately high significant relationship was found between appropriate practice and student achievement. The relationship between total trials and achievement was low and not significant.
Quality of Student Practice Trials	Increasing appropriate /successful student practice trials	Buck, Harrison, & Bryce (1991)	Tertiary educ. T: 2 highly skilled volley-ball players St: 58 (a few students were dropped from the analyses because of incomplete test data)	22 videotaped lessons were coded and analyzed based on the correct and incorrect trials for each student. A list of criteria for correct trials was given to the instructors to ensure that what was taught was designed to result in performance being measured. No other attempt was made to modify the instructional methods of the teacher.	Appropriateness of daily student trials and pre- and post-performance test (AAHPERD and Stanley spike tests-Volleyball set, forearm pass, serve and spike skills)	• The most consistent result was the importance of student total correct trials in determining achievement. It is also noted that outside-of-class participation increased the number of total correct trials for three out of four emphasized skills.
		Silverman (1985)	Tertiary education 17 to 31 years T: 5 certified swimming instructors St: 57	Two 15-minute videotaped lessons were coded and analyzed using a modified version of the ALT-PE coding system for motor and cognitive engagement, and an event-recording instrument for the quantity, type, and difficulty level of practice trials.	Pre- and post-performance test (Survival float-swimming skill)	• Whole-appropriate trials were positive predictors of achievement, whereas whole-inappropriate trials were negative predictors of achievement. Thus, practicing at an appropriate level of difficulty enhances achievement. The total number of trials was predictive of achievement for high-skilled students.

Table 2.8 Continued

Content-Specific Teaching Factor	Elements of Teaching Factor	Study	Stage of Schooling and Participants: Teachers (T) Students (St)	Methods/Procedures	Performance Measures (Task/Skill)	Results
Quality of Student Practice Trials (continued)	Increasing appropriate/ successful student practice trials (continued)	Silverman (1990, 1993)	Secondary: middle and junior high school T: 7 specialists St: 194 for forearm pass and 192 for serve skill	Seven consecutive regularly scheduled videotaped lessons were coded and analyzed using an author-developed observation system that captured the quantity and quality of individual student practice trials.	Pre- and post-performance test (AAHPERD and Brumback tests-Volleyball-underhand serve and forearm pass skills)	• Total and appropriate practice trials were positively related to achievement, whereas inappropriate trials were negatively related to achievement for both skills. A plateau and curvilinear relationship occurred after many practice trials (however, author warns that this is a usual finding that occurs in laboratory studies) (Silverman, 1990). • Low-skilled boys received the greatest number of appropriate trials, which had a strong correlation with achievement. Low-skilled students had significantly more inappropriate trials than medium- or high-skilled students, and girls had significantly more than boys. Consequently, skill level and gender are important variables for planning and implementing instruction (Silverman, 1993).

Notes. [a] The above review table is not exhaustive.
[b] Where review articles are presented, studies that preceded those reviews are not mentioned in the table.

Combining the Two Types of Teaching Practices

Following the above rationale, recently a few researchers in the field of PE and in other disciplines as well, have attempted to combine generic and content-specific teaching practices to develop and validate more comprehensive classroom observational instruments and to explore how these instruments may be associated with improved student achievement. These efforts have taken at least three different forms.

First, researchers tried to encompass generic and content-specific practices in single observation instruments. In the field of PE, Chen and colleagues' (2011) attempt to develop and validate a comprehensive classroom observational instrument for assessing quality teaching in PE (i.e., *Assessing Quality Teaching Rubrics, AQTR*) represents a prominent example of this kind of research work. Although not explicitly referred to generic and content-specific teaching practices as such, these researchers focused on key quality teaching practices found to contribute to student learning across subject areas, and tried to situate those practices into the context and content of PE. In particular, this instrument consists of four main generic dimensions each of which involves a different number of teaching components (17 overall) that are specific to the context and content of PE lessons (see Table 2.9). Through a series of studies *AQTR* has been determined as a psychometrically supported measure evaluating either pre-service (Chen, Hendricks, & Archibald, 2011) or in-service teachers (Chen, Hammond-Bennett, Upton, & Mason, 2014; Chen, Mason, Hammond-Bennett, & Zlamout, 2014). Moreover, using this instrument, these scholars found teachers to score differently on each of the four main dimensions of the instrument (Chen, Mason, Staniszewski, Upton, & Valley, 2012). Additionally, students of teachers who scored high quality levels among instrument's dimensions were significantly more competent (based on a final assessment) (Chen et al., 2016) and more physically active (Chen, Hypnar, Mason, Zlamout, & Hammond-Bennett, 2014) than students whose teachers scored low on it.

A similar attempt refers to the development of the Singapore *Physical Education Lesson Observation Tool* (*PELOT*). As Rink (2013) explains, drawing on the *Framework for Teaching* (Danielson Group, 2013)—a comprehensive generic-practices instrument—and adapting its descriptors to be suitable for PE lessons, the stakeholders in Singapore have developed a comprehensive observation tool for observing PE lessons. However, a review for studies that validated or employed this instrument to assess PE lessons did not yield any results.

Table 2.9

Description of the Four Dimensions of the AQTR [a]

Dimension	Teaching Component
Task Design	- The teacher offers students developmentally appropriate and challenging tasks. - The teacher provides students with maximally engaging tasks. - The teacher provides students with learning tasks that build on the previous tasks in a clear progression.
Task Presentation	- The teacher presents the tasks in a clear, concise, and accurate manner. - The teacher links the task presentation to game situations to help students understand the rationale for learning/using a skill/tactic. - The teacher demonstrates the correct form of the skill or a tactical concept. - The teacher presents the learning cues of the emphasized skill. - The teacher facilitates student understanding of the task by either asking questions or re-emphasizing important elements before student practice.
Management	- The teacher uses strategies/routines to gain/keep students' attention. - The teacher/students use(s) strategies/routines to collect/return equipment. - The teacher uses strategies/routines to form students into pairs/groups. - The teacher minimizes transition time among learning tasks.
Responses	- The teacher uses strategies to monitor the whole class and stop off-tasks behaviors. - The teacher stops the entire class to adjust or re-emphasize the important elements of a task, when the majority of students are not able to perform the task successfully. - The teacher provides students with positive/general feedback. - The teacher provides students with specific performance feedback based on students' movement responses. - The teacher engages students in reflecting on their performance and/or how to successfully perform a task.

Note. [a] Based on Chen et al. (2011).

Such attempts to develop comprehensive frameworks encompassing generic and content-specific teaching practices are also observed in other content areas. For instance, in Language Arts, Grossman and colleagues (2013) developed the *Protocol for Language Arts Teaching Observation* (*PLATO*) instrument which involves generic and content-specific practices. Using this instrument, these scholars documented differences between more and less effective teachers (based on their value-added scores) with respect to several instructional dimensions related to either generic or content-specific practices. Another such attempt concerns the study of Matsumura and colleagues' (2008), who developed the Instructional Quality Assessment rubric for measuring Mathematics and Language Arts instruction. Particularly, these researchers identified three common broad constructs (i.e., cognitive demand of tasks, classroom talk, and teacher expectations) that characterized the quality of instruction in both disciplines under consideration. Those three constructs involve some common teaching practices that capture teaching quality among the two disciplines, and some other teaching practices which intend to gauge teaching behaviors that are more pertinent to each of the two subject matters under consideration.

The second form of studies concerns the simultaneous employment of generic and content-specific observational instruments to investigate the instructional quality of the same lessons. The most notable work of this kind is the MET study, which explored how comprehensive generic and content-specific instruments may be associated with improved student achievement. This study employed five observation instruments (i.e., two generic, two content-specific for Mathematics and one content-specific for Language Arts) to code Mathematics or Language Arts lessons of nearly 3,000 teacher-volunteers in public schools across the U.S.A. The findings of this study were very supportive for combining both types of teaching practices in exploring teaching quality for two reasons. First, all five instruments have been shown to be positively associated with student achievement gains, suggesting that each instrument can capture certain teaching practices that help explain a proportion of the variance in student performance. Second, and in conjunction with the above, the fact that lower correlations were identified for the estimates yielded across generic and content-specific instruments rather than within the two categories of instruments (see Kane & Staiger, 2012, p. 32), suggests that the one type of teaching practices cannot substitute for the other.

The third form of studies concerns meta-analyses examining the relative contribution of *both* types of practices to student learning. Seidel and Shavelson's (2007) work represents a prominent example of this form of studies. Synthesizing the results of 112 studies, this meta-analysis corroborated the above claim for combining both types of teaching practices, as its results showed both generic and content-specific practices to contribute to student learning (i.e., learning processes, cognitive outcomes, and motivational-affective outcomes) regardless of the stage of schooling (elementary or secondary) or the discipline examined (Mathematics, Science, and Language Arts).

Collectively, then, the above results reveal the importance of combining these two perspectives for better examining the effects of teaching on student learning. Although this work is still at a nascent level, and more research work is needed to better understand how the combination of these two perspectives contributes to student learning, these examples seem quite promising in helping researchers to "unpack the black box" of teaching. Having this in mind and given that such efforts are largely absent in the field of PE, this study aims at combining generic and content-specific teaching practices to explore the quality of teaching PE and, in turn, its effects on student psychomotor learning. Moreover, taking into consideration that Chen et al.'s

(2016) research work focused only on student *performance* rather than student *learning* (since only a post-test was administered to students), the present study also addresses this gap by measuring students' psychomotor *learning gains* through the use of a pre- and post-test measure.

CHAPTER 3: METHODOLOGY[1]

Abstract

To answer the research questions of this study, data were collected through a pre- and post- student performance test, classroom observations using generic and content-specific observation instruments, and a student survey capturing the extent to which certain teaching practices were used in everyday lessons. The student test was administered to the 3rd, 4th, and 5th grade students (N=944) at the beginning and culmination of the school year to measure their psychomotor growth. Three classroom observations (one per each trimester) were conducted for each of the 49 participating teachers. Each observation was carried out by three observers: one using the generic instrument of *DMEE* and the other two employing the content-specific instrument of *mTSS*. Observers and scorers were carefully selected and underwent intensive and comprehensive training before entering the study. In addition, a student survey was also used as a supplementary measurement approach of instructional quality. The survey captured students' appraisals about the extent to which their teacher used certain generic and content specific teaching practices in their PE lessons. Data analyses included four different advanced statistical techniques. First, Item Response Theory (IRT) models and Structural Equation Modeling (SEM) analyses were run to test the construct validity and the psychometric properties of the performance test and the four observation forms. Multi-level model analyses were then run to explore the individual and joint effects of generic- and content-specific practices on student learning. Finally, after classifying teachers into three categories (i.e., most-effective, typical, and least-effective), a discriminant analysis was conducted to predict which generic and content-specific teaching practices could discriminate among these three categories of teachers.

Research Setting and Participants

Research Setting

Considering that early PE experiences contribute to the development of fundamental psychomotor skills, which in turn, lay the building blocks for future physical activity and sport participation (Graham, Holt/Hale, & Parker, 2013; Griggs, 2007; MacNamara et al., 2011;

[1] This study was part of the project "*Integrating Generic and Domain-Specific Factors in Exploring the Association between the Quality of Instruction and Student Learning*" that took place in Cyprus and aimed at examining the contribution of generic and content-specific instructional aspects to student learning in Mathematics and Physical Education. During this research project I had the role of the assistant coordinator and I was one of the members of the research teams who observed and coded PE lessons by using the modified Task Structure System observation instrument and administrated the pre- and post- psychomotor tests. As an assistant coordinator, I was also responsible for ensuring the smooth operation of the daily work of the project, which involved, among others, getting in contact with the participating teachers from schools around Cyprus, assembling the schedule and promptly notifying all the members of the research team who were conducting lesson observations or administrating tests/questionnaires.

Pickup, 2012), the present study focused on elementary school PE. The main aim of elementary PE in Cyprus is to provide all students with equal opportunities to develop, improve and perform various psychomotor skills within the context of five areas of activities (i.e., educational gymnastics, dance, games, track and field, and life activities) and to simultaneously form an integrated personality through the development of fair-play principles, the cultivation of moral and social values, and the emphasis of respecting rules (Ministry of Education and Culture, 2010). Particularly, elementary PE curriculum is organized around six main aims (see Table 3.1), which cover the essential knowledge, skills, values, attitudes, and behaviors that students are expected to develop through their participation in PE lessons (Ministry of Education and Culture, 2010).

Table 3.1

The Six Main Aims of the Elementary PE Curriculum in Cyprus

No.	Aim
The student should:	
1	Develop and competently perform psychomotor skills
2	Acquire essential PE-related knowledge (e.g., principles, rules, strategies) and apply it, in order to adequately participate in present and future opportunities of physical activity
3	Acquire knowledge related to and develop a health-enhancing level of physical fitness
4	Obtain positive experience through participating in physical activity opportunities and develop self-expression and social interaction
5	Understand and respect diversity and cooperate with all the students
6	Demonstrate responsible athletic and social behavior while participating in physical activities

In all elementary grades in Cyprus (Grade 1-6) PE is a required subject. Students participate in two 40-minute periods of PE weekly during the academic year[2], which comprises 37 weeks. Typically, PE lessons are structured and delivered based on the multi-activity model (i.e., a series of short units) (Siedentop & Tannehill, 2000), something that is usually observed in

[2] At the time of the study the time allocated for PE was two 40-minute periods for all grade levels. However, from the next school year (2015-2016) and onwards, the Ministry of Education and Culture decided to increase the allocated PE time for the 5th and 6th graders to three 40-minute periods per week.

other countries as well (Hardman, 2008; Tinning, 2005). In most elementary schools in Cyprus, the outdoor PE facilities and resources are moderate to adequate. Given that a large number of schools have no gymnasiums/multipurpose halls--or if they have them, they are often used for other purposes--PE lessons are usually conducted in outdoor facilities, which typically involve a soccer ground and one or two open-air basketball grounds, which also serve as volleyball, handball, and multi-activity grounds. Consequently, teaching quality might be affected by weather conditions, since during inclement weather PE either takes place in the classroom, where simple games can be practiced or a relevant PE topic is discussed; alternatively PE lessons are cancelled (Tsangaridou & Yiallourides, 2008).

Moreover, as is the case with elementary PE around the world (Fletcher & Mandigo, 2012), PE in Cypriot elementary schools is taught by generalist teachers. According to Kyriakides and Tsangaridou (2008), generalist teachers who teach PE in Cypriot elementary schools can be classified into three groups. The first group refers to classroom teachers who teach most of the subjects to the pupils of their class including PE. The second group concerns classroom teachers who teach some subjects to the pupils of their class including PE, but they are also teaching PE to pupils of other classrooms. The third group involves PE coordinators who teach only PE lessons to students of one or more schools. As far as their initial teacher education is concerned, the majority of these teachers attended one required course during their undergraduate studies, which covered both the content and pedagogy for delivering PE in elementary schools--something that is observed in other countries as well (Graber, Locke, Lambdin, & Solmon, 2008; Hardman, 2008). Moreover, some of them had attended PE professional development sessions organized by the Ministry of Education and Culture, and a few of them have masters in PE.

As relevant literature underlines, one of the most frequent challenges that the first two abovementioned groups of teachers face when it comes to teaching PE concerns the crowded curriculum (DeCorby et al., 2005; Morgan & Hansen, 2008a). Classroom teachers are responsible for planning and delivering a number of different subject-matters, something which adversely impacts preparation time, teaching quality, and teachers' willingness to teach PE (Kyriakides, 2011; Morgan & Hansen, 2008b). In conjunction with this, having in mind that in Cyprus PE is considered a low priority subject-matter, especially when compared to Mathematics and Language Arts (Christodoulou, 2010), often times Cypriot classroom teachers use PE lessons to teach other subject matters, trying to "cover" the prescribed curriculum (Yiallourides, 1998); research suggests that this phenomenon is not uncommon in other countries (Ennis, 2006; Morgan & Hansen, 2008a).

Participants

Although the original intention of the study was to sample 50 teachers, a number that would provide enough power to run the multilevel analyses described below (Cools, De Fraine, van der Noortgate, & Onghena, 2009), a sample of 51 Cypriot generalist elementary school teachers expressed their willingness to voluntarily participate in the study and signed a related informed-consent form. However, two of the participated teachers withdrew from classroom observations for personal reasons, but they participated in all other parts of the study (i.e., student psychomotor pre- and post-measurement, and student survey). The sampled teachers were teaching the subject of PE to students of Grades 3rd, 4th, or 5th (ages: 8-11).[3]

The teacher sample included more males (87.7%) than females and teachers differed in

[3] The sample was limited to these three grades, because the psychomotor pre- and post-test administered to students involved tasks/skills that were not related to objectives set in the national PE curriculum for the first two grades, and due to difficulties in recruiting enough 6th-grade PE teachers to participate in the study.

years of teaching experience: 34.7% had up to 5 years of experience teaching PE, 22.4% had 6 to 10 years of experience, 32.7% had 11 to 15 years of experience, and 10.2% had more than 15 years of experience. Although no official data were obtained on the gender composition or the teaching experience of the teacher population teaching PE, anecdotal evidence suggests that the study's sample largely represents the teacher population under consideration, given that typically female teachers and more seasoned teachers opt to not teach PE. Teachers were enrolled in 42 schools across Cyprus and their distribution in urban (45.24%) and suburban/rural (54.76%) schools, was also representative of these types of elementary schools in the country (x^2=2.38, df=1, p=0.12). In addition, the average PE class size was 19.39 students, ranging from 11 to 25 students.

After receiving teachers' consent to participate in the study, informed-consent letters were sent to students' parents/guardians. Once all parents/guardians' signed consent/denial forms were collected, a total of 944 students (46.29% girls) consented to participate in the study. The student sample was representative of the population in terms of gender (x^2=0.002, df=1, p=0.97). Although this study also investigates variables that pertain to the socioeconomic status of students' family, no official data were available about these characteristics of the population of elementary students. As a consequence, it was not possible to examine whether the sample was representative in terms of these characteristics.

Instrumentation

Given that all instruments are to some extent limited in validity, scope, and utility of the data they yield in terms of measuring the complex and multifaceted phenomenon of teaching (Peterson, 2000; Pianta & Hamre, 2009; Veal & Anderson, 2011), and following the suggestions for combining research instruments that have been shown to have the best predictive value (i.e., observations and student surveys) (Cantrell & Kane, 2013; Coe, Aloisi, Higgins, & Major, 2014), this study used four main sets of instruments (i.e., student performance tests, generic observation instruments, content-specific observation instruments, and student survey) to reach its objectives.

Student Performance Test

To measure students' psychomotor growth a criterion-referenced performance test (pre- and post-test) measuring 3rd-, 4th-, and 5th-grade students' psychomotor skills in PE was used. Criterion-reference tests are designed to measure student performance against a predetermined set of learning standards (i.e., what students are expected to know and be able to do at a specific stage of their education) and are considered to be more appropriate than norm-referenced tests, which are standardized tests that are designed to compare and rank students in relation to one another (Hambleton & Rogers, 1991; Zhu et al., 2011). Admittedly, performance tests may not reflect performance in a realistic setting (Siedentop & Tannehill, 2000) and might not capture all facets of student learning in PE; however, they are deemed objective, and are recognized as important indicators of achievement by educators and policymakers (Rockoff, 2004; Strong, 2011).

The criterion-reference performance PE test asked students to perform six tasks (see Appendix A). Collectively, the test pertained to 13 fundamental psychomotor skills, including locomotor, non-locomotor, and manipulative skills (Gallahue & Donnelly, 2003; Graham et al., 2013) (see Table 3.2), related to objectives set in the national PE curriculum for 3rd-, 4th-, and 5th-grade students (Ministry of Education and Culture, 2013). Based on the unique nature of each skill, scoring rubrics with performance indicators (0-3) for each discrete skill were separately developed (see Appendix A), including both quantitative (i.e., how accurately the students perform a skill—e.g., accuracy of a pass) and qualitative (i.e., how correctly the students perform a skill by applying its basic learning cues) measures. Every student had one trial for each psychomotor skill and his/her trial was assessed using the constructed 0-3 rating scale. This performance test was opted to be used in this study, instead of using other international validated tests (e.g., NASPE, 2010), because it has been used in a previous study in Cyprus and has been shown to have good psychometric properties in terms of both validity and reliability (Kyriakides & Tsangaridou, 2008).

Table 3.2

The 13 Psychomotor Skills Included in the Criterion-Reference Test

Type of Skill	Skill
Locomotor	- Sliding/side gallop, skipping, standing long jump/jump forward (form), standing long jump/jump forward (distance), jumping hurdles.
Non-Locomotor	- Dodging (i.e., sharp change of direction from original line of movement), static balance on one foot, arabesque balance.
Manipulative	- Dribbling with hand (basketball), chess-passing (basketball), dribbling with foot (soccer), passing (soccer), forehand (tennis).

Observation Instruments

Given that in the field of PE classroom observations have led to important findings about the nature of effective teaching (Siedentop & Tannehill, 2000), and that they are considered an excellent measurement approach which provides contextually rich data (McKenzie & van der Mars, 2015), this study employed classroom observations as the main measurement approach to capture teaching quality for two reasons. First, regardless of being more expensive from other measurement approaches (e.g., teacher or student surveys, teacher logs, classroom artifacts) (Douglas, 2009; Peterson, 2000), and although being a complex approach accompanied with a number of complications and limitations that need to be promptly considered and encountered (McKenzie & van der Mars, 2015; Strong, 2011), classroom observations comprise the “gold standard” for studying instruction (Matsumura et al., 2006). Second, they have been found to be more prone to identifying the effects of teaching on student learning than other measurement approaches (e.g., teacher self-reports) (Seidel & Shavelson, 2007).

Therefore, to capture teaching quality in generic teaching practices, the *DMEE* framework (Creemers & Kyriakides, 2008) was employed. This model was selected for five reasons. First and foremost, the eight teaching factors incorporated in this model have been shown to predict student learning in several subject matters (e.g., Mathematics, Language Arts, Science, Religious Education) (Kyriakides & Creemers, 2008a, 2008b, 2009; Kyriakides et al., 2009; Panayiotou et al., 2014). Thus, it seems reasonable to assume that it can be successfully employed in PE as well. Second, as described in the literature review section, several teaching practices incorporated in this model have been described by PE scholars as effective principles for teaching PE, but their effect on students' PE learning still awaits empirical validation. Third, the eight factors of the *DMEE* refer to different instructional approaches (e.g., direct teaching, mastery learning); hence they can capture teaching quality irrespective of the approach employed by the teacher (Kyriakides, Christoforou, & Charalambous, 2013). Fourth, this model can distinguish between effective and less effective teachers in promoting student learning, based on the teaching practices they implement during instruction (Kyriakides et al., 2009); and fifth it is considered as the most up-to-date multilevel model of educational effectiveness (Scheerens, 2013).

Particularly, two observation forms of the *DMEE*--one high-inference and one low-inference--were used (see Appendix B). The high-inference form, which is filled out after the completion of the lesson, requires from observers a high degree of subjective judgment; the low-inference form--which is filled out during the lesson--constrains such interpretations by focusing on more readily observable behaviors (Kennedy, 2010; Veal & Anderson, 2011). Together, these two observation forms capture seven of the eight teaching factors discussed above (i.e., they do not measure the assessment factor). In particular, the low-inference observation instrument involves five teaching factors (i.e., orientation, structuring, teaching modeling, questioning, and application), each of which is captured along the five dimensions that were previously mentioned. The high-inference instrument captures seven teaching factors (with the exception of assessment), and observers are expected to complete a Likert scale anchored by 1 to 5, to indicate how often each teacher behavior is observed. The good psychometric properties of both forms of this instrument have been determined in previous national studies (e.g., Kyriakides & Creemers, 2008a; Kyriakides et al., 2009), which have also shown that these observation forms yield reliable estimates of teaching quality when used to code three lessons per teacher, each coded by a single observer.

Besides *DMEE*, to capture teaching quality in content-specific teaching practices, a modified version of the *TSS* observational rubric (Siedentop et al., 1994) was employed. The *TSS* was selected as the guiding content-specific framework for four reasons. First, it can be considered as one of the most comprehensive content-specific frameworks, since it incorporates and refines aspects from other content-specific instruments/frameworks [e.g., the *OSCD-PE* (Rink, 2010) and the *QMTPS* (Werner & Rink, 1989)]. Second, as was the case with the *DMEE*, the teaching practices involved in this framework are considered effective for teaching PE (Siedentop & Tannehill, 2000). Third, this framework takes into consideration both teacher and student behaviors that occur in PE lessons, something that a comprehensive observational tool is expected to accomplish (especially in PE) (Rink, 2013; Siedentop & Tannehill, 2000; Wood et al., 2014). Fourth, given that *TSS* is a flexible tool that can be altered to better suit researchers' purposes (Veal & Anderson, 2011), a modified version (*mTSS*) was used in this study.

Specifically, the dimension of demonstration was added to the original version of the instrument, as the review of literature suggested that it is considered a crucial teaching skill for teaching psychomotor skills (see Table 2.7 above and Table 3.3 below for a full picture of the teaching dimensions involved in the *mTSS* instrument).

Table 3.3

Description of the Demonstration Dimension of the mTSS Instrument

Dimension	Codes
Use of demonstration: captures whether and how the teacher or a mediating agent (e.g., student, video or poster) demonstrated the desired performance or described its main learning cues/critical features.	• *Partial:* refers to an incomplete demonstration of skill performance (e.g., exhibiting only part of desired movement or mentioning insufficient learning cues/critical features). • *Full:* pertains to a complete demonstration of skill performance (e.g., exhibiting full model of desired movement or mentioning all main learning cues/critical features). • *Verbal:* relates only to communicating the learning cues/critical features of the desired movement skill. • *Practical:* refers only to visual information modeling desired movement skill, without referring any learning cues/critical features. • *Combination:* relates to both practical and verbal demonstration of the desired movement skill.

As was the case with the *DMEE*, a high-inference and a low-inference observation forms were used (see Appendix C) to capture content-specific teaching practices. Beyond the five content-specific teaching practices outlined in the literature review section, the *mTSS* low-inference form also involved some generic teaching instructional aspects (e.g., time devoted to managing the class, student time on task). However, these generic teaching aspects were not dropped, as it was deemed appropriate to check whether there was consistency among these dimensions and the corresponding ones captured with the *DMEE* (something that is discussed in the next chapter). As was the case with the generic instrument which included a high-inference form, it was also considered important to develop such a form for capturing teachers' content-specific practices. The development of this high-inference form was based on the low-inference form of *mTSS* (see Tables 2.7 and 3.3) and the *AQTR* (see Table 2.9), which is a comprehensive observation instrument that involves teaching aspects that are specific to the content and context of PE teaching (Chen et al., 2011). In particular, the high-inference form involved two generic (i.e., time and classroom management) and four content specific (i.e., task progression, demonstration, congruent and specific feedback, and quality of student practice) teaching dimensions involved in the *AQTR* and the *mTSS* low-inference form[4]. Observers were asked to fill in the high-inference form using a four-point Likert-type scale anchored by 0 to 3 to indicate the degree to which the incidents described by the statements occurred in the lesson or the quality of the teacher behaviors observed. For all the codes except for those associated with the factor related to the quality of student practice, the raters were focusing on teachers' behaviors and interactions with the students; for the latter codes, the raters were each observing three specific students (of low, mid, and high level of psychomotor skills) suggested by the teacher (this process is further explained below).

The original low-inference form of *TSS* as well as the modified version were used in a series of studies (e.g., Dyson, Linehan, & Hastie, 2010; Dyson & Strachan, 2004; Hastie, 2000; Jones, 1992; Siedentop et al., 1994) yielding empirical evidence supporting the construct validity of the rubric. However, in line with recommendations for instrument adaption for use in different cultural contexts (e.g., Hambleton, 2005), both the low- and the high-inference forms were validated before being used in the Cypriot context. In addition to backwards translation, a focus-group was held with experts in sports pedagogy and practicing teachers to examine whether the practices under exploration aligned with the teaching of PE in the Cypriot educational context

[4] The content-specific teaching dimension of task explicitness was not included in the high-inference form for two reasons. First, it was not involved in the *AQTR*, and second the one statement that could be developed for capturing this practice would not be enough to form a separate factor in the Confirmatory Factor Analysis. Thus, this teaching practice was only captured through the *mTSS* low-inference form. In a similar manner, the dimension of congruent and specific feedback (represented by only one statement in the high-inference form), was assimilated into the teaching dimension of demonstration, as both of them emphasize the appropriate use of the key learning cues (see data analysis below).

(i.e., content-validation). The adapted version of these forms that resulted from the previous step were then applied to capture and code a sample of six live PE lessons to confirm that teaching practices involved in the instrument were relevant and observable to the Cypriot educational context. Results also partly confirmed the notion that much information is lost when coding live (Veal & Anderson, 2011), something (along with budget constraints) that led to the decision of using two observers per lesson observation to achieve more reliable estimates.

Student Survey

A student-survey was also used as a supplementary measurement approach of instructional quality (see Appendix D). Having in mind that teaching is a very complex process, and given that to produce a reasonably balanced and comprehensive picture of what really happens in the classroom, teachers' daily work needs to be captured (Hastie & Siedentop, 1999), students' appraisals about the extent to which certain teaching practices were used in everyday lessons were elicited.

Student surveys were preferred instead of teacher ratings, as they aggregate the impressions of many individuals who have spent many hours with a teacher, balancing in that way each other's biases (MET, 2012; Peterson, 2000). In contrast, teacher ratings are considered to suffer from the biases associated with self-report data (i.e., teachers often overestimate/inflate their use of effective practices) (Strong, 2011). As a reliable, cost- and time- efficient measurement approach (Little, Goe, & Bell, 2009; Peterson, 2000), student surveys can enhance data collected through observations (Kane & Staiger, 2012), because they can capture information pertaining to teachers' everyday teaching practices that might not be observed during the classroom observations. Despite the fact that student surveys are difficult to complete with young students due to rating leniency and halo effects, when they involve the right questions asked in appropriate ways, even third-graders can provide important, reliable and valid information on the quality of teaching and the learning environment of a classroom (Cantrell & Kane, 2013; Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Follman, 1995).

The student survey was developed based on the 12 teaching practices (i.e., seven generic and five content-specific) involved in the *DMEE* and *mTSS* instruments. For time reasons (as the survey should be completed in only one 40-minute period), only two statements were developed for each of the seven generic factors that included in the *DMEE*, the two generic factors (i.e., *time* and *classroom management*) included in the high-inference *mTSS* instrument, and two out of five content-specific (i.e., c*ongruent and specific feedback*, *quality of student practice*) teaching dimensions involved in the *mTSS*. However, one statement was developed for the task explicitness factor, because to adequately capture this teaching practice, all three aspects of providing fully explicit tasks must be involved in the same statement, whereas four statements for each of the content-specific practices of *demonstration* and *task progression* were developed to capture all the facets of these practices. In particular, for *demonstration* one statement captured the extent to which this teaching practice was employed by the teacher (frequency), two statements captured the extent in which the teacher practically and verbally (i.e., emphasizing the important learning cues) demonstrated the skill to be learned (quality) and the fourth statement captured whether the teacher emphasized the key learning cues during the whole lesson (stage). Measuring *task progression*, was deemed to be harder, because this teaching practice is not directly observable to the students. Therefore, four statements were developed, one for each type of instructional tasks (i.e., *informing, refining, extending,* and *applying*). Finally, some additional statements were developed to capture: a) teaching practices that were included in the *mTSS* low-inference instrument and might relate to student learning (i.e., the extent to which students were engaged in warm-up activities that were relevant or irrelevant to lesson's aim(s)--two statements), and b) practices that could not be observed during the three scheduled observations (i.e., students were given the opportunity to engage in free play or PE lessons were canceled because other activities had to take place--three statements); thus yielding 36 overall statements.

The survey comprised two parts (see Appendix D). The first part involved validated questions pertained to the socioeconomic status of student's family (Kyriakides & Creemers, 2008b) and students' participation in out-of-school PE activities. In the second part students were asked to answer the 36 questions related to the extent to which their teacher used certain generic and content specific teaching practices in their PE lessons.

Data Collection

Data collection processes are presented in chronological order (see Figure 3.1). First, the recruitment and training of observers and scorers[5] is discussed, followed by the process of teachers' recruitment. Then the procedure of administrating the pre-test is outlined followed by the process of conducting the classroom observations. Finally, the administration of the PE post-test and student survey is described.

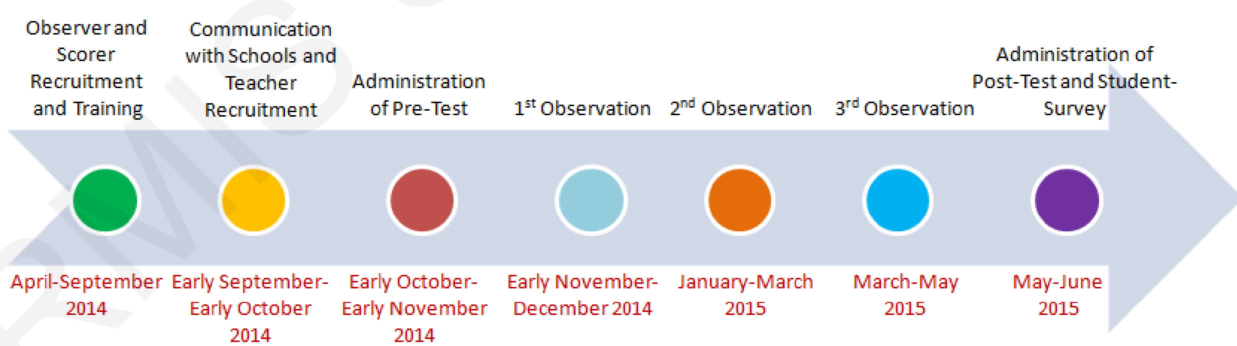


Figure 3.1. Design of the study (presented in chronological order).

[5] The term "observer" refers to the persons who conducted the classroom observations, whereas the term "scorer" is related to the persons who assessed students' ability when performing the psychomotor activities during the pre- and post-test.

Observer Recruitment and Training

Having in mind that data credibility can be ensured only when observers have completed proper training (McKenzie & van der Mars, 2015), observers were carefully selected and underwent intensive and comprehensive training before entering the study. For the purposes of this study 10 overall observers were recruited; six for using the *DMEE* instrument to capture generic teaching practices and four for using the *mTSS* instrument to capture content-specific teaching practices. Five out of six observers coding generic teaching practices were either master's or doctoral students in post-graduate degrees in Educational Administration and Evaluation, while the sixth observer was a master's student in a post-graduate degree in PE and sport pedagogy. On the other hand, all four observers recruited for coding content-specific teaching practices were either master's or doctoral students in post-graduate degrees in PE and sport pedagogy.

After expressing their willingness to participate in the study, these ten observers were invited to participate in a first meeting, where they were informed about the research project in general and were given some general directions (e.g., the need for objectivity and maintaining confidentiality were highlighted). Then, the observers were divided into two separate training groups. The first group involved the six observers who used the *DMEE* instrument and the second group the four observers that used the *mTSS* instrument.

The five out of the six observers of the first group were well trained in using both forms (low- and high-inference) of the *DMEE* instrument, as they had used them in previous studies in other content-areas. The other observer (i.e., the PE and sport pedagogy master's student) underwent a training program. In particular, she attended three two-hour training sessions with an expert user of the instrument under consideration. Before the first meeting, she read the observation manual to get familiar with all the codes. During the meetings, each dimension, code, item and scale was discussed in detail and the expert observer provided concrete examples for each code; the trainee also coded selected segments from videotaped lessons and compared/discussed her codes with expert codes. Then, the observer was asked to code two 40-minute live PE lessons, each followed by a one-hour discussion during which she compared her codes with the expert coder and they discussed possible discrepancies. Finally, the observer took an examination on the use of both forms of the *DMEE* instrument, by coding again a live PE lesson. Her ratings and expert coder ratings were in agreement with master/expert coders in more than 80% of the cases; hence, this observer was certified and included in the observer pool for the DMEE instruments.

Following Rink's (2013) suggestion for reliably using generic instruments in PE, before starting the observations, all observers of the first group participated in a two-hour training session, during which they discussed and gave examples of how the descriptors of the *DMEE* instrument can be implemented in the context of PE. For instance, during this meeting, we exhaustively discussed whether the modeling factor included teacher's demonstration of the desired psychomotor skill to students. After all observers expressed their opinion, it was agreed that the modeling factor is mainly concerned with high-order thinking skills and problem-solving. As such it should not incorporate teacher demonstration, because this teaching behavior does not challenge students as to how they can best perform the desired skill. Moreover, it was agreed that the warm-up/cool-down activities that usually take place either at the beginning or the end of the lesson should not be coded as an application activity.

Turning to the second group of observers, all four of them were trained in using both forms of the *mTSS* instrument through observing and coding videotaped and actual PE lessons.

Particularly, the observers originally attended six two-hour training sessions on using the instrument, during which they received the observation protocol including all dimensions/codes' definitions. They thoroughly discussed each dimension/code/item and scale with examples, coded selected segments from videotaped PE lessons, and compared their codes with master-coder[6] codes. Subsequently, they were asked to rate two 20-minute videotaped PE lessons taught by two different teachers using the two forms of the observation instrument at hand. Capitalizing on the observers' scores on this examination, a two-hour session was then held, during which observers discussed their discrepancies from master scores. During this session individualized as well as whole-group feedback was given to observers. After discussing and resolving these discrepancies, a series of four 40-minute live-coding lessons ensued, each followed by a one-hour discussion during which the observers compared their scores with master scores. Finally, all four observers were asked to take an inter-rater test on the use of the *mTSS* observation instrument, during which they had to code a 40-minute lesson, again in a live-coding condition. All four observers achieved the threshold of 80% of agreement with master-coder ratings, and thus, they were certified[7].

Scorer Recruitment and Training

As was the case with observers, performance scorers were also carefully selected and underwent an intense training program. In particular, eight master's or doctoral students in post-graduate PE and sport pedagogy degrees [8] were recruited and underwent training through five

[6] The master-coder was a PE expert who also had the role of the co-principal investigator of the research project: "*Integrating Generic and Domain-Specific Factors in Exploring the Association between the Quality of Instruction and Student Learning.*"
[7] The exact percentages for each observer were the following (for the low- and high-inference form, respectively): Observer A: 83% and 93%; Observer B: 95% and 87%; Observer C: 81% and 87%; Observer D: 88% and 93%.
[8] Five (out of the eight) performance raters were also observers. Four of them were using the *mTSS* instrument, while the fifth was using the *DMEE*.

two-hour sessions. In the first two meetings, scorers were informed about the general purpose of the study and about their specific role. Moreover, the scoring rubrics with performance indicators (0-3) for each discrete skill (see Appendix A) were given to them; each task/skill along with their associated indicators were thoroughly discussed. After the discussion of each skill, scorers used the rubric to score videotaped student performances relevant to the skills under consideration, and their codes were compared with master-coder codes; a session during which individualized as well as whole-group feedback was given to the scorers ensued. At the end of the second session, a number of videos were sent to the scorers via email, and the scorers were asked to code all the videotaped performances and send their scores back to the master-coder[9] before the third session. These scores were discussed at the first hour of the third session. During the second hour scorers took an inter-rater test, but the agreement with master-coder ratings fluctuated between 57% to 71%. Thus a second round of discussion was followed, and once all scorers' discrepancies with master-coder were discussed and questions were answered, all scorers took a second inter-rater test, achieving this time the minimum threshold of 80% agreement with master-coder ratings, and thus, they were certified[10].

Once certified, scorers participated in a fourth session, which took place at a multipurpose hall and aimed to inform scorers on how to set up the test's materials. In addition, during this session, scorers were divided into two groups according to their background experiences. The first group, which comprised four scorers, was responsible for coding the skills related to the first three test tasks (i.e., tasks 1-3: locomotor and non-locomotor skills), whereas the second group (another four scorers) was responsible for coding the skills pertained to the other three tasks (i.e., tasks 4-6: basketball, soccer, tennis) (see Appendix A). These groups were

[9] See footnote 6.
[10] The exact percentages for each scorer were the following: Scorer A: 86%; Scorer B: 80%; Scorer C: 80%; Scorer D: 83%; Scorer E: 86%; Scorer F: 83%; Scorer G: 80%; Scorer H: 83%.

maintained during the administration of both pre- and post- tests. The final session, which aimed at familiriazing the scorers with the live-conditions of the test administration, took place at a school setting where each group pilot coded the live performances of students from three different classes (one 3rd-, one 4th-, and one 5th-grade). Finally, due to the extended break between the pre- and the post-test, before the administration of the post-test, all scorers participated in a two-hour retraining session.

Teacher Recruitment

After obtaining the relevant permissions for conducting the study from the Ministry of Education and Culture of Cyprus, the Centre of Educational Research and Evaluation of Cyprus, and the Department of Education of University of Cyprus, all principals serving in public elementary schools were contacted via telephone and informed about the aim and the procedures of the study. In addition, it was explained to them that the study sought to recruit a specific sample, namely, a group of classroom teachers who taught both PE and Mathematics to the pupils of their class, and a group of teachers who taught only or mainly PE to students of their school[11]. After each phone conversation, an email was sent to each principal enclosing a letter which informed participants about the aim and the procedures of the study, the written permissions from the three abovementioned organizations, and the teacher and students' parents/guardians consent form. The principals were requested to talk to their personnel and encourage them to participate in the study. After receiving a signed consent form, teachers were contacted via phone and they were informed about the study (general purpose) and its logistics (e.g., the consent letters that students' parents/guardians needed to sign and return).

[11] This requirement, which was set from the larger project, part of which was this study, largely reduced the sample of available teachers and made the whole process of teacher recruitment quite difficult.

Pre-Test Administration

Once the study sample was recruited, all the students of the participating teachers, whose parents had consented to participate in the study, were administered the PE psychomotor performance test. During each administration, the performance of each student was coded by a pair of scorers, following the procedure described above. In addition, for minimizing waiting time, students of each class were split into two groups (according to their serial number in the class roster) while taking the test: the first group performed the first three test tasks (i.e., tasks 1-3, see Appendix A), at the same time that the second group performed the other three tasks (i.e., tasks 4-6); the two groups then alternated.

Lesson Observations

From early November to early May three 40-minute announced lesson observations were conducted per each participating teacher. As can be seen in Figure 3.1, for each teacher, the lessons were interspersed into the academic year, so that one lesson was observed per each trimester. Teachers were informed a week beforehand about the visit and were asked to teach typical daily lessons without any particular preparations that they would not have done, had our observers not been there. Teachers were asked to do so for two reasons; first, to avoid overloading their timetable; second, to capture what really happens during typical PE lessons, something that would help extract meaningful implications for practice, based on the findings of this study. The only stipulation given to these teachers was that their PE lessons should cover different topics from the curriculum in order to ensure that teachers would not only teach lessons from a curriculum area in which they felt more efficacious and confident.

Each lesson was observed and coded independently by three observers: an observer using both the low- and the high-inference forms of *DMEE* and two observers using both the low- and the high-inference forms of *mTSS*. As previously mentioned, decisions concerning the number of observers as well as the number of lessons per teacher to be observed using the two forms of the *DMEE* instrument were taken based on previous studies using these instruments and yielding reliable estimates of teaching quality (e.g., Kyriakides & Creemers, 2008a). The corresponding decisions concerning the two forms of the *mTSS* instrument were based on the results of the pilot-study discussed above, showing that one observer might lose important information; thus, to obtain reliable data, it was decided to have two observers per lesson and to code three lessons per teacher using these two forms of the *mTSS* instrument. Moreover, although not always feasible, each teacher's lessons was coded by a different triad of observers, avoiding, in that way, a nested relationship: observers being nested within teachers.

Finally, considering that student gender and skill level is a key moderator of learning in PE (Silverman, 2009), each of the two observers coding the content-specific practices was instructed to code the responses/trials of three students of different skill level and gender (see Table 3.4). In particular, before the beginning of each lesson, the teacher was asked to distribute numbered and colored pinafores to facilitate student identification during the coding process. To simplify the process of distribution, one of the observers advised the teacher beforehand, that the six students whose responses were to be observed should wear a certain color of pinafores (e.g., green), which should be distributed according to the Table 3.4 (the table was handed to the teacher before the lesson). The remaining numbered pinafores (having different colors) were distributed randomly or according to the needs of the lesson. Importantly, targeted students were not aware of their selection and assumed they were being observed along with the entire class.

Table 3.4

The Distribution of the Six Pinafores to Targeted Students

Pinafore's Number	Student's Gender and Skill Level
1	Male – High Level
2	Female – Low Level
3	Male – Middle Level
4	Female – High Level
5	Male – Low Level
6	Female – Middle Level

Post-Test and Student Survey Administration

At the culmination of the school year (around mid May), students of the participating teachers were administered the PE psychomotor performance test again. They were also administered a survey capturing the extent to which certain generic and content-specific teaching practices were used by their teacher in PE lessons during the school year. For the distribution of the post-test, the scorers applied the exact same procedure followed in the pre-test. As far as the student survey is concerned, this was distributed to the students of each class by a member of the research team (i.e., an observer or a scorer). After ensuring students that their anonymity will be maintained and that the teacher of the classroom will not have access to their survey, the member of the research team explained the scale of the survey by giving an example and letting the students work individually. If any further explanations were needed, the person in charge could provide additional help. For the third-graders, the member of the research team, read aloud each statement, then explained what the statement was asking them, and finally asked students to circle the answer before moving to the next statement.

Issues of Validity and Reliability

Despite that in the previous section of data collection and the following section of data analysis, several issues of validity and reliability are discussed, this section summarizes how issues of validity and reliability were addressed in this study. To begin with, issues of validity were ensured by examining the construct validity and the psychometric properties of all the employed instruments. For certain instruments different analyses and criteria were used. In particular, the validity of the performance test and the two low-inference observation forms was examined through the use of Item-Response-Theory models against a set of specific criteria (Bond & Fox, 2012), while the validity of the two high-inference observation forms and the student survey was explored throught the use of Confirmatory Factor Analyses against a set of fit indices (Hu & Bentler, 1999). In addition, as discussed above, content validity was established for the content-specific observation instrument (i.e., *mTSS*) by backward translating this instrument, holding a focus-group with experts in sports pedagogy and practicing teachers to examine whether the practices involved in this instrument aligned with the teaching of PE in the Cypriot educational context, and by pilot-testing this instrument within the Cypriot educational context (Hambleton, 2005).

Turning to issues of reliability, as previously mentioned, all observers and scorers were carefully selected, underwent intensive and comprehensive training, and were certified when at least an 80% of agreement was obtained between their ratings and master-coder ratings. Moreover, although that a check for drift was not undertaken due to budget and time constraints, in terms of having all coders code a certain number of lessons with the master-coder, after each observation, all the observers were typing their observations and sending an electronic copy of the data along with a summary (description) of the lesson to the master-coder. The master-coder carefully checked the data and in cases where a misinterpretation of a category definition or coding convention was noticed, the master-coder provided feedback to observers[12]. Furthermore, due to the extended break between the administration of the pre- and the post-test, a two-hour retraining session was held for all the scorers, to ensure the reliability of the data (McKenzie & van der Mars, 2015).

Data Analysis

To analyze the gamut of collected data, four different advanced statistical techniques were utilized, as described below.

Item-Response-Theory (IRT) Analyses

The psychometric properties of the performance test were tested by developing an IRT scale. IRT models enable the researcher to set all items and participants on the same scale (Zhu et al., 2011), and thus, determine participants' ability levels at which the test functions best (Safrit, Cohen, & Costa, 1989; Thomas, Nelson, & Silverman, 2015). In particular, the Extended Model of Rasch was utilized (Andrich, 1988) and a number of fit statistics was used to evaluate data quality. These fit statistics included whether fit mean squares (i.e., outfit and infit) of the emerging scale were close to 1; the normalized infit-*t* and outfit-*t* values had a mean of zero; the separability for test items and students was higher than 0.75; and the person estimates were well targeted against the item fit estimates (Bond & Fox, 2012).

[12] Regarding the inter-rater reliability among the two observers who used the *mTSS* during the main phase of the study, this was calculated for both the low- and the high-inference forms. As far as the low-inference form is concerned, the inter-rater reliability was calculated only for the teaching aspects of task progression-diversity of tasks, demonstration, congruent and specific feedback, and task explicitness (see Table 3.9 below). On the contrary, for the teaching aspects that were measured through time-intervals, the inter-rater reliability was not calculated, for reasons related to the complexity of such process. In addition, this percentage was not calculated for the quality of student practice, because, as described above, the two observers coded the practice trials of different students. Turning to the high-inference form, the inter-rater reliability was calculated by taking the ratio of the number of statements that were identically scored by the two observers to the number of statements that were differently scored. The overall reliability for the low-inference form was 72.01%, and for the high-inference form was 67%. Despite the latter percentage seems to be somewhat low, in the 99.99% of the observations, the observers' differences did not exceed the 1-point difference in the four-point Likert-type scale.

Table 3.5 provides a summary of the scale statistics that emerged based on the performance of 944 students to the 13 psychomotor skills. The entire sample scale was found to have high reliability for test items (r=0.96) and students (r=0.81), infit and outfit mean squares close to one, and infit and outfit t's close to zero. In addition, the fact that the separate scales for boys and girls and for third-, fourth-, and fifth-graders had acceptable values reinforced the goodness of the psychometric properties of the scale. For each subgroup the indices of reliability of cases and item separation were all above the acceptable threshold of 0.75, while the infit and outfit mean squares were close to one and the values of the infit t- and outfit t-scores were close to zero. There were only three exceptions, which were in the range of 0.10-0.16 (absolute values)--hence, they did not depart remarkably from zero.

Table 3.5

Item and Student Parameter Estimates for the Scale Developed to Capture Student Psychomotor Learning

Parameter Estimates		**Entire Sample (n=944)**	**Boys (n=507)**	**Girls (n=437)**	**Third-Graders (n=184)**	**Fourth-Graders (n=401)**	**Fifth-Graders (n=359)**
Mean	(items)	0.00	0.00	0.00	0.00	0.00	0.00
	(students)	-0.20	0.05	-0.58	-0.33	-0.16	-0.18
SD	(items)	0.37	0.42	0.68	0.40	0.36	0.40
	(students)	0.80	0.92	0.88	0.92	0.89	0.86
Reliability	(items)	0.96	0.94	0.96	0.80	0.90	0.91
	(students)	0.81	0.82	0.78	0.82	0.82	0.81
Infit mean square	(items)	1.00	1.00	1.00	1.00	1.00	1.00
	(students)	1.01	1.00	1.01	1.00	1.01	1.01
Outfit mean square	(items)	1.00	1.01	1.01	1.00	1.00	1.01
	(students)	1.00	1.01	1.01	1.00	1.00	1.01
Infit t	(items)	-0.07	-0.16	-0.09	-0.03	-0.10	-0.07
	(students)	-0.05	-0.07	-0.03	-0.05	-0.03	-0.06
Outfit t	(items)	-0.07	-0.02	0.11	-0.06	-0.07	0.00
	(students)	0.01	0.00	0.02	0.00	0.01	0.00

```
--------------------------------------------------------------------------------
Item Estimates (Thresholds)
(N =944 L = 13 Probability Level= .50)
--------------------------------------------------------------------------------
High achievement in test             Difficult items
  4.0                            |
                                 |
                                 |
                                 |
                                 |
                                 |
  3.0                            |
                                 |
                                 |
                                 |
                                 |       9.3
                                 |       3.3
  2.0                        X   |       2.3   8.3
                             X   |       1.3   10.3
                                 |       5.3
                             X   |       7.3   11.3
                            XX   |       6.3   12.3
                       XXXXXXX   |
  1.0                     XXXX   |
                         XXXXX   |
                         XXXXX   |       4.3
                        XXXXXX   |       1.2
               XXXXXXXXXXXXXXX   |       7.2   9.2   13.3
                     XXXXXXXXX   |       2.2    6.2   8.2   11.2
   .0              XXXXXXXXXXX   |       3.2
          XXXXXXXXXXXXXXXXXXXX   |       5.2   10.2   12.2
                   XXXXXXXXXXX   |
                    XXXXXXXXXX   |
                    XXXXXXXXXX   |       4.2   13.2
            XXXXXXXXXXXXXXXXXX   |       7.1
                      XXXXXXXX   |
 -1.0                   XXXXXX   |       1.1
                        XXXXXX   |       6.1   12.1
                         XXXXX   |       5.1
                           XXX   |      11.1
                                 |
                           XXX   |       9.1   10.1
 -2.0                       XX   |       2.1   8.1   13.1
                                 |       4.1
                             X   |
                                 |
                                 |       3.1
                             X   |
 -3.0                            |
                                 |
                                 |
                                 |
                                 |
                                 |
 -4.0                            |
Low achievement in test              Easy items
--------------------------------------------------------------------------------
  Each X represents   5 students
================================================================================
```

Figure 3.2. The one-parameter IRT scale developed for students' psychomotor learning.

Furthermore, as can be observed from Figure 3.2, which presents the scale for the 13 items of the student performance test for the 944 students, there was a relatively good match between the item difficulty and the person parameter estimates. Although two items (i.e., item 3.3 and item 9.3)[13] were somewhat more difficult compared to the sample's ability, students' scores ranged from –2.95 to 2.87 logits[14], while item difficulties ranged from –2.75 to 2.35 logits, suggesting that the test items sufficiently captured different difficulty levels.

To investigate whether there were differences among the performance of the three grades (i.e., 3rd, 4th, and 5th) in psychomotor test, one-way ANOVA analysis was conducted. Results indicated that there was a statistically significant difference in student performance from grade to grade (for pre-test: $F_{(2,898)}$=30.97, p<.01; for post test: $F_{(2,884)}$=34.98, p<.01). In particular, the Scheffe post-hoc test showed that the 5th graders (pre-test: $\bar{x}$ =-.31, SD=.80; post-test: $\bar{x}$ =.40, SD=.80) performed significantly (p<.01) better than the 3rd (pre-test: $\bar{x}$ =-.84, SD=.71; post-test: $\bar{x}$ =-.16, SD=.61) and 4th (pre-test: $\bar{x}$ =-.65, SD=.74; post-test: $\bar{x}$ =.10 SD=.75) graders in pre- and post-test, and that 4th graders performed significantly better than the 3rd graders (pre-test: p<.05; post-test: p<.01). Similarly, a t-test was run to identify whether there were statistically significant differences between boys and girls. The results showed that boys (pre-test: $\bar{x}$ =-.31, SD=.81 ; post-test: $\bar{x}$ =.40, SD=.81) performed significantly better than girls (pre-test: $\bar{x}$ =-.84, SD=.70; post-test: $\bar{x}$ =-.12, SD=.62) (pre-test: *t*=-10.45, df=898, p<.001; post-test: *t*=-11.04, df=871, p<.001). Collectively, the above results were not surprising, as it was anticipated that older students would perform better than the younger ones, and that boys will perform better than girls,

[13] The decimal numbers represent the item thresholds, which correspond to the four performance indicators (i.e., 0-3) (see appendix A).
[14] Values that were smaller than -3 (i.e., five scores ranged from -4.2 to -3.43 logits) or larger than 3 (four scores, all of which were equal to 3.33 logits) were not included in the this range, as they were deemed particularly low or high (total percentage of values excluded: 0.95%).

something that is in par with previous studies in elementary PE (e.g., Butterfield, Angell, & Mason, 2012; McKenzie, Alcaraz, Sallis, & Faucette, 1998; Zhu et al., 2011).

IRT was also employed to check the construct validity of the *DMEE and mTSS* low-inference observation forms. IRT was preferred instead of the Confirmatory Factor Analysis, because of two reasons. First, low-inference forms involved interval data that were turned into ordinal data, and second, initial exploratory analyses suggested the data largely formed one factor per instrument. Therefore, it was more preferable to analyze them using the IRT analysis. As far as the *DMEE* low-inference observation form is concerned, descriptive statistics showed that five teaching aspects captured by this observation form had scarcely been employed by the teachers. Four of these teaching aspects relate to the dimension of *differentiation*. In particular, the dimension of *differentiation* in *structuring* and *orientation* factors was not observed in any lesson, while in *questioning techniques* and *modeling* factors, differentiation was observed in four and one lesson(s) (out of 147 lessons), respectively. The fifth teaching aspect refers to the time that teachers had waited before students answer a question, and which was observed in only five lessons. Therefore, these five teaching aspects were dropped from Rasch analysis.

Table 3.6 provides a summary of the scale statistics that emerged based on the performance of the 49 teachers (49 teachers * 3 PE lessons = 147 lessons) to the seven generic teaching factors (26 instructional aspects) of the *DMME* low-inference. The scale was found to have good reliability for test items (r=0.87) and lessons (r=0.80), infit/outfit mean squares close to one, and infit/outfit t's close to zero (Bond & Fox, 2012). Figure 3.3 presents the scale for the 26 generic teaching aspects of the *DMEE* low-inference form (see Table 3.7 for the exact teaching aspects) for the 49 teachers (49 teachers * 3 PE lessons = 147 lessons). Teachers' scores ranged from –2.47 to 1.01 logits, while teaching practices employment ranged from –2.00 to

1.89 logits, suggesting that participating teachers used these practices at different frequencies and degrees.

Table 3.6

Item and Teacher Parameter Estimates for the Scale Developed to Capture Teaching Practices with DMEE Low-Inference Instrument

Parameter Estimates		**Total (N=147)**
Mean	(items)	0.00
	(lessons)	-0.61
SD	(items)	0.83
	(lessons)	0.78
Reliability	(items)	0.87
	(lessons)	0.80
Infit mean square	(items)	1.00
	(lessons)	0.99
Outfit mean square	(items)	1.01
	(lessons)	1.01
Infit t	(items)	-0.10
	(lessons)	-0.09
Outfit t	(items)	-0.03
	(lessons)	0.00

However, as can be observed, there were four lessons (taught by different teachers) that did not achieve the minimum threshold of the frequently observed teaching practices. On the other hand, nine teaching aspects (i.e., 1.3, 6.3, 9.3, 13.2, 18.2, 21.2, 26, 28.2 and 30.2) were rarely employed by the sampled teachers. Four of these aspects refer to the dimension of *stage* of the *structuring, questioning, modelling,* and *orientation* factors. This means that teachers did not employ these practices during all the three main phases of a lesson (e.g., introduction, core, end of the lesson). The other three teaching aspects are related to the dimension of *focus* of the *questioning, application,* and *orientation* factors, which implies that when teachers employed these teaching practices, they were more frequently relating them to one of the following aspects: a certain aim of the lesson, the entire lesson, the unit/a series of lessons, and more occasionally to two or even all three of these aspects. Finally, the other two teaching aspects pertain to the dimension of *quality* of the *questioning* and *modelling* factors. In particular, as far as the

```
--------------------------------------------------------------------------------
Item Estimates (Thresholds)
(N = 147 L = 26 Probability Level= .50)
--------------------------------------------------------------------------------
        High achievement in DMEE
        low-inference form            Not frequently observed teaching practices
  2.0                            |
                                 |      17.2
                                 |
                                 |       9.3
                                 |
                                 |      20.2
                                 |       1.3
                                 |      25
                                 |      29.2
                                 |      13.2   27.2
                                 |       6.3
  1.0                        X   |      10
                           XXX   |      21
                            XX   |
                                 |
                           XXX   |      19
                          XXXX   |
                           XXX   |
                            XX   |       2.3    9.2   12.2   24
                          XXXX   |       3.2   22     23
                       XXXXXXX   |       6.2   20.1   28     30
                            XX   |      18.2   27.1   29.1
   .0                    XXXXX   |
                        XXXXXX   |       7.2
                         XXXXX   |
                        XXXXXX   |
                           XXX   |       1.2    7.1
                                 |      12.1
                      XXXXXXXX   |       2.2   11
                    XXXXXXXXXX   |
                        XXXXXX   |
                       XXXXXXX   |
                 XXXXXXXXXXXXX   |       6.1    9.1   13.1   16.2
 -1.0                            |
                         XXXXX   |
                        XXXXXX   |
                       XXXXXXX   |      17.1
                                 |
                     XXXXXXXXX   |
                                 |
                         XXXXX   |       2.1   18.1
                                 |       1.1
                           XXX   |       3.1   15
                                 |       4
 -2.0                      XXX   |      16.1
                                 |
                             X   |
                                 |
                                 |
                           XXX   |
                                 |
                                 |
                                 |
                                 |
                                 |
 -3.0                            |
        Low achievement in DMEE
        low-inference form           Frequently observed teaching practices
--------------------------------------------------------------------------------
  Each X represents    1 lesson
================================================================================
```

Figure 3.3. The one-parameter IRT scale capturing teachers' use of generic teaching practices.

questioning factor is concerned, it seems that teachers indicated to students whether their answers were correct or incorrect more frequently than to invite students to give comments on the given answer. The results for the *modelling* factor indicate that teachers more frequently presented students a strategy for solving a problematic situation before asking them to engage into and try to solve this situation than after.

Table 3.7

Variables Used From the DMEE Low-Inference Classroom Observation Instrument in IRT Analysis

	DMEE Low-Inference Instrument
V.1	Structuring-Stage
V.2	Structuring-Frequency
V.3	Structuring-Focus
V.4	Structuring-Quality
V.5	Structuring-Differentiation
V.6	Questioning Techniques-Stage
V.7	Questioning Techniques-Frequency
V.8	Questioning Techniques-Frequency for waiting time before answering
V.9	Questioning Techniques-Focus
V.10	Questioning Techniques-Quality-Type of Question
V.11	Questioning Techniques-Quality-Teacher reaction when answer is not given
V.12	Questioning Techniques-Quality-Feedback when answer is given
V.13	Questioning Techniques-Quality-Reaction about the answer
V.14	Questioning Techniques-Differentiation
V.15	Application-Stage
V.16	Application-Frequency (no. of activities) [a]
V.17	Application-Focus
V.18	Application-Quality
V.19	Application-Differentiation
V.20	Modeling-Stage
V.21	Modeling-Frequency
V.22	Modeling-Focus
V.23	Modeling-Quality-Teacher's role
V.24	Modeling-Quality-Appropriateness of the model
V.25	Modeling-Quality-Lesson's stage that is observed
V.26	Modeling-Differentiation
V.27	Orientation-Stage
V.28	Orientation-Frequency
V.29	Orientation-Focus
V.30	Orientation-Quality
V.31	Orientation-Differentiation

Notes. [a] To measure the frequency dimension of the application factor, the number of instructional activities has been used instead of the total application time of the lesson. This is because, in lessons that involved student practice in waiting lines, the observers did not subtract the waiting time when capturing total application time, and as a result, this teaching aspect was not representative of the total practice time of students.

Turning to the low-inference observation form of the *mTSS* instrument, Table 3.8 provides a summary of the scale statistics that emerged based on the performance of the 49 teachers (49 teachers * 3 PE lessons* 2 raters = 294 scores) to the 7 teaching factors (18 teaching aspects) of this instrument. The scale was found to have satisfactory reliability for test items (r=0.84), infit/outfit mean squares close to one, and infit/outfit t's close to zero (Bond & Fox, 2012). However, the reliability for lessons was a bit lower (r=0.70) than the acceptable threshold of 0.75, something suggesting that the scale developed could discriminate among teachers (and the respective quality of their lessons) somewhat lower than the scale developed for the *DMEE*.

Table 3.8

Item and Teacher Parameter Estimates for the Scale Developed to Capture Teaching Practices with mTSS Low-Inference Instrument

Parameter Estimates		**Total (N=294)**
Mean	(items)	0.00
	(lessons)	-0.08
SD	(items)	0.54
	(lessons)	0.49
Reliability	(items)	0.84
	(lessons)	0.70
Infit mean square	(items)	1.00
	(lessons)	1.00
Outfit mean square	(items)	1.00
	(lessons)	1.00
Infit t	(items)	-0.09
	(lessons)	-0.08
Outfit t	(items)	-0.02
	(lessons)	0.00

Figure 3.4 presents the scale for the 18 teaching aspects of the *mTSS* low-inference form (see Table 3.9 for the exact teaching aspects) for the 49 teachers (49 teachers * 3 PE lessons* 2 raters = 294 scores). Teachers' scores ranged from –1.52 to 1.48 logits, while teaching practices employment ranged from –1.91 to 1.44 logits, suggesting that the teaching practices involved in this instrument were employed at different frequencies and degrees from the participating teachers. However, as one can notice, two teaching aspects (i.e., 3.1, 7.1) were more easily

```
--------------------------------------------------------------------------------
Item Estimates (Thresholds)
(N =294 L = 19 Probability Level= .50)
--------------------------------------------------------------------------------
 High achievement in mTSS
 low-inference form                   Not frequently observed teaching practices
 2.0                             |
                                 |
                                 |
                                 |
                                 |
                                 |
                                 |       8.3
                             X   |
                                 |       2.4
                            XX   |      12.2   17.3
                                 |       3.3    4     9.4
                                 |       1.3    5.3   11.3   16.3
                             X   |
                           XXX   |
  1.0                            |
                            XX   |
                          XXXX   |      10.4   14.3   15.3
                             X   |      9.3
                             X   |
                        XXXXXX   |      18.2
                      XXXXXXXX   |
                     XXXXXXXXX   |       2.3    7.3
                     XXXXXXXXX   |       1.2
                XXXXXXXXXXXXXX   |       6.3   10.3   11.2
                XXXXXXXXXXXXXX   |
               XXXXXXXXXXXXXXX   |       8.2
      XXXXXXXXXXXXXXXXXXXXXXXX   |      12.1   14.2
   .0      XXXXXXXXXXXXXXXXXXX   |       2.2
         XXXXXXXXXXXXXXXXXXXXX   |
                 XXXXXXXXXXXXX   |       3.2    5.2
                 XXXXXXXXXXXXX   |       7.2   16.2
                 XXXXXXXXXXXXX   |      9.2    15.2   17.2
  XXXXXXXXXXXXXXXXXXXXXXXXXXXX   |      11.1
                     XXXXXXXXX   |       6.2   10.2
                   XXXXXXXXXXX   |
                        XXXXXX   |
                  XXXXXXXXXXXX   |       2.1
                  XXXXXXXXXXXX   |       1.1   13.2
                        XXXXXX   |
                       XXXXXXX   |
                                 |      10.1
 -1.0                        X   |       9.1
                           XXX   |
                             X   |       6.1    8.1   15.1   16.1
                                 |
                          XXXX   |      14.1
                                 |       5.1   18.1
                                 |
                             X   |      13.1   17.1
                                 |
                                 |
                                 |
                                 |       3.1
                                 |
                                 |       7.1
 -2.0                            |
Low achievement in mTSS
low-inference form                    Frequently observed teaching practices
--------------------------------------------------------------------------------
  Each X represents    1 rater's coding of a lesson
================================================================================
```

Figure 3.4. The one-parameter IRT scale capturing teachers’ use of content-specific teaching practices.

employed compared to the sample's ability. This was expected for both teaching aspects. In particular, as far as the first step of *task progression diversity* is concerned (i.e., 3.1), this refers to teacher's implementation of just two out of four instructional tasks (i.e., *informing, refining, extending, applying)* during the whole lesson, something that is easily implemented. Similarly, the *classroom management* (i.e., 7.1) refers to issues that are irrelevant to the instructional activities (i.e., roll taking, announcements, disciplinary issues), and as such, these issues were avoided by the teachers during the classroom observations. Finally, one teaching aspect (i.e., 8.3) was somewhat above sample's ability. This teaching aspect relates to the *stage* dimension of the *skill demonstration* factor, and results showed that teachers rarely demonstrated the emphasized skill in all three main phases (i.e., introduction, core, end of lesson) of the lesson.

Table 3.9

Variables Used From the mTSS Low-Inference Classroom Observation Instrument in IRT Analysis

mTSS Low-Inference Instrument	
V.1	Time on Task (Waiting Time)
V.2	Time on Task (Total Practice Time)
V.3	Task Progression (Diversity of Tasks)
V.4	Task Progression (Practice Time in Refining Tasks) [a]
V.5	Task Progression (Practice Time in Extending Tasks) [a]
V.6	Classroom Management (Transition Time)
V.7	Classroom Management (Classroom Disorder Time)
V.8	Skill Demonstration (Stage)
V.9	Skill Demonstration (Frequency)
V.10	Skill Demonstration (Quality)
V.11	Congruent and Specific Feedback (Stage)
V.12	Congruent and Specific Feedback (Frequency)
V.13	Congruent and Specific Feedback (Quality)
V.14	Task Explicitness (Stage)
V.15	Task Explicitness (Frequency)
V.16	Task Explicitness (Quality)
V.17	Quality of Student Practice (% of Students' On-Stated Trials) [b]
V.18	Quality of Student Practice (% of Students' Appropriate Trials) [b]

Notes. [a] The variables related to practice time in informing and applying tasks were not involved in task progression factor, as according to (Rink, 2010), it is the cycle of refining-extending tasks that leads to students' psychomotor development.

[b] Quantity of student trials was not included in the IRT model, as it could not be reliably measured. This is because teachers taught a large gamut of psychomotor skills, some of which were discrete (i.e., having a clear beginning and end; e.g., throwing, kicking, passing), whilst others were continuous (i.e., having an arbitrary beginning and/or ending point; e.g., basketball dribbling, running) and it was difficult for the observers to distinguish when a student who performs a continuous skill has completed a whole trial.

Confirmatory Factor Analyses (CFA)

CFA analyses were applied to the high-inference forms of the *DMEE* and *mTSS* observation instruments and the student survey data, to test for consistency with the observed/reported data. Having in mind that the present study employed the two abovementioned frameworks and the student survey as the guiding frameworks for capturing teaching quality in this study, the purpose of this analysis was to verify the a priori factor structure of these constructs and the relationship among the observed variables and their corresponding latent factor. Therefore, it was deemed appropriate to run a CFA. The analyses were carried out using Structural Equation Modeling (SEM) techniques and the EQS 6.1 software (Bentler, 1995). Particularly, maximum likelihood (ML) method and cutoff values of a set of fit indices (i.e., the ratio of chi-square to its degrees of freedom <1.96, the Comparative Fit Index-CFI > .90, and the Root Mean Square Error of Approximation-RMSEA < .06) were used to evaluate the extent to which data fit each theoretical model under investigation (Hu & Bentler, 1999).

The *DMEE* high-inference observation form consisted of 40 items that related to 7 generic teaching factors (i.e., *orientation, structuring, application, questioning, modeling, time management,* and *classroom as a learning environment*). Before running a CFA analysis, a preliminary descriptive statistical analysis was conducted of all 40 items to investigate the extent to which each practice had a satisfactory spread across the scale (i.e., Likert 1-5) used. The results indicated that the majority of the items had been scored variously. However, five items (i.e., 16, 17, 18, 22, 25) of the *classroom as a learning environment* factor (and more particularly of the *classroom disorder* sub-factor), had been given the score “1” for the vast majority (at least 90%) of the lessons. This was not surprising, as these items refer to the extent to which students’ misbehavior was serious (i.e., had the form of verbal or bodily intimidation) or to teacher’s inability to solve students' misbehavior. As expected, these behaviors were rare, probably because of the presence of the observers. As a consequence these items were removed from CFA analysis.

Then a first order CFA model was run with the remaining 35 items to examine the construct validity of this observation form. However, during this attempt, ten items (i.e., 6, 9, 11, 14, 19, 20, 26, 31, 32, and 35, see Appendix B for the exact statements) had a relatively low loading to the targeted factor (i.e., 0.35 at most). Thus, it was decided to remove those statements and run another CFA model with the remaining 25 items. The fit indices of this attempt supported a good fit of the data to the model [χ^2=346.05, df=229, RMSEA=.059, (90% Confidence Interval of RMSEA=.046-.071), CFI=.953] and the resulting factors had at least two items each. Yet, the items related to the *classroom as a learning environment* factor were found to belong to two different factors (thus yielding eight overall factors): one measuring the types of interactions that exist in the classroom (i.e., teacher-student and student-student interactions) and the other capturing the teacher's ability to deal with student misbehavior; a result that is in par with Panayiotou et al.'s (2014) findings. However, collectively, the basic theoretical assumptions of the model were not violated.

Before exploring whether these eight factors could form one or more second-order factors, Kline's (2011) advice for determining whether the fit of simpler models was comparable was followed. In particular, a comparison was made between the delta chi-square to the delta dfs of a single-factor and a two-factor (i.e., factor A involved statements that were related to direct instruction, while factor B involved statements pertained to constructivism) first-order models[15]. However, in both simpler models, the quotient of the delta chi-square to the delta dfs was significantly above two (i.e., 10.81 and 1.77 respectively), thus the eight-factor structure was

[15] Other simpler models were not tested, due to the inherent complexity that this task involves.

considered more appropriate (single-factor model: χ^2=739.23, df=197, RMSEA=.137, (90% Confidence Interval of RMSEA=.126-.147), CFI=.702; two-factor model: χ^2=734.67, df=196, RMSEA=.137, (90% Confidence Interval of RMSEA=.126-.147), CFI=.704).

Therefore, the next step of the CFA analysis was to explore whether these eight factors could form one or more second-order factors, since one of the main assumptions of the *DMEE* is that teaching factors are interrelated (Kyriakides et al., 2009). Thus, a second-order factor solution was examined with all eight factors loading to one second-order factor, that could represent the overall quality of teaching. Table 3.10 shows that all parameter estimates were statistically significant (p.< .05), and all the standardized factor loadings were positive and moderate or moderately high, with standardized values ranging from 0.41 to 0.93 for the first-order factor structure and from 0.40 to 0.99 for the second-order factor structure; supporting, thus, the validity of this model.

Table 3.10

Maximum Likelihood Estimates for an Eight-Factor Solution of the DMEE High-Inference Classroom Observation Data and Fit Statistics for the Model

Parameter	Unstandardized	SE	Standardized
	Factor loadings		
Orientation, F1([a])			
Item 1	1.000[b]	---	.452
Item 27	1.509	0.274	.805
Item 28	2.020	0.355	.899
Structuring, F2([a])			
Item 2	1.000 [b]	---	.444
Item 3	1.288	0.265	.811
Application, F3([a])			
Item 4	1.000 [b]	---	.533
Item 5	0.797	0.088	.411
Item 7	0.654	0.177	.445
Questioning Techniques, F4([a])			
Item 33	1.000 [b]	---	.540
Item 34	1.106	0.219	.827
Modeling, F5([a])			
Item 36	1.000 [b]	---	.702
Item 37	1.713	0.164	.908
Item 38	1.705	0.211	.890
Item 39	1.941	0.182	.930
Item 40	0.434	0.067	.570
Time Management, F6([a])			
Item 8	1.000 [b]	---	.804
Item 29	0.881	0.106	.687
Item 30	1.426	0.156	.872
Teacher-Student and Student-Student Interactions, F7([a])			
Item 10	1.000 [b]	---	.740
Item 12	0.993	0.164	.510
Item 13	1.486	0.180	.717
Classroom Disorder, F8([a])			
Item 15	1.000 [b]	---	.724
Item 21	1.427	0.139	.869
Item 23	1.608	0.145	.903
Item 24	1.741	0.196	.751
Quality of Teaching, F9([a])			
F1	0.494	0.089	.990
F2	0.638	0.132	.884
F3	0.583	0.109	.813
F4	0.700	0.144	.739
F5	0.358	0.064	.515
F6	0.508	0.053	.853
F7	0.630	0.068	.921
F8	0.238	0.056	.400

Table 3.10 Continued

	Measurement error variances		
Item 1	0.966	0.115	.892
Item 27	0.308	0.043	.594
Item 28	0.242	0.049	.439
Item 2	2.119	0.261	.896
Item 3	0.449	0.148	.585
Item 4	1.292	0.217	.846
Item 5	1.604	0.220	.912
Item 7	0.892	0.123	.896
Item 33	2.181	0.294	.842
Item 34	0.506	0.190	.562
Item 36	0.496	0.062	.712
Item 37	0.299	0.044	.418
Item 38	0.367	0.058	.456
Item 39	0.284	0.045	.368
Item 40	0.188	0.023	.822
Item 8	0.195	0.037	.595
Item 29	0.309	0.039	.727
Item 30	0.228	0.059	.490
Item 10	0.387	0.058	.672
Item 12	1.318	0.160	.860
Item 13	0.976	0.148	.697
Item 15	0.332	0.041	.690
Item 21	0.242	0.048	.496
Item 23	0.214	0.046	.429
Item 24	0.854	0.125	.660
F1	0.005	0.010	.142
F2	0.113	0.082	.467
F3	0.174	0.153	.582
F4	0.407	0.150	.674
F5	0.353	0.075	.857
F6	0.097	0.029	.522
F7	0.071	0.039	.391
F8	0.309	0.061	.919

χ^2_M	df_M	χ^2_M / df_M	*RMSEA (90% CI)*	*CFI*	*SRMR*
372.22	248	1.50	0.058 (0.046 - 0.070)	0.950	0.070

Notes. [a] See Appendix B for the exact statements used in each factor.
[b] Not tested for statistical significance.

Turning to the *mTSS* high-inference observation form, this consisted of 19 statements and pertained to five instructional dimensions, two of which were generic (i.e., *classroom* and *time management*), and three of which were content-specific (i.e., *task progression*, *demonstration of desired movement skills*, and *quality of student practice*). Thus, a first step was to conduct a first-order CFA model to examine whether the results provided support to the construct validity of this observation form. Fit indices supported a good fit of the data to the model [χ^2=215.11, df=136, RMSEA=.045, (90% Confidence Interval of RMSEA=.033-.055), CFI=.963]. However, following Kline's (2011) advice for determining whether the fit of a simpler model was comparable, a comparison was made between the delta chi-square to the delta dfs of the simpler models. In all simpler models, the quotient of the delta chi-square to the delta dfs was above 2 (see Table 3.11), thus, the five-factor structure was considered as the most appropriate.

Table 3.11

Values of Selected Fit Statistics for Alternative mTSS First-Order Models

Model	χ^2_M	df_M	χ^2_D [b]	df_D	RMSEA (90% CI)[a]	CFI	SRMR
5f-model	215.11	136	---	---	.045 (.033 - .055)	.963	.049
1f-model	801.11	146	586.00	10	.124 (.115 - .132)	.697	.100
2f-model	562.88	145	347.77	9	.099 (.090 - .108)	.807	.097
3f-model	396.87	143	181.76	7	.078 (.069 - .087)	.883	.072
4f-model	388.80	140	173.69	4	.078 (.069 - .087)	.885	.083

Notes. [a] CI: Confidence Interval.
[b] All differences significant at $p < .001$

After determining the first-order factor structure, then a second-order factor solution was tested, to examine whether the two generic and the three content-specific factors could form two second-order factors. As can be observed in Table 3.12, the results supported this structure. All parameter estimates were statistically significant ($p < .05$) for both the first- and the second-order factor structure, and all the standardized factor loadings were positive and moderate or moderately high, with standardized values ranging from 0.40 to 0.89 for the first-order factor structure and from 0.63 to 0.87 for the second-order factor structure, with the only exception of quality of student practice factor (i.e., 0.32). However, the loading of this factor was expected to be somewhat lower than the other two factors, because this factor refers to what students do (i.e., quality of student practice), whereas the other two factors of the content-specific second-order factor pertain to teacher behaviors.

Table 3.12

Maximum Likelihood Estimates for a Five-Factor Solution of the mTSS High-Inference Classroom Observation Data and Fit Statistics for the Model

Parameter	Unstandardized	SE	Standardized
	Factor loadings		
Classroom Management, F1([a])			
Item 4	1.000[b]	---	.636
Item 5	0.620	0.088	.548
Item 6	1.145	0.143	.727
Item 8	0.688	0.105	.499
Time Management, F2([a])			
Item 7	1.000 [b]	---	.892
Item 12	0.954	0.084	.890
Item 19	0.854	0.061	.761
Task Progression, F3([a])			
Item 13	1.000 [b]	---	.789
Item 14	0.451	0.059	.565
Item 15	1.036	0.101	.822
Item 16	1.139	0.051	.805
Demonstration of the desired movement skill, F4([a])			
Item 1	1.000 [b]	---	.505
Item 2	1.344	0.176	.742
Item 3	1.368	0.176	.769
Item 9	0.921	0.121	.714
Item 10	1.013	0.142	.625
Quality of student practice, F5([a])			
Item 11	1.000 [b]	---	.404
Item 17	2.393	0.716	.580
Item 18	2.642	0.751	.463
Generic Instructional Aspects Practices, F6([a])			
F1	0.272	0.045	.630
F2	0.425	0.058	.631
Content-Specific Teaching Practices, F7([a])			
F3	0.441	0.056	.690
F4	0.476	0.069	.872
F5	0.035	0.013	.323
	Measurement error variances		
Item 4	0.276	0.031	.772
Item 5	0.168	0.016	.836
Item 6	0.219	0.031	.687
Item 8	0.268	0.025	.867
Item 7	0.116	0.037	.451
Item 12	0.109	0.033	.456
Item 19	0.240	0.035	.648

Table 3.12 Continued

Item 13	0.248	0.038	.615
Item 14	0.177	0.018	.825
Item 15	0.210	0.040	.570
Item 16	0.286	0.048	.593
Item 1	0.868	0.078	.863
Item 2	0.439	0.054	.670
Item 3	0.386	0.050	.640
Item 9	0.243	0.027	.700
Item 10	0.477	0.046	.781
Item 11	0.059	0.006	.915
Item 17	0.130	0.022	.815
Item 18	0.292	0.034	.886
F1	0.113	0.028	.777
F2	0.273	0.053	.776
F3	0.213	0.043	.723
F4	0.072	0.036	.490
F5	0.010	0.004	.947
		Factor covariances	
F6 ↔F7	0.716	0.087	.716

χ^2_M	df_M	χ^2_M / df_M	*RMSEA (90% CI)*	*CFI*	*SRMR*
231.54	140	1.65	0.047 (0.036 - 0.058)	0.958	0.055

Notes. [a] See Appendix C for the exact statements used in each factor.
[b] Not tested for statistical significance. All other unstandardized estimates are statistically significant at p< .05.

Finally, as explained above, the student survey consisted of 36 items that related to the 12 teaching factors (i.e., seven generic and five content-specific practices) involved in the two observation instruments employed, and some other teaching aspects as mentioned above. However, ten items were removed from the CFA analysis for different reasons. Particularly, the five items (i.e., 1, 6, 20, 24, 32) that pertained to other teaching aspects than the 12 involved in the observation instruments and the one (i.e., 15) that pertained to teachers' explicitness of tasks[16] were removed because they were not expected to load into any factor. In addition, from the administration it was noticed that the four items that referred to the task progression (i.e., 16, 30, 31, 33) were not understood by a number of students (especially third-graders); thus it was

[16] These items were removed from the CFA analysis. However, their contribution to student psychomotor learning was investigated, by entering each individual item into the multi-level analysis.

decided to be removed from the analysis. Thus, the remaining 26 items were analyzed with CFA, to investigate the construct validity of the survey. Yet, beyond the *classroom as a learning environment* and *time management* factors which were represented by more than two items, since similar items were also developed based on the *mTSS* instrument, the other five generic teaching practices were represented with only two statements. As a consequence, these pairs of items could not form single factors (Kline, 2011). Hence, it was decided to investigate whether all five generic teaching practices could form a single factor. In a similar manner, the two items that pertained to *congruent and specific feedback* were merged with the items that refer to *demonstration*, as both practices emphasize the key learning cues. Thus, the first-order factor model investigated the following construct: a) generic teaching practices, b) *demonstration* and *congruent and specific feedback*, c) *quality of student practice*, d) *time management*, and e) *classroom management.* After running this model, ten items had a rather low loading (i.e., <.30), thus it was decided to remove them from further analyses. Six of these items (i.e., 6, 8, 10, 14, 23, 25) pertained to *generic teaching practices*, one item (i.e., 22) to *time management*, one item (i.e., 9) to *demonstration and congruent specific feedback* and two items (i.e., 12 and 27) to *classroom management* factor. In addition, because the correlation between *classroom and time management* factors was high (i.e., $r > .80$) it was decided to test whether those two factors could form a single one. Therefore, another CFA model was run with the remaining 16 items, and which investigated whether a four-factor model could be constructed. Fit indices supported a good fit of the data to the model [χ^2= 276.77, df=97, RMSEA=.043, (90% Confidence Interval of RMSEA=.037-.049), CFI=.953]. To determine whether the fit of a simpler model was comparable, a comparison was made between the delta chi-square to the delta dfs of the simpler models, which confirmed that the four-factor solution was the most appropriate, as in all simpler models, the quotient of the delta chi-square to the delta dfs was above 2 (see Table 3.13).

Table 3.13

Values of Selected Fit Statistics for Alternative First-Order Models of Student Survey

Model	χ^2_M	df_M	χ^2_D [b]	df_D	RMSEA (90% CI)[a]	CFI	SRMR
4f-model	276.77	97	---	---	.043 (.037 - .049)	.953	.035
1f-model	598.90	103	322.13	6	.070 (.064 - .075)	.870	.051
2f-model (Factors 1,2,4, were put together)	478.03	102	201.26	5	.061 (.056 - .067)	.901	.046
3f-model_A (Factors 1 and 2 were put together)	417.53	100	140.76	3	.057 (.051 - .062)	.917	.042
3f-model_B (Factors 1 and 4 were put together)	343.76	100	66.99	3	.050 (.044 - .055)	.936	.039
3f-model_C (Factors 2 and 4 were put together)	371.22	100	94.45	3	.052 (.047 - .058)	.929	.041

Notes. [a] CI: Confidence Interval.
[b] All differences significant at p <.001

After testing the model against the alternative simpler models, a second-order factor solution was examined to explore whether the four first order factors were related to each other. The results indicated that the three out of the four first order factors (except from the factor pertained to quality of student practice) loaded on a second-order factor (see Table 3.14). In addition, all parameter estimates were statistically significant and the standardized factor loadings were positive and moderate to high.

Table 3.14

Maximum Likelihood Estimates for a Four-Factor Solution of the Student Survey Data and Fit Statistics for the Model

Parameter	Unstandardized	SE	Standardized
	Factor loadings		
Generic Teaching Practices, F1([a])			
Item 2	1.000 [b]	---	.592
Item 4	0.961	0.067	.593
Item 17	1.064	0.069	.666
Item 19	1.074	0.073	.624
Item 21	0.874	0.064	.559
Demonstration and Congruent and Specific Feedback, F2([a])			
Item 3	1.000 [b]	---	.579
Item 13	0.948	0.066	.598
Item 18	1.037	0.069	.642
Item 35	1.131	0.071	.704
Item 36	1.024	0.067	.653
Quality of Student Practice, F3([a])			
Item 28	1.000 [b]	---	.615
Item 34	1.321	0.105	.777
Time and Classroom Management,F4([a])			
Item 5	1.000 [b]	---	.433
Item 11	0.902	0.114	.403
Item 26	1.161	0.118	.603
Item 29	0.980	0.106	.488
Quality of Teaching, F5([a])			
F1	0.761	0.046	.870
F2	0.703	0.042	.920
F4	0.523	0.048	.803
	Measurement error variances		
Item 2	1.419	0.073	.806
Item 4	1.302	0.067	.805
Item 17	1.087	0.061	.746
Item 19	1.388	0.074	.782
Item 21	1.291	0.065	.830
Item 3	1.160	0.058	.816
Item 13	0.941	0.048	.801
Item 18	0.896	0.047	.767
Item 35	0.760	0.043	.710
Item 36	0.823	0.044	.757
Item 28	0.096	0.165	.789
Item 34	0.197	0.024	.629
Item 5	0.228	0.038	.902
Item 11	1.775	0.091	.915
Item 26	1.000	0.065	.798
Item 29	1.304	0.068	.873
F1	0.187	0.035	.494
F2	0.090	0.024	.393
F4	0.150	0.034	.595
	Factor Covariates		
F3 ↔F5	0.457	0.037	.681

χ^2_M	df_M	χ^2_M / df_M	*RMSEA (90% CI)*	*CFI*	*SRMR*
311.025	99	3.14	0.047 (0.041 - 0.052)	0.944	0.038

Notes. [a] See Appendix D for the exact statements used in each factor.
[b] Not tested for statistical significance. All other unstandardized estimates are statistically significant at $p < .05$.

Multilevel Analyses

After running the above analyses, multi-level model analyses were run (Luke, 2004) by using ‘MLwiN’ software (Rasbash, Steele, Browne, & Goldstein, 2012), to examine the value that teachers have added to student psychomotor learning over the academic year (Amrein-Beardsley, 2008; Strong, 2011); a key aim in this analysis was also to explore the individual and joint effects of generic- and content-specific practices on student learning. Initially, due to the hierarchical structure of the data, a two-level model (students nested within teachers) was run with students’ post-test performance as the dependent variable. This model would provide the opportunity to explore where (at what level) and how effects are occurring. Omitting the school level was reasonable both because there were several schools in which only one PE teacher was sampled (i.e., 34 out of 42) and because of the present study’s interest in exploring the effect of different teaching practices, which are situated at the teacher rather than at the school level. Once the empty model was established, some explanatory variables found to explain most of the variance at the student level (Scheerens & Bosker, 1997) were then introduced into Model 1 (i.e., student pre-test performance and background characteristics--age, gender, and socio-economic status). Moreover, to control for out-of-school PE experiences and classroom effects, in Model 2, the variable pertaining to student’s participation in out-of-school PE activities and the variables pertaining to classroom composition (i.e., class size, aggregated pre-test performance from student level data, percentage of girls in the classroom, aggregated socio-economic status) were additionally entered.

After controlling for variables entered in Models 1 and 2, the variables obtained through observations and student ratings, and which pertain to generic and content-specific teaching practices were entered in Models 3a to 3*n*[17]. By comparing each of the latter models to Model 2, it was possible to investigate the extent to which generic and content-specific practices, *individually*, explained a greater proportion of the variance in student learning compared to that explained without using any such variables; thus addressing the first research question. Then, in Model 3_{n+1}, both generic and content-specific variables found to have a statistically significant contribution to student learning in Models 3a and 3*n* were entered. The comparison of Model 3_{n+1} to Models 3a to 3*n* helped answer the second research question, which related to the *added value* of exploring *both* types of teaching practices as opposed to considering each type in isolation.

Discriminant Analysis

The third research question was addressed by employing a discriminant analysis to determine those teaching practices (generic and/or content-specific) that could discriminate among most-, typical, and least- effective teachers. In particular, a discriminant analysis was undertaken (Burns & Burns, 2008) to identify those teaching practices (from among all the practices found to significantly contribute to student psychomotor learning in the multilevel analyses) that could contribute most to this separation. Differently put, it was explored whether the differences among the effectiveness of these three groups could be explained by the employment of certain teaching practices.

To this end, teachers were classified into three categories (i.e., most-effective, typical, and least-effective) based on the residuals at the teacher level (Goldstein, 2003), resulting from Model 2 described above (before entering any variables related to teaching practices). In

[17] Due to multicollinearity problems, a staged process was followed, by adding one by one the teaching factors/ aspects of each instrument, and comparing each alternative model to the Model 2 (see Creemers, Kyriakides, & Sammons, 2010 for a similar approach). Therefore, *n* represents a different number of models, as the exact number of alternative models depends on how many teaching factors/aspects were involved in the instrument under exploration.

particular, using the residuals of each teacher's estimate score, teachers were classified as follows: least-effective (N= 15): residual ≤ -1SD; typical (N=24): -1SD ≤ residual ≥ 1SD; most-effective: residual ≥ 1SD (N=10).

CHAPTER 4: RESULTS

Abstract

The present study had three main aims. The first aim was to investigate the individual contribution of generic and content-specific teaching practices to student psychomotor learning. The study's results that relate to this aim corroborated existing research findings which highlighted the contribution of certain generic and content-specific teaching practices to student psychomotor learning. Moreover, the study's findings provided empirical evidence pointing to the effect of some underexplored generic and content-specific teaching practices on student psychomotor learning. In particular, orientation, time management, and questioning (from generic teaching practices) as well as demonstration and congruent and specific feedback (from content-specific teaching practices) were found to have the largest effects on student psychomotor learning. The second main aim of this study was the investigation of the joint contribution of generic and content-specific teaching practices to student psychomotor learning. Results that relate to this aim indicated that the combination of the two types of practices could explain more variance at the teacher level compared to that explained when considering either type of practices in isolation. The third and final aim of this study was to explore which teaching practices could serve as the best predictors that could discriminate among teachers based on their level of effectiveness. Findings showed that certain generic and content-specific practices could predict allocation of teachers to most-effective or non-most effective (i.e., typical or least-effective).

This chapter presents the study results, and therefore, addresses the three research questions that the study sought to answer. However, before presenting the results that relate to the three research questions, some descriptive statistics are presented, concerning the degree to which each generic and content-specific teaching practice was employed by the teachers, followed by a brief presentation of the student learning variance decomposition, and the contribution of student- and teacher-background characteristics to student learning. Then, the presentation of results is organized per each research question. In particular, the contribution of generic teaching practices on student psychomotor learning is presented, followed by the examination of the contribution of content-specific teaching practices. In doing so, the first research question, which seeks to determine the *individual* contribution of either type of practices to student psychomotor learning, is being addressed. Second, the extent to which the combination of generic and content-specific teaching practices explains a greater proportion of variance in student learning is explored; this exploration helps address the second research question, which searches for the *joint* contribution of both types of practices to student psychomotor learning. Finally, the degree to which the employment of certain generic and/or content-specific teaching practices can discriminate among teachers based on their level of effectiveness is examined, thus addressing the third research question.

Descriptive Statistics

Tables 4.1 and 4.2, respectively, present how frequently each of the five generic teaching practices involved in the low-inference form of *DMEE* and each of the seven generic and content-specific practices included in the low-inference form of *mTSS* were employed by the participating teachers[18]. However, it is noted that within this section only some notable patterns or observations that emerged from these two tables will be discussed, as the descriptive statistics of both generic and content-specific teaching practices are presented along with the presentation of study's findings in the following sections.

To begin with, examining the degree to which the five generic teaching practices of the *DMEE* occurred in PE lessons (see column *did not occur* of *stage* dimension in Table 4.1), one can observe that questioning, modeling, and orientation were only observed in a small percentage of lessons, whereas structuring was observed in the majority of the lessons and application opportunities were offered in all the observed lessons. Focusing on the stage dimension, except from the teaching practice of application, which was typically observed in all

[18] The descriptive statistics refer only to the low-inference form of the two instruments, and not to the high-inference form, since the former captures more readily observable behaviors than the latter, which is filled out after the completion of the lesson and requires from observers a high degree of subjective judgment (Kennedy, 2010).

Table 4.1

Descriptive Statistics for the Five Factors and Their Constituent Aspects/Dimensions of the DMEE Low-Inference Form

Stage Dimension				
	Did not occur	Occurred in one main juncture (i.e., introduction, core, end) of the lesson	Occurred in two main junctures (i.e., introduction, core, end) of the lesson	Occurred in all three main junctures (i.e., introduction, core, end) of the lesson
Structuring	23.8%	35.4%	32.7%	8.2%
Questioning Techniques	43.5%	30.6%	17.0%	8.8%
Application	0.0%	2.7%	22.4%	74.8%
Modeling	70.1%	23.8%	5.4%	0.7%
Orientation	68.7%	21.8%	7.5%	2.0%
Frequency Dimension				
	Did not occur	Occurred once during the lesson	Occurred twice during the lesson	Occurred three or more times during the lesson
Structuring	23.8%	30.6%	23.8%	21.8%
Questioning Techniques	43.5%	14.3%	23.8%[a1]	18.4%[a2]
Questioning Techniques – Waiting time before answering	96.6%[b1]	2.0%[b2]	1.4%[b3]	0.0%[b4]
Application (no. of activities)	15.0%[c1]	28.5%[c2]	32.6%[c3]	23.8%[c4]
Modeling	70.1%	10.2%	7.5%	12.2%
Orientation	68.7%	14.3%	11.6%	5.4%
Focus Dimension				
	Did not occur	Related to one of the following aspects: a) a certain aim of the lesson, b) the entire lesson, and c) the unit/a series of lessons)	Related to two of the following aspects: a) a certain aim of the lesson, b) the entire lesson, and c) the unit/a series of lessons)	Related to all three of the following aspects: a) a certain aim of the lesson, b) the entire lesson, and c) the unit/a series of lessons)
Structuring	23.8%	49.0%	25.2%	2.0%
Questioning Techniques	43.5%	34.0%	18.4%	4.1%
Application	0.0%	35.4%	57.8%	6.8%
Modeling	70.1%	27.2%	2.7%	0.0%
Orientation	68.7%	23.1%	8.2%	0.0%
Quality Dimension				
	Did not occur	1 (see notes at the end of the table)	2 (see notes at the end of the table)	3 (see notes at the end of the table)
Structuring	23.8%	2.0%[d]	74.1%[d]	-
Questioning Techniques – Type of Question	43.5%	38.8% [e]	13.6% [e]	4.1% [e]
Questioning Techniques – Teacher reaction when answer is not given	43.5%	0.0% [f]	11.5% [f]	45.0% [f]

Table 4.1 Continued

Quality Dimension (continued)				
	Did not occur	1 (see notes at the end of the table)	2 (see notes at the end of the table)	3 (see notes at the end of the table)
Questioning Techniques – Feedback when answer is given	43.5%	10.9% [g]	19.8% [g]	25.8% [g]
Questioning Techniques – Reaction about the answer	43.5%	0.0% [h]	53.8% [h]	2.7% [h]
Application	0.0%	68.0% [i]	32.0% [i]	-
Modeling –Teacher's role	70.1%	16.3% [j]	12.2% [j]	1.4% [j]
Modeling – Appropriateness of the model	70.1%	2% [k]	27.9% [k]	-
Modeling –Lesson's stage that is observed	70.1%	16.4% [l]	13.5% [l]	-
Orientation	68.7%	10.9% [m]	20.4% [m]	0.0% [m]

Differentiation Dimension				
	Did not occur	Occurred once during the lesson	Occurred twice during the lesson	Occurred three or more times during the lesson
Structuring	100%	0.0%	0.0%	0.0%
Questioning Techniques	97.3%	1.4%	0.7%	0.7%
Application	76.2%	10.2%	8.2%	5.4%
Modeling	99.3%	0.7%	0.0%	0.0%
Orientation	100%	0.0%	0.0%	0.0%

Notes.[a1] Occurred twice or thrice during the lesson, [a2] Occurred four or more times during the lesson.
[b1] 0 minutes, [b2] 1 minute, [b3] 2 minutes [b4] three or more minutes.
[c1] 1-3 activities, [c2] 4-5 activities, [c3] 6-7 activities, [c4] more than 8 activities.
[d] During the lesson the teacher provided structuring activities that were:1= not clear for the students, 2= clear for the students.
[e] During the lesson the teacher mainly posed: 1= product questions, 2= mixed of product-process questions, 3= process questions.
[f] During the lesson the teacher mainly: 1= moves to another question or answers the question him/herself, 2= poses an easier question or restates a question (easier words), 3= no reaction needed-answer was given.
[g] During the lesson the teacher provided: 1= no comments or negative comments to incorrect and partly correct answers, 2= positive comments to correct answers only, 3= positive comments to correct answers and constructive comments to incorrect and to partly correct answers.
[h] During the lesson the teacher mainly: 1= ignores the answer, 2= indicates that the answer is correct or partly correct or incorrect, 3= invites students to give comments on the answer.
[i] During the lesson the teacher mainly: 1= uses the same activity to help students find a specific result, 2= activates students' certain cognitive processes for the solution of more complex activities.
[j] During the lesson the teacher mainly: 1= gives the strategy to students, 2= engages students in guided discovery, 3= engages students in discovery.
[k] During the lesson the teacher mainly provides: 1= an unsuccessful model, 2= a successful model.
[l] During the lesson the teacher initiates a strategy mainly: 1= before a problematic situation, 2= after a problematic situation.
[m] During the lesson the teacher mainly engages students in orientation activities that: 1= are typical, 2= are related to learning, 3= ask for them to specify the aim(s).

Table 4.2

Descriptive Statistics for the Seven Factors and Their Constituent Aspects/Dimensions of the mTSS Low-Inference Form

Stage Dimension				
	Did not occur	Occurred in one main juncture (i.e., introduction, core, end) of the lesson	Occurred in two main junctures (i.e., introduction, core, end) of the lesson	Occurred in all three main junctures (i.e., introduction, core, end) of the lesson
Skill Demonstration	19.7%	39.1%	31.6%	9.5%
Congruent and Specific Feedback	41.8%	23.8%	22.4%	11.9%
Task Explicitness (fully explicit tasks)	16.7%	41.2%	21.4%	20.7%
Frequency Dimension				
	Did not occur	Occurred once during the lesson	Occurred twice during the lesson	Occurred three or more times during the lesson
Skill Demonstration	19.7%	22.8%	21.1%	36.4%
Congruent and Specific Feedback	41.8%	15.0%	13.3%	29.9%
Task Explicitness (fully explicit tasks)	16.7%	27.6%	16.7%	39.1%
Quality Dimension				
	0 or Did not occur (see notes at the end of the table)	1 (see notes at the end of the table)	2 (see notes at the end of the table)	3 (see notes at the end of the table)
Skill Demonstration	19.7%	20.1%[a]	36.4%[a]	23.8%[a]
Congruent and Specific Feedback	1.7%[b]	23.8%[b]	46.9%[b]	27.6%[b]
Task Explicitness	0.0%[c]	65.6%[c]	34.4%[c]	-
Time on Task (Waiting Time)	29.9%[d]	37.8%[d]	20.4%[d]	11.9%[d]
Time on Task (Total Practice Time)	28.9%[e]	24.8%[e]	18.4%[e]	27.9%[e]
Task Progression (Diversity of Tasks)	9.5%[f]	39.8%[f]	36.4%[f]	14.3%[f]
Task Progression (Practice Time in Refining Tasks)	78.9%[g]	21.1%[g]	-	-
Task Progression (Practice Time in Extending Tasks)	15.0%[h]	33.0%[h]	37.0%[h]	15.0%[h]
Classroom Management (Transition Time)	3.7%[i]	29.0%[i]	54.7%[i]	12.6%[i]
Classroom Management (Classroom Disorder)	29.9%[j]	38.8%[j]	23.2%[j]	8.2%[j]
Quality of Student Practice (% of Students' On-Stated Trials)	2.7%[k]	5.1%[k]	19.1%[k]	73.1%[k]
Quality of Student Practice (% of Students' Appropriate Trials)	10.9%[l]	26.9%[l]	35.3%[l]	26.9%[l]

Notes.[a] During the lesson the teacher or a student mainly demonstrated the skill by making: 1= a partial reference to the critical cues or a partial practical demonstration, 2= a combination of partial reference to the critical cues and

partial practical demonstration or only a full reference to the critical cues without practical demonstration or a full practical demonstration without any reference to the critical cues, 3= a combination of full reference to the critical cues and full practical demonstration.
[b] During the lesson teacher's accountability of student practice mainly involved: 0= no supervision, 1= just monitoring, 3= reinforcement, 4= reinforcement and congruent and specific feedback.
[c] During the lesson the teacher mainly provided task directions that were: 0= implicit, 1= partially explicit, 2= fully explicit.
[d] The waiting time of the lesson: 0= exceeded the 40% of lesson time, 1= ranged from 25.01%-40% of lesson time, 2= ranged from 15.01%-25% of lesson time, 3= did not exceed the 15% of lesson time.
[e] The total practice time of the lesson: 0= did not exceed the 10% of lesson time, 1= ranged from 10.01%-20% of lesson time, 2= ranged from 20.01%-30% of lesson time, 3= exceed the 30% of lesson time.
[f] During the lesson the teacher provided: 0= just one type of instructional tasks, 1= two types of instructional tasks, 2= three types of instructional tasks, 3= all four types of instructional tasks.
[g] During the lesson the teacher: 0= did not engage students in refining tasks, 1= engaged students in refining tasks.
[h] During the lesson the teacher: 0= did not engage students in extending tasks, 1= engaged students in extending tasks for up to two minutes, 2= engaged students in extending tasks between two to eight minutes, 3= engaged students in extending tasks more than eight minutes.
[i] The transition time of the lesson: 0= exceeded the 30% of lesson time, 1= ranged from 20.01%-30% of lesson time, 2= ranged from 10.01%-20% of lesson time, 3= did not exceed the 10% of lesson time.
[j] The time devoted to issues that were not relevant to the instructional activities (e.g., announcements irrelevant to lesson's goals, roll taking): 0= did not occur, 1= did not exceed one minute, 2= ranged from one to two minutes, 3= exceeded the two minutes of lesson time.
[k] The on-stated trials of targeted students: 0= did not exceed the 50% of their total trials, 1= ranged from 50.01%-75% of their total trials, 2= ranged from 75.01%-90% of their total trials, 3=ranged from 90.01%-100% of their total trials.
[l] The appropriate trials of targeted students: 0= did not exceed the 50% of their total trials, 1= ranged from 50.01%-75% of their total trials, 2= ranged from 75.01%-90% of their total trials, 3=ranged from 90.01%-100% of their total trials.

three main junctures of a lesson, all other teaching practices mainly occurred in one or two main junctures of the lesson. The same pattern was obvious for the frequency and focus dimensions, since these teaching practices were usually employed once or twice during a PE lesson, and were related to one or two of the three aspects that pertain to the dimension of focus. As far as the quality dimension is concerned, descriptive statistics suggested that teachers have employed these teaching practices in different ways, attending to different properties of each practice, as these are outlined in the notes of Table 4.1. Interestingly, when no answer was given to a posed question, teachers never moved to another question or answered the question themselves, but they posed easier questions or restated the questions with easier words. Similarly, teachers never ignored a given answer. Instead, they either indicated that the answer was correct or not, or invited students to comment on the given answer. Finally, differentiation was scarcely observed in any generic teaching practice, with the exception of differentiation of application, which occurred in about the one fourth of the lessons in different frequencies per lesson.

Turning to the content-specific teaching practices (see Table 4.2), demonstration and task explicitness (i.e., provision of fully explicit tasks) were frequently employed and adequately allocated to the three main phases of PE lessons. On the contrary, congruent and specific feedback did not occur in a large percentage (i.e., 41.8%) of observed lessons. When employed, however, it mainly occurred in one or two main phases of the PE lesson. As far as the quality dimension is concern, similarly to the generic teaching practices, all the practices involved in the *mTSS* low-inference form were employed in different ways and degrees by the teachers. A careful examination of the quality dimension of Table 4.2, leads to a couple of interesting observations. First, a significant percentage of teachers (i.e., about 25%) did not supervise or they just monitored student practice, without giving any feedback to students; and second, as far as the task explicitness is concerned, no implicit tasks were provided to students throughout the 147 observed lessons.

Variance Decomposition and the Contribution of Student- and Teacher-Background Variables to Student Psychomotor Learning

Tables 4.3 - 4.7 present the results of the multilevel analyses, each of which was run by using a different set of data captured with each of the five instruments employed in this study. However, before analyzing the results related to the individual and joint contribution of generic and content-specific teaching practices, some other important observations are highlighted within this section and further discussed at the next chapter. First, as can be observed from the second column of these tables (Model 0), although the greatest portion of variance is situated at the student level (about 85%), yet, a significant percentage also exists at the teacher level (about

15%), which implies that the teacher has an important role in helping students develop their psychomotor skills. Second, turning to Model 1, in which student-level factors were examined, one can notice that the variables entered in this model helped explain approximately 70% of the total variance of student achievement. As was expected, most of this explained variance was related to student level, as the variables entered in this model pertained to students' initial performance, students' gender, parents' education level and job social status. Third, as far as the contribution of gender is concerned, boys were found to perform significantly better than girls. Fourth, the only variable related to the socioeconomic status of students' family that was found to significantly contribute to student psychomotor learning pertained to the social status of father's job. Particularly, students whose father's job belonged to the upper-middle class category exhibited better performance than those whose father's job had a middle or working class status; similarly, students' whose father had a middle class status job performed better than those whose father's job had a working class status. Fifth, a percentage of 24% of the total variance remained unexplained at the student level, which implies that other variables related to student level could contribute to student psychomotor learning as well.

Sixth, and turning to Model 2, one such variable refers to student out of school participation in physical activities. As can be observed from the corresponding column (see Column 4-Model 2), students who participated in out of school physical activities, performed better than their classmates who did not do so. This variable explained approximately 1% of student achievement, thus leaving about 23% of the total variance unexplained at the student level. In addition, when this variable was entered in Model 2, the effect of the socioeconomic status variable (i.e., father's job social status) either disappeared or became statistically significant at level $\alpha=0.10$, something that might be due to multicollinearity (i.e., the presence of linear or near linear relationship among these two variables). All other variables entered in Model 2 (i.e., variables pertain to classroom composition effects and teacher characteristics) did not have any significant contribution to student psychomotor development, thus leaving a percentage of 6.13% and 23.18% unexplained at the teacher and student level, correspondingly. Finally, it is worth mentioning that in all the above cases, the deviance likelihood statistic between the previous and subsequent models (e.g., Model 0 to Model 1; Model 1 to Model 2) showed a significant change ($p<.001$), which justifies the selection of the subsequent model.

The Contribution of Generic Teaching Practices to Student Psychomotor Learning

At the next step of the analysis, generic teaching aspects were entered individually in Model 3, thus creating different versions of this model. Tables 4.3 and 4.4 present the results pertaining to the contribution of generic teaching aspects as captured either with the low- and high-inference form of *DMEE and mTSS* or the student ratings[19]. As one can observe from both tables, five out of the seven teaching factors of *DMEE* (i.e., *orientation, questioning, modeling, time management,* and *classroom as a learning environment*) were found to significantly contribute to student psychomotor learning, since the likelihood statistic (X^2) showed a significant change between Model 2 and the corresponding alternative Models 3 ($p< .001$). This means that the variables measuring these five teaching factors had a significant effect on student psychomotor learning. In addition, two teaching factors (i.e., *orientation,* and *questioning*) and one dimension (i.e., *quality* of *orientation* and *questioning*) were consistently found to contribute to student psychomotor development irrespective of the type of *DMEE* observation form

---

[19] As mentioned in the data analysis section, beyond the content-specific factors, the low- and high-inference forms of *mTSS,* also captured two generic teaching factors (i.e., *time* and *classroom management*). In addition, the student ratings were also capturing both types of practices.

Table 4.3

Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes--Generic Teaching Practices Captured With DMEE and mTSS Low-Inference Forms [a]

Factors	Model 0	Model 1[b]	Model 2[b]	Model 3a	Model 3b	Model 3c	Model 3d	Model 3e	Model 3f	Model 3g
Fixed Part (Intercept)	0.170(0.048)	0.075(0.038)	-0.034(0.048)	-0.030(0.048)	-0.030(0.048)	-0.027(0.048)	-0.027(0.048)	-0.028(0.048)	-0.028(0.049)	-0.026(0.048)
STUDENT LEVEL										
Context										
Initial performance		0.769(0.021)	0.757(0.021)	0.763(0.022)	0.763(0.022)	0.764(0.022)	0.765(0.022)	0.765(0.022)	0.764(0.022)	0.762(0.022)
Gender (girls=0, boys=1)		0.139(0.031)	0.143(0.031)	0.130(0.031)	0.130(0.031)	0.130(0.031)	0.129(0.031)	0.130(0.031)	0.130(0.031)	0.128(0.031)
Father's job social status (working class=0, middle class=1)		0.073(0.031)	0.053(0.031)+	0.063(0.031)	0.061(0.031)	0.061(0.031)	0.061(0.031)	0.061(0.031)	0.061(0.031)	0.061(0.031)
Father's job social status (working class=0, upper middle class=1)		0.147(0.080)+	NSS	NSS	NSS	NSS	NSS	NSS	NSS	NSS
Opportunity to learn										
Participation in out-of-school PE activities (no=0, yes=1)			0.138(0.038)	0.132(0.038)	0.133(0.038)	0.131(0.038)	0.132(0.038)	0.131(0.038)	0.131(0.038)	0.131(0.038)
CLASSROOM/TEACHER LEVEL										
Teaching Quality										
Questioning Stage (*DMEE*)				0.077(0.039)						
Questioning Frequency (*DMEE*)					0.053(0.024)					
Questioning Quality(reaction if no answer is given) (*DMEE*)						0.060(0.031)+				
Questioning Quality(type of feedback to students' answers) (*DMEE*)							0.053(0.027)			
Questioning Quality (reaction about the answer) (*DMEE*)								0.136(0.062)		
Modelling Stage (*DMEE*)									0.139(0.082)+	
Orientation Stage (*DMEE*)										0.198(0.068)
Variance components										
Teacher	14.57%	6.13%	6.13%	5.79%	5.79%	5.79%	5.79%	5.79%	5.96%	5.13%
Student	85.43%	24.01%	23.18%	22.85%	22.85%	22.85%	22.85%	22.85%	22.85%	22.85%
Absolute	0.60	0.18	0.18	0.17	0.17	0.17	0.17	0.17	0.17	0.17
Percentage Explained		69.87%	70.70%	71.36%	71.36%	71.36%	71.36%	71.36%	71.19%	72.02%
Significance test										
X^2	1999.97	756.89	709.71	671.95	671.22	672.14	672.01	671.14	673.02	668.09
Reduction		1243.07	47.18	37.76[c]	38.50[c]	37.57[c]	37.70[c]	38.57[c]	36.69[c]	41.63[c]
Degrees of freedom		4	1	1	1	1	1	1	1	1
p value		.001	.001	.001	.001	.001	.001	.001	.001	.001

Table 4.3 Continued

Factors	Model 3h	Model 3i	Model 3j	Model 3k	Model 3l	Model 3m
Fixed Part (Intercept)	-0.026(0.048)	-0.026(0.048)	-0.027(0.047)	-0.028(0.047)	-0.033(0.047)	-0.031(0.049)
STUDENT LEVEL						
Context						
Initial performance	0.763(0.022)	0.763(0.022)	0.762(0.021)	0.761(0.021)	0.756(0.022)	0.757(0.022)
Gender (girls=0, boys=1)	0.128(0.031)	0.127(0.031)	0.127(0.031)	0.130(0.031)	0.134(0.031)	0.134(0.031)
Father's job social status (working class=0, middle class=1)	0.062(0.031)	0.062(0.031)	0.063(0.031)	0.064(0.031)	0.062(0.031)	0.059(0.031)+
Father's job social status (working class=0, upper middle class=1)	NSS	NSS	NSS	NSS	NSS	NSS
Opportunity to learn						
Participation in out-of-school PE activities (no=0, yes=1)	0.131(0.038)	0.131(0.038)	0.131(0.038)	0.131(0.038)	0.135(0.038)	0.132(0.038)
CLASSROOM/TEACHER LEVEL						
Teaching Quality Factors						
Orientation Frequency (*DMEE*)	0.145(0.055)					
Orientation Focus (*DMEE*)		0.172(0.071)				
Orientation Quality (*DMEE*)			0.177(0.052)			
Time Management: Waiting Time (Recoded) (*mTSS*)				0.145(0.043)		
Time Management: Total Practice Time (*mTSS)*					0.109(0.033)	
Classroom Management: Management Time *(mTSS)*						0.055(0.030)+
Variance components						
Teacher	5.46%	5.46%	4.80%	4.80%	4.97%	5.79%
Student	22.85%	22.85%	22.85%	22.85%	22.85%	22.85%
Absolute	0.17	0.17	0.17	0.17	0.17	0.17
Percentage Explained	71.69%	71.69%	72.35%	72.35%	72.19%	71.36%
Significance test						
X^2	669.37	670.16	665.57	665.76	666.12	672.62
Reduction	40.34 [c]	39.55 [c]	44.14 [c]	43.96 [c]	43.59 [c]	37.09 [c]
Degrees of freedom	1	1	1	1	1	1
p value	.001	.001	.001	.001	.001	.001

Notes. [a] For space reasons the above table involves only teaching aspects found to significantly contribute to student psychomotor learning. All other generic teaching aspects' contribution involved in *DMEE and mTSS* low-inference forms was found to be NSS.

[b] Mother's job social status and father's and mother's education level (i.e., compulsory, upper-secondary, tertiary) were also entered in Model 1, but their contribution was NSS. In model 2, variables pertain to classroom composition effects (i.e., aggregated pre-test performance at the classroom level, percentage of girls in classroom, aggregated father's and mother's education level and job social status) and teacher characteristics (i.e., gender, PE coordinator/classroom teacher, teaching experience in PE) were also tested but found to be NSS.

[c] For each alternative Model 3 (i.e., Models 3a up to 3m) the reduction is estimated in relation to the deviance of Model 2.

[+] p. value < .10, all other effects significant at p< .05 unless otherwise stated (NSS=Not Statistically Significant).

Table 4.4

Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes--Generic Teaching Practices Captured With DMEE and mTSS High-Inference Forms and Student Ratings

Factors	Model 0	Model 1[a]	Model 2[a]	Model 3a	Model 3b	Model 3c	Model 3d	Model 3e	Model 3f
Fixed Part (Intercept)	0.170(0.048)	0.075(0.038)	-0.034(0.048)	-0.034(0.047)	-0.034(0.048)	-0.034(0.048)	-0.031(0.047)	-0.029(0.048)	-0.033(0.048)
STUDENT LEVEL									
Context									
Initial performance		0.769(0.021)	0.757(0.021)	0.758(0.021)	0.757(0.021)	0.757(0.021)	0.764(0.021)	0.762(0.022)	0.758(0.022)
Gender (girls=0, boys=1)		0.139(0.031)	0.143(0.031)	0.130(0.031)	0.143(0.031)	0.143(0.031)	0.128(0.031)	0.130(0.031)	0.133(0.031)
Father's job social status (working class=0, middle class=1)		0.073(0.031)	0.053(0.031)+	0.064(0.031)	0.053(0.031)+	0.053(0.031)+	0.062(0.031)	0.061(0.031)	0.062(0.031)
Father's job social status (working class=0, upper middle class=1)		0.147(0.080)+	NSS	NSS	NSS	NSS	NSS	NSS	NSS
Opportunity to learn									
Participation in out-of-school PE activities (no=0, yes=1)			0.138(0.038)	0.135(0.038)	0.138(0.038)	0.138(0.038)	0.135(0.038)	0.132(0.038)	0.136(0.038)
CLASSROOM/TEACHER LEVEL									
Teaching Quality-Factors									
F1: Orientation (*DMEE*)				0.137(0.039)					
F2: Structuring (*DMEE*)					NSS				
F3: Application (*DMEE*)						NSS			
F4: Questioning (*DMEE*)							0.104(0.031)		
F5: Teaching Modeling (*DMEE*)								NSS	
F6: Time Management (*DMEE*)									0.134(0.054)
Variance components									
Teacher	14.57%	6.13%	6.13%	4.80%	6.13%	6.13%	4.97%	6.13%	5.46%
Student	85.43%	24.01%	23.18%	22.85%	23.18%	23.18%	22.85%	23.18%	22.85%
Absolute	0.60	0.18	0.18	0.17	0.18	0.18	0.17	0.18	0.17
Percentage Explained		69.87%	70.70%	72.35%	70.70%	70.70%	72.19%	70.70%	71.69%
Significance test									
X^2	1999.97	756.89	709.71	665.01	-	-	665.48	-	670.08
Reduction		1243.07	47.18	44.70 [b]	-	-	44.23 [b]	-	39.63 [b]
Degrees of freedom		4	1	1	-	-	1	-	1
p value		.001	.001	.001	-	-	.001	-	.001

Table 4.4 Continued

Factors	**Model 3g**	**Model 3h**	**Model 3i**	**Model 3j**	**Model 3k**
Fixed Part (Intercept)	-0.031(0.049)	-0.033(0.049)	-0.033(0.048)	-0.028(0.048)	-0.040(0.047)
STUDENT LEVEL					
Context					
Initial performance	0.760(0.022)	0.753(0.022)	0.757(0.022)	0.760(0.022)	0.753(0.021)
Gender (girls=0, boys=1)	0.131(0.031)	0.132(0.032)	0.133(0.031)	0.131(0.031)	0.147(0.031)
Father's job social status (working class=0, middle class=1)	0.061(0.031)	0.069(0.032)	0.060(0.031)+	0.061(0.031)	0.054(0.031)+
Father's job social status (working class=0, upper middle class=1)	NSS	NSS	NSS	NSS	NSS
Opportunity to learn					
Participation in out-of-school PE activities (no=0, yes=1)	0.133(0.038)	0.129(0.039)	0.135(0.038)	0.132(0.038)	0.144(0.038)
CLASSROOM/TEACHER LEVEL					
Teaching Quality Factors					
F7: Classroom as a Lear. Envir.: T-S & S-S Interactions (*DMEE*)	0.077(0.045)+				
F8: Classroom as a Lear. Envir.: Classroom Disorder (*DMEE*)		0.110(0.038)			
F1: Classroom Management (*mTSS*)			0.223(0.091)		
F2: Time Management (*mTSS*)				0.120(0.062)	
Time and Classroom Management (*student ratings*)[c]					0.203(0.066)
Variance components					
Teacher	5.96%	5.30%	5.46%	5.79%	4.80%
Student	22.85%	22.85%	22.85%	22.85%	23.18%
Absolute	0.17	0.17	0.17	0.17	0.17
Percentage Explained	71.19%	71.85%	71.69%	71.36%	72.02%
Significance test					
X^2	672.99	651.26	670.14	672.14	701.19
Reduction	36.72 [b]	58.45 [b]	39.58 [b]	37.57 [b]	8.52 [b]
Degrees of freedom	1	1	1	1	1
p value	.001	.001	.001	.001	.01

Notes. [a] Mother's job social status and father's and mother's education level (i.e., compulsory, upper-secondary, tertiary) were also entered in Model 1, but their contribution was NSS. In model 2, variables pertain to classroom composition effects (i.e., aggregated pre-test performance at the classroom level, percentage of girls in classroom, aggregated father's and mother's education level and job social status) and teacher characteristics (i.e., gender, PE coordinator/classroom teacher, teaching experience in PE) were also tested but found to be NSS.

[b] For each alternative Model 3 (i.e., Models 3a up to 3k) the reduction is estimated in relation to the deviance of Model 2.

[c] Student data aggregated to teacher level.

[+] p. value < 0.10, all other effects significant at p.<05 unless otherwise stated (NSS=Not Statistically Significant).

employed (i.e., low- or high-inference)[20]. Despite the relatively low percentage (i.e., 6.13%) that remained unexplained at the teacher level after running Models 1 and 2, the abovementioned five generic teaching practices helped explain from 0.17% (i.e., *stage of modeling* and *teacher-student and student-student interactions*) to 1.33% (*quality of orientation, time management: waiting*, and *orientation*) of the unexplained teacher-level variance. Although these percentages are deemed small, they represent the 2.77% and 21.70%, respectively, of the variance that remained unexplained at the teacher level after introducing variables that pertain to student level (i.e., see Models 1 and 2)[21].

A careful examination of the contribution of each generic teaching practice leads to an important finding. Two out of the three generic teaching practices found to have the largest contribution to student psychomotor learning (i.e., *orientation*, *time management*, and *questioning*)[22] were among those teaching practices that have been identified as underexplored in the literature review section. In particular, orientating students toward learning goals (and specifically the *quality* dimension, see Table 4.3) was found to have the highest contribution among generic teaching practices captured with *DMEE* to student psychomotor learning[23]. Although challenging students to identify the reasons for which a certain activity or lesson occurred (this represents one of the three aspects of the *quality* dimension of *orientation* --see

---

[20] To examine the contribution of each dimension to student psychomotor learning and whether there was consistency among the variables of *DMEE* low- and high-inference forms that were found to significantly contribute in explaining additional variance, each individual statement of the *DMEE* high-inference form was entered in Model 3. However, Table 4.4 presents only the factor scores of the SEM models that were entered in alternative Models 3.

[21] These percentages were calculated by taking the ratio of the additional explained teacher-level variance when shifting from Model 2 to alternative Models, to the unexplained teacher-level variance in Model 2.

[22] The fact that these teaching generic practices were found to have the largest contribution to student psychomotor learning, does not necessarily imply that these practices are the most important generic practices for teaching PE. Rather, this might relate to the large variance that existed in the data of these generic practices.

[23] As can be observed from Tables 4.3 and 4.4, Models 3j and 3a, respectively, explain the most variance than any other alternative Model 3 that involves variables of the *DMEE* instrument.

note ‘m’ in Table 4.1) was not observed in any lesson, teachers who provided the importance for which a specific task or lesson was taking place in a typical manner, were found to be more effective than those who did not do so. In addition, students whose teachers related this importance to lesson’s goals, had even better outcomes. Moreover, three other dimensions of the *orientation* factor were found to statistically contribute to student psychomotor development. For instance, teachers who employed *orientation* at two or all three main phases of the lesson (*stage*), were found to be more effective than their counterparts who did not employ this teaching practice or had employed it only once. Similarly, as far as the *frequency* dimension is concerned, the study results suggested that students whose teachers employed *orientation* at least twice in a lesson, had larger learning gains than those whose teachers did not employ it at all or they employed it only once. Furthermore, teachers who were relating *orientation* to at least two of three *focus* aspects (i.e., a certain aim of the lesson, the entire lesson, and the unit/a series of lessons), were found to be more effective than those who were doing it for just one aspect or were not doing it at all. Interestingly, despite its importance, *orientation* was only observed in 46 (i.e., about 31%) out of 147 classroom observations.

The other generic teaching practice that was found to have a significant impact on student psychomotor learning and whose impact has not been adequately validated through research studies in PE refers to *questioning*. Results indicated that three of its dimensions (i.e., *stage, frequency, and quality*) had a similar effect on student psychomotor learning. Specifically, teachers who were asking questions during all three main junctures (i.e., introduction, core, end) of the lesson (*stage*) helped their student develop their psychomotor skills in a greater extent than those who did it in fewer phases of the lesson. Likewise, teachers who asked at least four[24]

[24] The number of questions (i.e., four) is indicative and implies that these teachers were asking more questions than the others.

questions (*frequency*) during the lesson, restated a question using easier words when no answer was given (*quality*), positively commented on correct student answers and constructively dealt with incorrect student answers (*quality*), as well as indicated that the answer was correct/partly correct/incorrect (*quality*), helped their students develop their psychomotor skills to a greater degree than teachers who did fewer than four questions, negatively commented on incorrect student answers or did not comment at all and/or ignored students' answers. Moreover, as was the case with *orientation*, although *questioning* is considered an important teaching skill, it had not been observed in 43.5% of total lessons; whereas 38% of lessons had one to three questions and only 18.5% had at least four questions.

The third generic teaching factor (second in magnitude) which was found to have significant effects upon student achievement gains refers to *time management*. The importance of this factor and its constituent aspects was consistently corroborated by the data captured from all four instruments/forms employed to measure this factor (i.e., *DMEE* high-inference, *mTSS* low- and high-inference, and student ratings). Particularly, *minimizing waiting time* and *maximizing practice time* related to the emphasized skill were among the generic teaching aspects that were found to contribute most to student psychomotor learning (see Models 3k and 3l in Table 4.3). This finding seems to resonate with data captured with student survey, as the conglomerate factor of *time and classroom management* (i.e., minimizing waiting and transition time, starting lesson on time) was found to contribute the most to student psychomotor learning, among factors that were rated by students (see Model 3k in Table 4.4). Likewise, *time management* factor, as captured with the high-inference forms of *DMEE* (e.g., maximizing on-task behavior) and *mTSS* (e.g., organizing small-sided games to maximize participation and practice) was found to have a significant effect on student learning outcomes.

Modeling and *classroom as a learning environment/classroom management* were the other two generic teaching practices found to significantly contribute to student psychomotor learning. Specifically, students of teachers who employed *modeling* during one or two different phases of the lesson (*stage*) were found to perform better than their counterparts whose teachers did not employ this teaching practice. However, the other dimensions of *modeling* as well as the corresponding high-inference factor were not found to significantly contribute to student achievement, something that is probably associated with the fact that this teaching practice was not used in about 70% of observed lessons. Interestingly, as far as the *quality* of this teaching practice is concerned, approximately 12.2% of the observed lessons mainly involved guided discovery and only about 1.4% included discovery (see Table 4.1). Turning to the two sub-factors of the *classroom as a learning environment* factor (i.e., *teacher-student*/*student-student interactions* and *classroom disorder*), both of them were found to have a statistically significant impact on student psychomotor learning. Specifically, students who interacted with their teacher and classmates in a way that contributed to achieving the lesson's goals had greater learning gains than students who did not experience such interactions. As far as the *classroom disorder/management* is concerned, the results yielded from data captured by different instruments (i.e., *DMEE* high-inference, *mTSS* low- and high-inference, and student ratings) converged in showing that this teaching factor and its aspects significantly contribute to student psychomotor learning. For instance, the analysis of the data captured with the *DMEE* high-inference form showed that teachers who effectively dealt with students' misbehavior were more effective than those who were not successful in doing so; whereas the analysis of the data captured with *mTSS* high-inference indicated that students whose teachers used effective management or transitional routines (e.g., distributing equipment, assigning student to pairs/groups, dealing with disciplinary problems) achieved greater learning gains than their counterparts whose teachers did not use such routines. In addition, the analysis of the data captured with *mTSS* low-inference form pointed to statistically significant differences in favor of students whose teachers minimized the time devoted to issues that were not relevant to the instructional activities (e.g., announcements irrelevant to lesson's goals, roll taking).

Unlike the abovementioned generic teaching factors, which were found to have a statistically significant effect on student psychomotor learning, there were two generic teaching factors which did not have any significant impact on student achievement[25]. Namely, *application* and *structuring* were not found to contribute to student psychomotor development. Interestingly, this outcome was consistent among the data captured with both forms of the *DMEE*. Scrutinizing the data captured with the low-inference form, descriptive statistics showed that both teaching factors were employed by the teachers in different degrees and shades (see Table 4.1). In particular, *structuring* comments were used in about 76% of the lessons observed, while *application* activities ranged from 1 to 12 per lesson. Nevertheless, this finding does not imply that these teaching factors are not important for teaching psychomotor skills in PE. Instead, this outcome might be related to observers' unfamiliarity with PE teaching and their insufficient content-knowledge, which, in turn, led to difficulties in correctly coding these two teaching practices; an issue that is discussed in more detail in the next chapter.

Before shifting to discussing the contribution of the content-specific teaching practices to student psychomotor learning, a couple of observations about the five dimensions of the *DMEE* are in order. First, each of the four out of the five dimensions (except *differentiation*) was found to have a significant impact on student psychomotor learning through at least one teaching factor.

[25] In addition, variables that captured with the student survey and pertained to the extent to which students were given the opportunity to engage in free play or PE lessons were canceled because other activities had to take place were not found to have any significant impact on student psychomotor learning.

For instance, the *stage* dimension of *orientation, questioning,* and *modeling*, the *frequency* dimension of *questioning, orientation, time management*[26], and *classroom disorder*, the *focus* dimension of *orientation* and *application*, and the *quality* dimension of *orientation, questioning, structuring, modeling, teacher-student and student-student interactions,* and *classroom disorder* were found to significantly contribute to student psychomotor learning. This outcome implies that beyond the *frequency* and *quality* dimensions which are usually captured by the observation instruments in general, other dimensions are equally important for measuring instruction in PE, something that is further discussed in the next chapter.

Second, the dimension of *differentiation* was not found to have significant effect on student psychomotor development. However, this result should not be dissociated from the fact that *differentiation* was rarely employed by the teachers in study's sample. Particularly, *differentiation* of *structuring, modeling,* and *orientation* were not observed in any lesson and *differentiation* of *questioning* was captured only in four lessons (out of 147, 2.72%). Therefore, the dimension of *differentiation* should not be deemed as unimportant for student psychomotor learning. Rather, the fact that this teaching dimension was infrequently employed by the teachers of the study implies that teachers might not employ *differentiation* because they find it difficult to do so, a fact that deserves further scrutiny and one that is considered in the next chapter.

The Contribution of Content-Specific Teaching Practices to Student Psychomotor Learning

To examine the individual contribution of content-specific teaching practices, alternative Models 3 were also created involving content-specific teaching aspects. Tables 4.5 and 4.6 present the results of analyzing the data captured with the low- and high-inference forms of the

[26] See footnote 20 for the process followed to examine the contribution of each dimension of the *DMEE* high-inference factors to student psychomotor learning.

mTSS instrument, correspondingly, and student ratings. Collectively, four out of the five content-specific teaching factors (i.e., *task progression, demonstration, congruent and specific feedback, quality of student practice*) were found to have a significant effect on student psychomotor learning in PE, since the likelihood statistic (X^2) showed a significant change between Model 2 and the corresponding alternative Models 3 (p< .001).

Checking for consistency among the two observation forms of *mTSS*, one can notice that two teaching factors (i.e., *task progression*, and *demonstration*) and all three dimensions[27] (i.e., *stage, frequency*, and *quality* of *demonstration* and *congruent and specific feedback*) were consistently found to contribute to student psychomotor development. Notably, the abovementioned four content-specific teaching practices explained an almost identical range of the unexplained teacher-level variance (i.e., from 0.17%-- *stage* dimension of the *skill demonstration* to 1.00%--*demonstration and congruent feedback* factor, see Model 3d in Table 4.6) to that explained by generic teaching practices. However, once again, these percentages should be considered with respect to the relatively low percentage (i.e., 6.13%) that remained unexplained before entering any variable pertaining to teaching practices. Following this approach, these percentages represent the 2.77% and 16.31%, correspondingly, of the remaining unexplained variance at the teacher level of Model 2[28].

Demonstrating how to perform a psychomotor skill and emphasizing the key learning cues before, during, and immediately after student practice (i.e., *congruent and specific feedback)*

[27] Although the *mTSS* low-inference asks from the observer to code only the dimensions of *frequency and quality* of the observed teaching factors, for the practices of *demonstration, congruent and specific feedback*, and *task explicitness* the dimension of *stage* was also calculated based on the data captured with this observation form and following the guidelines of the *DMEE* low-inference instrument. In addition, to examine whether there was consistency among *mTSS* low- and high-inference forms in terms of the statistical significance of each dimension, a similar process to that followed with the *DMEE* high-inference form was followed, by testing the statistical significance of each individual statement of the *mTSS* high-inference form in Model 3.

[28] See footnote 21 on how these percentages were calculated.

Table 4.5

Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes—Content-Specific Teaching Practices Captured With mTSS Low-Inference Form [a]

Factors	Model 0	Model 1[b]	Model 2[b]	Model 3a	Model 3b	Model 3c
Fixed Part (Intercept)	0.170(0.048)	0.075(0.038)	-0.034(0.048)	-0.028(0.048)	-0.029(0.048)	-0.028(0.048)
STUDENT LEVEL						
Context						
Initial performance		0.769(0.021)	0.757(0.021)	0.760(0.022)	0.761(0.022)	0.761(0.022)
Gender (girls=0, boys=1)		0.139(0.031)	0.143(0.031)	0.130(0.031)	0.131(0.031)	0.134(0.031)
Father's job social status (working class=0, middle class=1)		0.073(0.031)	0.053(0.031)+	0.060(0.031)+	0.061(0.031)	0.060(0.031)+
Father's job social status (working class=0, upper middle class=1)		0.147(0.080)[+]	NSS	NSS	NSS	NSS
Opportunity to learn						
Participation in out-of-school PE activities (no=0, yes=1)			0.138(0.038)	0.132(0.038)	0.133(0.038)	0.131(0.038)
CLASSROOM/TEACHER LEVEL						
Teaching Quality Factors						
Task Progression: Diversity of Tasks (*Informing, Refining, Extending, Applying*)				0.146(0.051)		
Task Progression: Time Practicing Extending Tasks					0.070(0.031)	
Task Progression: Time Practicing Refining Tasks						0.289(0.139)
Variance components						
Teacher	14.57%	6.13%	6.13%	5.30%	5.63%	5.79%
Student	85.43%	24.01%	23.18%	22.85%	22.85%	22.85%
Absolute	0.60	0.18	0.18	0.17	0.17	0.17
Percentage Explained		69.87%	70.70%	71.85%	71.52%	71.36%
Significance test						
X^2	1999.97	756.89	709.71	668.25	670.88	671.63
Reduction		1243.07	47.18	41.46 [c]	38.83 [c]	38.08 [c]
Degrees of freedom		4	1	1	1	1
p value		.001	.001	.001	.001	.001

Table 4.5 Continued

Factors	Model 3d	Model 3e	Model 3f	Model 3g	Model 3h
Fixed Part (Intercept)	-0.026(0.049)	-0.027(0.048)	-0.030(0.048)	-0.030(0.048)	-0.030(0.049)
STUDENT LEVEL					
Context					
Initial performance	0.762(0.022)	0.762(0.022)	0.758(0.022)	0.759(0.022)	0.759(0.022)
Gender (girls=0, boys=1)	0.129(0.031)	0.130(0.031)	0.132(0.031)	0.132(0.031)	0.131(0.031)
Father's job social status (working class=0, middle class=1)	0.061(0.031)	0.061(0.031)	0.059(0.031)+	0.059(0.031)+	0.060(0.031)+
Father's job social status (working class=0, upper middle class=1)	NSS	NSS	NSS	NSS	NSS
Opportunity to learn					
Participation in out-of-school PE activities (no=0, yes=1)	0.131(0.038)	0.131(0.038)	0.134(0.038)	0.134(0.038)	0.133(0.038)
CLASSROOM/TEACHER LEVEL					
Teaching Quality Factors					
Skill Demonstration: Stage	0.084(0.048)				
Skill Demonstration: Frequency		0.065(0.029)			
Specific and Congruent Feedback: Stage			0.093(0.038)		
Specific and Congruent Feedback: Frequency				0.061(0.022)	
Specific and Congruent Feedback: Quality					0.049(0.027)+
Variance components					
Teacher	5.96%	5.63%	5.46%	5.30%	5.79%
Student	22.85%	22.85%	22.85%	22.85%	22.85%
Absolute	0.17	0.17	0.17	0.17	0.17
Percentage Explained	71.19%	71.52%	71.69%	71.85%	71.36%
Significance test					
X^2	672.78	670.83	670.14	668.80	672.63
Reduction	36.93 [c]	38.88 [c]	39.57 [c]	40.91 [c]	37.09 [c]
Degrees of freedom	1	1	1	1	1
p value	.001	.001	.001	.001	.001

Notes. [a] For space reasons the above table involves only teaching aspects found to significantly contribute to student psychomotor learning. All other teaching aspects' contribution involved in *mTSS* low-inference was found to be NSS.
[b] Mother's job social status and father's and mother's education level (i.e., compulsory, upper-secondary, tertiary) were also entered in Model 1, but their contribution was NSS. In model 2, variables pertain to classroom composition effects (i.e., aggregated pre-test performance at the classroom level, percentage of girls in classroom, aggregated father's and mother's education level and job social status) and teacher characteristics (i.e., gender, PE coordinator/classroom teacher, teaching experience in PE) were also tested but found to be NSS.
[c] For each alternative Model 3 (i.e., Models 3a up to 3h) the reduction is estimated in relation to the deviance of Model 2.
[+] p. value < 0.10, all other effects significant at p.<05 unless otherwise stated (NSS=Not Statistically Significant).

Table 4.6

Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes—Content-Specific Teaching Practices Captured With mTSS High-Inference Form and Student Ratings

Factors	Model 0	Model 1[a]	Model 2[a]	Model 3a	Model 3b	Model 3c	Model 3d	Model 3e
Fixed Part (Intercept)	0.170(0.048)	0.075(0.038)	-0.034(0.048)	-0.030(0.048)	-0.030(0.048)	-0.034(0.048)	-0.034(0.047)	-0.032(0.048)
STUDENT LEVEL								
Context								
Initial performance		0.769(0.021)	0.757(0.021)	0.760(0.022)	0.758(0.022)	0.757(0.021)	0.759(0.021)	0.758(0.021)
Gender (girls=0, boys=1)		0.139(0.031)	0.143(0.031)	0.131(0.031)	0.131(0.031)	0.143(0.031)	0.141(0.031)	0.142(0.031)
Father's job social status (working class=0, middle class=1)		0.073(0.031)	0.053(0.031)+	0.060(0.031)+	0.060(0.031)+	0.053(0.031)+	0.054(0.031)+	0.053(0.031)+
Father's job social status (working class=0, upper middle class=1)		0.147(0.080)+	NSS	NSS	NSS	NSS	NSS	NSS
Opportunity to learn								
Participation in out-of-school PE activities (no=0, yes=1)			0.138(0.038)	0.135(0.038)	0.133(0.038)	0.138(0.038)	0.141 (0.038)	0.140(0.038)
CLASSROOM/TEACHER LEVEL								
Teaching Quality Factors								
F3: Task Progression (*mTSS*)				0.155(0.065)				
F4: Demonstration of desired movement skills (*mTSS*)					0.126(0.054)			
F5: Quality of Student Practice (*mTSS*)						NSS		
Content-Specific Teaching Practices (i.e., demonstration and congruent feedback). (*student ratings*)[c]							0.145(0.055)	
Quality of Student Practice (*student ratings*)[c]								0.171(0.083)
Variance components								
Teacher	14.57%	6.13%	6.13%	5.63%	5.63%	6.13%	5.13%	5.63%
Student	85.43%	24.01%	23.18%	22.85%	22.85%	23.18%	23.18%	23.18%
Absolute	0.60	0.18	0.18	0.17	0.17	0.18	0.17	0.17
Percentage Explained		69.87%	70.70%	71.52%	71.52%	70.70%	71.69%	71.19%
Significance test								
X^2	1999.97	756.89	709.71	670.35	670.67	-	703.22	705.61
Reduction		1243.07	47.18	39.36 [b]	39.05 [b]	-	6.49 [b]	4.10 [b]
Degrees of freedom		4	1	1	1	-	1	1
p value		.001	.001	.001	.001	-	.05	.05

Notes. [a] Mother's job social status and father's and mother's education level (i.e., compulsory, upper-secondary, tertiary) were also entered in Model 1, but their contribution was NSS. In model 2, variables pertain to classroom composition effects (i.e., aggregated pre-test performance at the classroom level, percentage of girls in classroom, aggregated father's and mother's education level and job social status) and teacher characteristics (i.e., gender, PE coordinator/classroom teacher, teaching experience in PE) were also tested but found to be NSS.

[b] For each alternative Model 3 (i.e., Models 3a up to 3e) the reduction is estimated in relation to the deviance of Model 2.

[c] Student data aggregated to teacher level.

[+] p. value < 0.10, all other effects significant at p.<05 unless otherwise stated (NSS=Not Statistically Significant).

were the two content-specific teaching practices found to explain the greatest percentage of variance in student learning[29]. Specifically, this finding resulted from the data captured with student ratings, and pertained to the content-specific factor that involved four statements capturing different aspects of *demonstration* (i.e., the way the skill was demonstrated--verbally or practically--and the emphasis paid on the learning cues) and one statement related to *congruent and specific feedback* (i.e., providing this type of feedback to individual students during practice). The importance of these two teaching practices was also corroborated by the data captured with the two observation forms of the *mTSS*. As far as the high-inference form is concerned, the corresponding factor, which involved three statements related to different aspects of *demonstration* (e.g., the way the skill was demonstrated--verbally or practically) and two statements pertaining to *congruent and specific feedback* (e.g., whether the teacher was providing individual and whole-class feedback) was found to endorse the abovementioned outcome.

Turning to the low-inference form, and focusing first at *demonstration*, results supported that teachers who used this teaching practice in at least two different phases of their lessons (i.e., beginning, core, end) (*stage*) had students who developed their psychomotor skills to a greater extent than those who did not employ *demonstration* or used it just once during the lesson. In addition, students of teachers who employed this teaching practice at least four times[30] (*frequency*) during a lesson had greater achievement gains than their counterparts whose teacher did so fewer times per lesson. Interestingly, about 36.4% of the observed lessons employed

[29] As was the case with generic teaching practices, the fact that these content-specific practices were found to have the largest contribution to student psychomotor learning, does not necessarily imply that these practices are the most important content-specific practices for teaching PE. Rather, this might relate to the large variance that existed in the data of these content-specific practices.

[30] As was the case with the number of questions, the times that a skill was demonstrated (i.e., four) is indicative and implies that these teachers were demonstrating the emphasized skill more times during a lesson than the others.

demonstration three or more times per lesson, while about 44% used it one or two times, and about 20% did not used this teaching practice at all. On the contrary, the *quality* dimension of this teaching practice was not found to significantly contribute to student achievement. However, this result does not necessarily imply that the way teacher demonstrates a psychomotor skill is not important for students. Rather, this outcome might be due to the different types of demonstration (i.e., whether a skill is demonstrated *verbally* and/or *practically* or if the demonstration is *partial* or *full*), and the inherent difficulty of classifying these categories from most to least important, something that is discussed with examples in the next chapter.

As far as the *congruent and specific feedback* is concerned, all three dimensions (*stage, frequency,* and *quality)* were found to significantly contribute to student psychomotor achievement. Specifically, teachers who reinforced their students' practice and they were providing at least three times (about 30% of observed lessons) (*frequency)* and throughout the three main phases of the PE lesson (*stage)* congruent and specific feedback (i.e., *quality*), had a larger impact on student learning than those who did so less than three times or at most two out of three main phases of a PE lesson. Likewise, teachers who were providing this type of feedback in two out of the three main phases of a lesson had a greater effect on student learning than those who did not employ this teaching practice (i.e., about 42% of observed lessons) or employed it just once during a lesson (i.e., 15%).

Another content-specific teaching practice which had a significant effect on student learning refers to the way in which teachers developed and presented the lesson content in terms of the focus of the motor task (i.e., *task progression*). Both the analyses of the *mTSS* low- and high-inference data revealed that this teaching factor significantly contributed to student psychomotor development. In particular, teachers who used three (about 36.5% of observed lessons) or all four (about 14%) types of instructional tasks (i.e., *informing, refining, extending,* and *applying)* during a single lesson, had better student outcomes than those who used just one (about 9.5%) or two (about 40%) types.

In addition, students who were engaged in refining tasks (about 21% of the observed lessons) exhibited better performance than those who did not. Interestingly, student practice in refining tasks ranged from few seconds (in cases where refining tasks were performed in waiting lines) to about 15 minutes out of 40 minutes that a typical PE lesson lasts. Similarly, student practice in extending tasks (observed in about 85% of total lessons) was found to have a significant effect on student psychomotor development. Specifically, teachers who offered practice opportunities in extending tasks that lasted on average between two to eight minutes[31] (out of 40 minutes) helped their student develop their psychomotor skills to a greater extent than those who did not provide this type of instructional task at all. However, no differences were found among teachers who offered on average extending tasks that lasted more than eight minutes and those who did not offered this type of task, a result probably associated with the fact that only a couple of teachers offered extending tasks lasting more than eight minutes in all three observed lessons. As far as the other two types of instructional task are concerned, descriptive statistics showed that applying tasks were offered in about half of the observed lessons and their duration ranged from few seconds (in cases where the applying task was the last task of the lesson and occurred at the very end of the lesson) to about 14 minutes, while informing tasks (observed in about 96% of lessons) ranged from few seconds (in cases where informing tasks were performed in waiting lines) to about 18.5 minutes.

[31] See footnote 'h' in Table 4.2 on how the classification of student practice time in extending task was made.

The fourth content-specific teaching factor which found to have significant effects upon student learning gains refers to the *quality of student practice*. This finding, however, was only supported by the data captured with student ratings. In particular, the corresponding teaching factor, which involved two statements that examined the degree to which student practice was appropriate and congruent to teacher's instruction, had a statistically significant effect on student psychomotor development. Nevertheless, this result was not corroborated by the data captured with the two content-specific observation forms, a result that is discussed in the next chapter.

The only content-specific teaching practice which was not found to have an effect on student psychomotor learning was *task explicitness*[32]. As mentioned in the methodology section, this teaching practice was only captured by the *mTSS* low-inference instrument and the one item involved in student survey. None of the three dimensions of this teaching practice (i.e., *stage*, *frequency*, and *quality*) had a significant effect on student achievement, a result that is probably related to the fact that the vast majority of teachers described the instructional tasks in a partly explicit manner (see Table 4.2); namely, by giving information about the task and the conditions under which students would practice the task, but not mentioning the criterion under which the success of a task or its completion could be judged. Therefore, this result does not necessarily imply that *task explicitness* is not an important teaching practice when teaching PE; yet, it might indicate that typically teacher directions are communicated to students in a similar manner, which does not include the criterion by which successful performance can be judged. Nevertheless, what might lead teachers to give partly explicit tasks is discussed in the next chapter.

[32] In addition, the two content-specific variables captured with the student survey and pertained to the extent to which students were engaged in warm-up activity that was relevant or irrelevant to lesson's aim(s) were not found to have any significant impact on student learning.

The Joint Contribution of Generic and Content-Specific Teaching Practices to Student Psychomotor Learning

The main hypothesis of this study was that the *joint* contribution of generic and content-specific teaching practices would help explain a greater percentage of the unexplained teacher-level variance than that explained when considering the *individual* contribution of either type of practices. This hypothesis was examined by employing three different approaches. First, by entering into a single multi-level model variables that pertained to both generic and content-specific teaching practices (see Table 4.7, Model 3h) and comparing the explained variance to that explained when individual variables/factors were entered into a model. Second, by comparing the percentage that got explained by the composite score of the *mTSS* low-inference instrument (see Table 4.7, Model 3i), which involved both generic and content-specific teaching practices to the explained percentages of individual variables. Third, by contrasting the contribution of factors that combined teaching aspects which pertained to both types of teaching practices (see Table 4.7, Model 3j) to the contribution of factors that involved a single teaching practice.

However, before examining the *joint* contribution of the two types of teaching practices, the cumulative contribution of each type of practices was explored. In other words, it was examined whether the combination of either generic or content-specific teaching practices could explain more variance than that explained when considering only the contribution of individual practices. Again, this was explored by following the three aforementioned approaches (when this was applicable): first, by entering into a single multi-level model more than one variables that pertained to generic (see Table 4.7, Model 3a) or content-specific (see Table 4.7, Model 3f) teaching practices and comparing the explained variance to the variance explained when only a

Table 4.7

Parameter Estimates and (Standard Errors) for the Analysis of Student Psychomotor Learning Outcomes--Combination of Generic and Content-Specific Teaching Practices Captured With Classroom Observation Forms and Student Ratings

Factors	Model 0	Model 1[a]	Model 2[a]	Model 3a	Model 3b	Model 3c	Model 3d	Model 3e
Fixed Part (Intercept)	0.170(0.048)	0.075(0.038)	-0.034(0.048)	-0.035(0.044)	-0.028(0.048)	-0.037(0.049)	-0.031(0.048)	-0.034(0.047)
STUDENT LEVEL								
Context								
Initial performance		0.769(0.021)	0.757(0.021)	0.752(0.021)	0.763(0.022)	0.752(0.022)	0.757(0.022)	0.762(0.021)
Gender (girls=0, boys=1)		0.139(0.031)	0.143(0.031)	0.132(0.031)	0.129(0.031)	0.131(0.032)	0.132(0.031)	0.140(0.031)
Father's job social status (working class=0, middle class=1)		0.073(0.031)	0.053(0.031)+	0.062(0.031)	0.061(0.031)	0.071(0.032)	0.061(0.031)	0.055(0.031)+
Father's job social status (working class=0, upper middle class=1)		0.147(0.080)+	NSS	NSS	NSS	NSS	NSS	NSS
Opportunity to learn								
Participation in out-of-school PE activities (no=0, yes=1)			0.138(0.038)	0.134(0.038)	0.132(0.038)	0.132(0.039)	0.133(0.038)	0.141(0.038)
CLASSROOM/TEACHER LEVEL								
Teaching Quality Factors_Combination of Generic Items								
Time Management: Waiting Time (Recoded)[c]				0.086(0.042)				
Classroom as a Learning Env.: Classroom Disorder [c]				0.079(0.024)				
Questioning (quality) [d]				0.119(0.049)				
Orientation (quality) [d]				0.140(0.050)				
DMEE (low-inference) Rasch Composite Score [f]					0.102(0.046)			
F9: Teaching Quality(*DMEE* High-Inference Second-Order Factor)						0.193(0.059)		
F6: Generic Instructional Aspects (F1,F2) (*mTSS* High-Inference)							0.221(0.085)	
Generic Teaching Practices (i.e., orientation, structuring, questioning) [e]								0.156(0.058)
Variance components								
Teacher	14.57%	6.13%	6.13%	2.81%	5.79%	5.13%	5.46%	5.13%
Student	85.43%	24.01%	23.18%	23.01%	22.85%	22.85%	22.85%	23.18%
Absolute	0.60	0.18	0.18	0.16	0.17	0.16	0.17	0.17
Percentage Explained		69.87%	70.70%	74.17%	71.36%	72.02%	71.69%	71.69%
Significance test								
X^2	1999.97	756.89	709.71	648.21	660.42	649.17	669.52	702.98
Reduction		1243.07	47.18 [b]	61.51 [b]	49.30 [b]	60.55 [b]	40.19 [b]	6.73 [b]
Degrees of freedom		4	1	4	1	1	1	1
p value		.001	.001	.001	.001	.001	.001	.01

Table 4.7 Continued

Factors	Model 3f	Model 3g	Model 3h	Model 3i	Model 3j
Fixed Part (Intercept)	-0.038(0.046)	-0.031(0.048)	-0.040(0.044)	-0.036(0.046)	-0.036(0.047)
STUDENT LEVEL					
Context					
Initial performance	0.761(0.021)	0.758(0.022)	0.752(0.021)	0.751(0.022)	0.759(0.021)
Gender (girls=0, boys=1)	0.129(0.031)	0.131(0.031)	0.131(0.031)	0.134(0.031)	0.142(0.031)
Father's job social status (working class=0, middle class=1)	0.061(0.031)	0.060(0.031)+	0.062(0.031)	0.060(0.031)+	0.054(0.031)+
Father's job social status (working class=0, upper middle class=1)	NSS	NSS	NSS	NSS	NSS
Opportunity to learn					
Participation in out-of-school PE activities (no=0, yes=1)	0.134(0.038)	0.135(0.038)	0.135(0.038)	0.136(0.038)	0.142(0.038)
CLASSROOM/TEACHER LEVEL					
Teaching Quality Factors_Combination of Content-Specific Items					
Task Progression: Diversity of Tasks [c]	0.104 (0.049)				
Demonstration [e]	0.181(0.061)				
F7: Content-Specific Instructional Aspects (F3, F4, F5) (*mTSS* High Inference)		0.175(0.065)			
Teaching Quality Factors_Combination of Generic and Content-Specific Items					
Time Management: Waiting Time (Recoded) [c]			0.081(0.041)		
Classroom as a Learning Environment: Classroom Disorder [c]			0.072(0.023)		
Questioning (quality) [d]			0.092(0.050)+		
Orientation (quality) [d]			0.112(0.051)		
Demonstration [e]			0.099(0.057)+		
mTSS (low-inference) Rasch Composite Score [f]				0.311(0.075)	
Teaching Quality (Second-Order Factor) [e]					0.201(0.064)
Variance components					
Teacher	4.14%	5.30%	2.48%	4.30%	4.80%
Student	22.85%	22.85%	23.01%	22.85%	23.18%
Absolute	0.16	0.17	0.15	0.16	0.17
Percentage Explained	73.01%	71.85%	74.50%	72.85%	72.02%
Significance test					
X^2	660.24	669.06	645.26	661.13	700.84
Reduction	49.47 [b]	40.65 [b]	64.45 [b]	48.58 [b]	8.88 [b]
Degrees of freedom	2	1	5	1	1
p value	.001	.001	.001	.001	.01

Notes. [a] Mother's job social status and father's and mother's education level (i.e., compulsory, upper-secondary, tertiary) were also entered in Model 1, but their contribution was NSS. In model 2, variables pertain to classroom composition effects (i.e., aggregated pre-test performance at the classroom level, percentage of girls in classroom, aggregated father's and mother's education level and job social status) and teacher characteristics (i.e., gender, PE coordinator/classroom teacher, teaching experience in PE) were also tested but found to be NSS.

[b] For each alternative Model 3 (i.e., Models 3a up to 3j), the reduction is estimated in relation to the deviance of Model 2.

[c] data captured with *mTSS* instrument, [d] data captured with *DMEE* instrument, [e] data captured with student ratings.

[f] A composite score was calculated for each teacher by taking each teacher's average lesson estimate score that emerged from the corresponding *DMEE or mTSS* low-inference observation forms across the three lessons.

[+] p. value < 0.10, all other effects significant at p.<05 unless otherwise stated (NSS=Not Statistically Significant).

single variable was entered in the model; second, by comparing the variance explained from the composite score of the low-inference form (see Table 4.7, Model 3b) to that explained from individual variables; and third, through contrasting the explained variance of factors that combined generic (see Table 4.7, Models 3c, 3d, 3e) or content-specific (see Table 4.7, Model 3g) teaching aspects to those that only involved a single practice. To start, as can be observed from Table 4.7, and specifically Model 3a, the combination of different generic teaching practices helped explain an even greater percentage of the unexplained variance at the teacher level than that explained by only examining the contribution of a single generic factor. Notably, the four out of five generic teaching practices that found to individually contribute to student psychomotor learning, were combined and explained 3.32% of the teacher level variance, a percentage that represents the 54.16% of the remaining unexplained variance at the teacher level of Model 2. Similarly, turning to the Model 3f, two out of four content-specific teaching practices were combined and explained 1.99% of the unexplained teacher variance (or 32.46% of the remaining unexplained variance at the teacher level of Model 2); a percentage that represents the largest explained variance than any other resulting from considering the contribution of a single content-specific teaching factor. The above results were partly corroborated by the second approach, as the content-specific second-order factor of the *mTSS* high-inference form explained a greater variance than the variance explained by single content-specific teaching factors; however, this was not the case with the generic second-order factor of this form.

Taking into consideration the above findings, the obvious result to be drawn is that the examination of the contribution of only one teaching aspect each time, might lead to a fragmented and piecemeal picture of what constitutes effective teaching in PE, as none teaching practice can be employed in isolation from other teaching practices. For example, when entering all the generic teaching factors found to significantly contribute to student achievement into Model 3a, the teaching factor of *modeling* was found to have a not statistically significant effect; this result was similar to that found with the content-specific factors of *congruent and specific feedback* and *student quality practice* (see Model 3f). Both results are probably related to multicollinearity; namely, the presence of linear or near linear relationship among factors, and which converts some factors from statistically significant to not statistically significant (Shavelson, 1996).

Turning to the examination of the *joint* contribution of generic and content-specific teaching practices, results were in line with the study's hypothesis that the combination of the two types of practices would help explain a greater proportion of variance in student learning. As can be observed in Table 4.7, when both types of practices were entered into the model (i.e., Model 3h), the explained variance increased. In particular, an additional percentage of 3.65% of the unexplained variance was explained, a percentage that is larger than that explained when combining only generic (i.e., 3.32%, Model 3a) or content-specific (i.e., 1.99%, Model 3f) teaching practices, respectively. Admittedly, the above percentage is small, but, having in mind that the available teacher-level variance to be explained after running Model 2 was 6.13%, then this percentage represents the 59.54% of the unexplained variance at the teacher level, as opposed to the 54.16% and 32.46% that was explained when only generic and content-specific teaching practices were respectively examined. Similarly, the second approach supported study's assumption, as the percentage explained when entering the *mTSS* low-inference form composite score, which involved both generic and content-specific teaching practices, into the model 3i (see Table 4.7) was larger than that explained when entering a single generic or content-specific teaching practice. On the contrary, the third approach did not confirm the study's hypothesis, as the percentage explained when entering the second-order factor of student ratings into Model 3j (see Table 4.7), was the same with the percentage explained when examining the contribution of *time and classroom management* factor.

Nevertheless, and irrespective of the amount of percentage explained, the fact that the two out of three approaches confirmed that the combination of these two types of practices can explain more variance at the teacher level compared to that explained when considering either type of practices in isolation corroborates the study's main assumption about the added value of combining generic and content-specific practices in exploring student psychomotor learning; and thus, it answers the second research question. The theoretical, methodological, and practical implications of this result are detailed in the next chapter.

Determining Generic and Content-Specific Teaching Practices That Discriminate Among Teachers Based on Their Level of Effectiveness

The third research question was intended to explore which teaching practices (generic and/or content-specific) could discriminate among teachers based on their level of effectiveness. To address this question, after classifying teachers into three groups (most effective, typical, and least effective, see the exact process in the data analysis section), a discriminant function analysis (Burns & Burns, 2008) was conducted to identify which generic and/or content-specific teaching practices were the best predictors of teachers' effectiveness.

In particular, discriminant function analysis produced two discriminate functions: the first function that distinguished among those teachers who were found to be the most-effective from the other two groups of teachers (i.e., typical, and least-effective), and the second function that distinguished between typical and least-effective teachers. However, the significance of Wilks' lambda indicated that only the first function was statistically significant ($p<.05$), thus, only the teaching practices that distinguished among most-effective and the other two groups of teachers were investigated.

The eigenvalue that emerged indicated that the first function accounted for 43.69% of the variance. In addition, as can be calculated from Table 4.8, the percentage of the teachers which were correctly classified was 69.39%[33]. However, six least-effective and four most-effective teachers were expected to be typical; whereas three typical teachers were misclassified as least-effective and one as most-effective. Finally, there was only an extreme case of one most-effective teacher who was misclassified as least-effective.

Table 4.8

Classification Results of Teacher Effectiveness

	Predicted Group Membership			
Groups of Teachers	Least-Effective	Typical	Most-Effective	Total
Least-Effective	9(60.0%)	6(40.0%)	0 (0.0%)	15
Typical	3(12.5%)	20(83.3%)	1(4.2%)	24
Most-Effective	1(10.0%)	4(40.0%)	5(50.0%)	10

Turning to the predictors, which are the emphasis of the third research question, as can be seen from Table 4.9, almost all the generic and content-specific teaching practices found to significantly contribute to student psychomotor learning were deemed important predictors of teacher effectiveness (except one generic--*modeling* and one content-specific--*task progression*). A careful examination of the importance of each predictor suggests that one generic and one content-specific teaching practices were found to be the strongest predictors. In particular, *questioning*[34] was the strongest predictor that helped distinguish among most-effective and non-most effective (i.e., typical or least-effective) teachers, while *student quality practice*[35] was next

[33] This percentage was calculated by taking the ratio of the number of teachers that were correctly classified (i.e., N=34, see teachers on diagonal), to the total number of participating teachers (N=49).
[34] Refers to the first-order factor of questioning of the *DMEE* high-inference form.
[35] Captured with student ratings.

in importance as a predictor. Differently put, these two teaching practices, stood out as those that strongly predicted allocation to the most-effective or non-most effective (i.e., typical or least-effective) teachers. In particular, what seems to mainly differentiate most-effective teachers from the non-most effective was the use of clear questions in terms of their content and the remediation of students’ wrong answers, and the increase of student quality and congruent practice trials. In addition, providing and relating to lessons’ goals the importance for which a specific task or lesson was taking place (i.e., *quality of orientation*), minimizing waiting time, efficiently dealing with student misbehavior, demonstrating the emphasized psychomotor skill and providing congruent and specific feedback during student practice, were important but less successful predictors of teachers’ effectiveness.

Table 4.9

Structure Matrix Based on Analysis of Teacher Effectiveness

Teaching Practice	Function 1 [a]
Questioning[c]	.734
Student Quality Practice [f]	.500
Orientation (quality)[b]	.457
Time management (waiting)[d]	.452
Classroom Disorder [c]	.440
Demonstration (frequency)[d]	.353
Congruent and specific feedback (frequency)[d]	.352

Notes. [a] Function 2 is not presented in the table as it was found to be not statistically significant, and thus, it was not used for interpretation.
[b] data captured with *DMEE* low-inference form; [c] data captured with *DMEE* high-inference form; [d] data captured with *mTSS* low-inference form, [e] data captured with *mTSS* high-inference form; [f] data captured with student ratings.

Summary of Findings with Respect to the Three Research Questions

Taking into consideration the above research findings, the obvious answer to the first research question is that *each type of practices seems to contribute to student psychomotor learning*. In particular, five out of seven generic teaching practices were found to contribute to student psychomotor learning, with some having a greater effect (i.e., *orientation, time management,* and *questioning*) than others (i.e., *classroom management* and *modeling*). However, two generic teaching factors (i.e., *application* and *structuring*) were not found to have a statistically significant effect on student learning. Turning to content-specific teaching practices, four out of five factors were found to contribute to student achievement. As was the case with generic teaching factors, their contribution had different degrees and shades. *Demonstration* and *congruent and specific feedback* were found to have the largest impact on student psychomotor development, while *task progression* and *quality of student practice* had a smaller contribution. *Task explicitness* was the only content-specific teaching factor that was not found to have statistically significant effects on student learning; yet, this might be due to the fact that the vast majority of teachers gave partially explicit or a combination of partially and fully explicit task descriptions. Collectively, after controlling for certain student background characteristics, the generic teaching practices were found to explain between 2.77% and 21.70%; whereas content-specific practices between 2.77% and 16.31% of the unexplained variance at the teacher level of Model 2[36], a fact that suggests that each type of teaching practices significantly contributes to student psychomotor development. Nevertheless, the reader is reminded that the purpose of this study is not to pit the one type of practices against the other, but to explore their *individual* and *joint* contribution to student psychomotor learning.

Turning to the second research question, the findings corroborated to a great extent the study's main assumption about the added value of combining generic and content-specific practices in exploring student psychomotor learning. In particular, the percentage of the variance that was explained when combining generic and content-specific teaching practices (first approach) or when using the *mTSS* composite score (second approach) which involved both generic and content-specific practices was larger than the corresponding percentage of explained variance that emerged when

[36] See footnote 21 on how these percentages were calculated.

combining only generic or content-specific teaching practices. However, study's hypothesis was not confirmed when employing the third approach of examining the joint contribution of generic and content-specific teaching practices. Specifically, during this approach the second-order factor of student-ratings that combined both generic and content-specific teaching practices was entered into the model and explained an equal percentage of variance to that explained when examining the contribution of one generic teaching factor (i.e., time and classroom management).

As far as the third research question is concerned, the study's outcomes underlined that both generic and content-specific teaching practices are important predictors that can help discriminate among teachers based on their level of effectiveness. Particularly, the use of clear questions and the remediation of students' misconceptions/wrong answers as well as the increase of on-stated-task and appropriate student practice trials were the teaching aspects that strongly predicted the allocation of teachers to the most-effective or non-most effective (i.e., typical or least-effective) categories. Moreover, providing and relating to lesson's goals the importance for which a specific task or lesson was taking place, minimizing waiting time, efficiently dealing with student misbehavior, demonstrating the emphasized psychomotor skill and providing congruent and specific feedback during student practice were found to be important but not as strong predictors of teachers' effectiveness as the abovementioned teaching aspects.

CHAPTER 5: DISCUSSION AND CONCLUSIONS

Abstract

This study attempted to combine generic and content-specific teaching practices to explore teaching quality in PE and its impact on student learning. Through this exploration, several important findings have emerged. First, the present study confirmed teachers' important role for student psychomotor learning. Second, it contributed to the existing literature in two ways: by corroborating existing research findings that underlined the importance of certain generic and content-specific practices, and by highlighting the contribution of other generic and content-specific teaching practices whose effect on student psychomotor learning has not been adequately explored so far. Third, the present study underscored the added value of combining generic and content-specific teaching practices to explore student learning, as opposed to considering each type of practices in isolation. Finally, the study's outcomes supported that teaching aspects from both types of practices could serve as important predictors that could discriminate among most-effective and non-most effective teachers. All these results, which should be interpreted in light of the study's limitations, have theoretical, methodological, and practical implications which are outlined along with suggestions for future research in the field of teaching effectiveness in PE.

The present study sought to investigate teaching quality in PE and its impact on student psychomotor learning. Through this investigation several key findings have emerged. This chapter is structured as follows: first, the important role of teachers and teaching on student psychomotor learning is discussed, followed by the contribution of certain student background characteristics to student psychomotor development. After discussing the dual contribution of study's findings to the existing literature of teacher effectiveness in PE, a preliminary attempt to exemplify how an integration of generic and content-specific teaching practices could be pursued to develop a comprehensive and parsimonious conceptual framework and its associated observation instrument is described. Then, the generic and content-specific teaching practices that distinguished among most-effective and non-most-effective teachers in PE are discussed. Finally, the theoretical, methodological, and practical implications are outlined along with additional suggestions for future research. The chapter ends up with a short conclusion.

Teachers and Teaching Do Matter For Student Learning

In line with previous studies in PE and other subject matters (e.g., Castelli & Rink, 2003; Hattie, 2009; Konstantopoulos, 2012; Rivkin et al., 2005), this study has confirmed that teachers are important for student learning, since a significant percentage (i.e., about 15%) of variance of student achievement was situated at the teacher level. Admittedly, this percentage is not large, something that is probably related to the teacher sample that participated in the study. As mentioned in the methodology section, the study sought to recruit a specific sample of generalist teachers (i.e., classroom teachers who taught PE and Mathematics to the pupils of their class and teachers who taught only or mainly PE to students of their school); a sample that might not exhibit large differences in the way they taught PE. Although situated at the lower threshold, this percentage seems to be in par with the percentages found by other studies that examined either psychomotor (in PE) or cognitive learning (in other subject matters), and which fluctuated between 15-30% of variance (Kyriakides, Campbell, & Gagatsis, 2000; Kyriakides & Creemers, 2008a, 2008b, 2009; Kyriakides & Tsangaridou, 2008; Panayiotou et al., 2014; Teddlie & Reynolds, 2000).

By examining what is it about teachers that matters most for student learning, the findings of this study pointed to the actual teaching process rather than teacher background characteristics, a result which is also consistent with previous studies (e.g., Creemers & Kyriakides, 2008; Kyriakides & Tsangaridou, 2008). In other words, teachers' employment of certain teaching practices was found to have a significant effect on student psychomotor learning, whereas none of the teacher characteristics examined (e.g., gender, teaching experience in PE) was found to significantly contribute to student learning.

Student Background Characteristics That Influence Student Psychomotor Learning

However, the large amount of variance (i.e., about 85%) was situated at the student level. The vast majority of this variance (i.e., about 70%) was explained by students' initial performance, which also explained a significant percentage of variance at the teacher level, leaving only about 6% at this level to be explained by other variables. Yet, this is a common finding in other studies as well, showing prior knowledge to have the strongest effect in predicting student achievement at the end of the school year (Kyriakides & Creemers, 2008a; Kyriakides & Tsangaridou, 2008; Panayiotou et al., 2014). In addition, student gender was found to have a significant effect on student psychomotor learning, and specifically, boys were found to perform significantly better than girls, a result which resonates with previous studies (e.g., Barnett, van Beurden, Morgan, Brooks, & Beard, 2010; Butterfield et al., 2012; McKenzie et al., 1998; Zhu et al., 2011). Some of these studies (e.g., Barnett et al., 2010; Butterfield et al, 2012) attributed these differences to heredity, lack of practice for girls, and encouragement and prompting of boys to participate in activities that mainly require the exercise of manipulative and object control skills (e.g., throwing, catching, kicking).

Moreover, the socioeconomic status of students' family, and particularly the social status of father's job, was found to significantly contribute to student psychomotor learning, a result that corroborates previous findings in PE, even in preschool age (e.g. Kyriakides & Tsangaridou, 2008; Venetsanou & Kambas, 2010). The differences in psychomotor development among students of different socioeconomic status might be attributed to the more out-of-school PE opportunities and experiences that students from families with higher socioeconomic status have. This assumption was made based on the linear or near linear relationship among these two variables, which was implied by the fact that when the variable of participation in out-of-school

PE activities was entered in Model 2, the effect of the socioeconomic status variable (i.e., father's job social status) either disappeared or became statistically significant at level $\alpha=0.10$. The practical implication of this finding along with the gender differences that were found in student psychomotor learning is discussed below, when considering how *differentiation* could help closing the gap among those groups of students.

Enhancing and Corroborating Existing Research Evidence

Another key finding of this study was that certain generic and content-specific teaching practices were found to have a significant effect on student psychomotor learning. By examining the individual effect of each teaching practice on student achievement, this study not only corroborated previous research findings related to the importance of employing certain generic and content-specific teaching practices when teaching PE, but also provided research evidence about the contribution of certain teaching practices that had not been adequately explored in PE so far. Beginning with the generic teaching practices, and specifically with the study's findings that provided support to previous research outcomes, this study underlined the important role that time and classroom management play when teaching PE, something that was highlighted by several previous PE research works (e.g., Fink & Siedentop, 1989; Hickson & Fishburne, 2004; Jones, 1992; Kyriakides & Tsangaridou, 2008; van der Mars, 2006). In particular, this study found that teachers who started PE lessons on time, minimized non-instructional tasks and waiting/transition time, maximized practice time, and effectively dealt with misbehavior, had students who experienced higher learning gains in their psychomotor skills than those teachers who did not effectively used these practices. Therefore, this study reaffirmed that maximizing instructional time and student on-task behavior are among the most strongly related variables to student (psychomotor) learning outcomes (Muijs et al., 2014; van der Mars, 2006). In addition, although classroom management is not directly related to the content to be taught, the study findings confirmed the essential role of establishing and maintaining an orderly and effective PE learning environment (Rasmussen et al., 2014; Rink, 2013; Siedentop & Tannehill, 2000).

Turning to the generic practices whose effect on psychomotor learning has been underexplored, this study provided research evidence about the important role that the practices of orientating students towards learning goals, posing good questions and encouraging students to develop and use their own strategies for solving problematic situations (*modeling*) have on student psychomotor learning. Despite the typical guidelines provided by scholars about the important role of these three practices, empirical evidence supporting the effect of these practices on student psychomotor learning is rather scarce. Interestingly, this study classified two of these practices (i.e., questioning and orientation) among the three most effective generic teaching practices that contributed to student psychomotor learning.

Specifically, students whose teachers were explaining at least twice in a lesson the importance for which a specific task or lesson was taking place and related this importance to lesson's goals, had better learning outcomes than their counterparts whose teachers did not use these practices. The importance of this teaching practice seems to lie on its capability of making tasks and lessons meaningful to students, something that encourages their active engagement in the lesson and makes student practice highly focused (Creemers & Kyriakides, 2008; Siedentop & Tannehill, 2000). However, although orientation has been endorsed as an important generic teaching practice in classroom disciplines, in PE has scarcely been explored, something that stresses the importance of this finding.

Similarly, although posing good questions is one of the most widely studied aspects of teaching in classroom effectiveness research (Muijs et al., 2014), in PE it has only received attention when examining the effectiveness of certain pedagogical models (e.g., Teaching Games for Understanding) that emphasize tactical awareness (i.e., selecting the appropriate responses to solve tactical problems that arise during a game). On the contrary, this study investigated how the use of effective questioning techniques could impact on student psychomotor learning and found that teachers who were asking clear questions during all three main lesson phases (i.e., introduction, core, end of the lesson) and were constructively dealing with students' incorrect answers, helped their students to achieve greater learning gains than those who did not. These findings expand the applicability of effective questioning techniques to the field of PE, and resonate with the principle that effective questioning techniques can give the opportunity to students to take responsibility for their own learning (Muijs et al., 2014), since through the use of questions, students are puzzled about the emphasized psychomotor skills and are encouraged to actively participate in the learning process by expressing their opinions.

As far as the practice of modeling is concerned, although only a small percentage of sampled teachers (i.e., 13.6%) employed the guided discovery or the discovery approach of modeling (see Table 4.1), it was found that the teachers who used this practice had students who performed better than their counterparts who did not (even when teachers were giving the strategy to students). This finding corroborates the admittedly few existing research findings (e.g., Chatzipanteli & Digelidis, 2011; Hodges & Franks, 2002; Lidor, 2004) that highlighted modeling and discovery learning methods as important for teaching psychomotor skills. In addition, it suggests that by encouraging students to use strategies and/or develop their own strategies to successfully perform a psychomotor skill or solve different problematic situations (Creemers & Kyriakides, 2008), teachers motivate students to undertake the responsibility of their own learning and develop useful life-long skills (i.e., self-regulation, meta-cognitive skills).

Nevertheless, the abovementioned findings that concern the contribution of orientation, questioning, and modeling to student psychomotor learning need to be supported by more research studies, as the literature review chapter underlined that the effect of these three generic teaching practices on student psychomotor development has been underexplored.

As was the case with the generic teaching practices, this study corroborated existing research evidence supporting the contribution of certain content-specific teaching practices to student psychomotor learning. For instance, the findings of this study that pertain to the contribution of skill demonstration and teacher's ability to increase the quality of student trials, support the existing research evidence which underlined these two practices as crucial when teaching psychomotor skills. For instance, in line with Chen et al.'s (2016) findings, the study's results underlined demonstration as one of the most significant PE-specific teaching practices, supporting its ability to help students transform the abstract concept of the emphasized skill to something more concrete and tangible (Hunt et al., 2009; Valentini, 2004). However, contrary to previous studies, which showed that students who received full demonstration accompanied with succinct learning cues performed better than their counterparts who received a partial demonstration (e.g., Janelle et al., 2003; Kwak, 2005), in this study the quality dimension of demonstration (i.e., whether the skill was presented verbally, practically or by combining these two ways; or if the skill was partly or fully presented) was not found to contribute to student psychomotor learning. This result is probably associated with the inherent difficulty of classifying these categories from most to least important and the fact that more than one of these categories might be present in a single lesson. In contrast to the methodology pursued by Janelle et al.'s (2003) and Kwak's (2005) studies, which followed an experimental design with straightforward groups (see Table 2.8), in this study teachers employed different combinations of the constituent components of this practice within a single PE lesson, something that complicated the classification of these aspects. For example, it sounds reasonable to assume that a full demonstration might be more helpful for students than a partial one; yet, some teachers might prefer to break the demonstration of the emphasized skill into smaller parts and present one part at each time, something that might be at least equally effective as demonstrating the whole skill. Similarly, it seems difficult to discern which of the following is more effective: a teacher who chooses to visually and verbally demonstrate a skill just once, and then remind the students of the key learning cues in a verbal manner or a teacher who chooses to visually and verbally demonstrate a skill twice in a lesson without reemphasizing the learning cues? Therefore, although previous research findings support that a full demonstration is typically better than a partial one (e.g., Kwak, 2005) or a demonstration that combines visual and verbal aids is usually more effective than just using either of the two ways (e.g., Janelle et al., 2003), this study suggests that the quality aspect of this teaching practice should be scrutinized in more depth, in order to unravel in which cases one approach of demonstrating a skill might be more effective than the other.

As far as the quality of student practice is concerned, in line with previous studies (Ashy et al, 1988; Buck et al., 1991; Silverman, 1985, 1990, 1993), this study underlined the importance of increasing students' appropriate and on-stated-task trials to student learning in PE. Although, a significant relationship between this teaching factor and student psychomotor learning was only established when using the data captured with student ratings, this teaching practice was found to be among the two most significant predictors that helped distinguish between most-effective and non-most effective (i.e., typical or least-effective) teachers. This result is consistent with PE scholars contention that this teaching practice represents the ultimate teaching skill in PE (e.g., Ennis & Chen, 2011; Rink, 2003; Silverman, 2011), since most other teaching practices apparently affect psychomotor learning through their influence on student practice. Nevertheless, why data captured through observations did not yield similar results might be due to three reasons that relate to the inherent difficulties and limitations of the process of coding student trials followed in this study. First, as explained in the data analysis section, low-inference data that pertained to the quantity of student practice were not included in the analysis, because of the difficulty in distinguishing when a continuous skill (e.g., basketball dribbling, running) is beginning and/or ending (see note b in Table 3.9). However, this might influence the data related to the quality of student practice, since the percentage of the appropriate/congruent trials was calculated by taking the ratio of the number of appropriate/on-stated-task trials to the total number of trials. Second, as far as the congruency dimension of student practice is concerned, a non statistically significant effect was found probably because the vast majority of student trials were performed as stated by the teacher (about 92% of observed lessons, captured with the low-inference form, involved at least 75% '*on-stated-task'* student trials; whereas about 82% of observed lessons captured with the high-inference form, coded that '*on-stated-task*' student practice was occurring to a great extent). Third, and turning to the appropriateness of student trials, this aspect has not been found to be statistically significant when using data captured with observation forms, probably because of the complexity of measuring it through observations. As mentioned in the methodology section, the six targeted students of different skill level, whose practice trials were coded, were chosen by the PE teacher based on his/her experience of teaching PE to the specific class. This process, in conjunction with the fact that students of different classes were practicing psychomotor skills that had an uneven degree of difficulty (e.g., some teachers might choose to teach psychomotor skills that were not familiar to their students, something that would probably yield more inappropriate trials, while others might choose to re-teach some skills, something that would lead to more appropriate trials) might have caused what is known as '*noise*' to the data.

Furthermore, the findings of this study supported one of the most consistent findings related to task progression, namely that refining tasks are the building blocks of successful skill development and that when employed in PE lessons they can lead to improvements in psychomotor learning (Masser, 1987; Pellett & Harrison, 1995a; Rikard, 1992; Rink et al., 1991). In addition, this study confirmed research works (e.g., French et al., 1991; Hebert et al., 2000) supporting that the teachers who are altering the conditions to make practice gradually more complex (i.e., extending task), help their students develop their psychomotor skills to a greater extent than their counterparts who are not doing so; thus, confirming Rink's (2010) argument that what eventually helps students to satisfactorily develop their psychomotor skills is the interplay among refining and extending tasks. Not surprisingly, however, descriptive statistics showed that the refining tasks were only employed in about 21% of the observed lessons, confirming previous reports underlining this type of instructional tasks as the most neglected (Hastie & Siedentop, 1999; Tannehill et al., 2015), and calling for paying more attention to this type of task in pre-service and in-service professional development opportunities; something that is discussed below.

Moreover, as was the case with certain generic teaching practices, this study has provided evidence about the significance of certain content-specific practices, whose impact on student psychomotor development has not been adequately explored. Such practices refer to the use of diverse instructional tasks (i.e., informing, refining, extending, and applying) and congruent and specific feedback. Considering first the use of the quartet of instructional tasks proposed by Rink (2010), although scholars support that there is not an ideal sequencing of these tasks, this study found that the use of at least three different types of instructional tasks within a single lesson could help students significantly improve their psychomotor skills; a finding which contributes to the limited research evidence that relates to the impact of this quartet of instructional tasks to student psychomotor learning. Interestingly, teachers who employed all four types of tasks within a single lesson, typically pursued the following sequence of tasks: first, they gave the opportunity to students to practice the emphasized psychomotor skill of the lesson without demonstrating the correct way of performing the skill (*informing)*. Then either a discussion on the correct form of the skill was followed or the teacher or a student demonstrated the appropriate form of executing the skill by giving emphasis to its critical learning cues and asking students to apply them during practicing the skill (*refining)*. After working on the critical learning cues of the skill, students were asked to engage in activities that required from them to perform the skill in more complex ways (*extending)*. Finally, usually at the end of the lesson, teachers asked students to apply the emphasized skill in a situation similar to that in which the skill is finally applied (e.g., a modified game). It seems that this course helped students gradually conquer the emphasized skill, as it provided to students the opportunity to experiment with the emphasized skill and work on its learning cues, which, in turn, might make them more confident to perform more complex tasks. However, it is underlined once again that this was a typical sequence which was observed in the lessons that involved all four types of instructional tasks, and not an ideal sequence of the quartet of tasks.

As far as the congruent and specific feedback is concerned, the results of this study supported the limited previous research findings underlining this type of feedback as crucial when teaching psychomotor skills (e.g., Pellett & Harrison, 1995b; Sariscsany et al., 1995;

Silverman et al., 1995). In particular, teachers who were providing congruent and specific feedback during student practice had a larger impact on student achievement than those who did not. As a consequence, although research findings that pertain to the generic nature of feedback in PE have been inconsistent (cf. Lee et al., 1993), it seems that these two specific aspects (i.e., congruency and specificity) of the multidimensional concept of feedback have a significant effect on student learning in PE. The importance of these two aspects of feedback lies on the fact that both of them target the critical learning cues of the skill and direct students' attention to successfully apply these cues, and therefore, to perform correctly the emphasized skill. In particular, specificity and congruency relate to the extent to which feedback is specific and matches the emphasized learning cues of the psychomotor skill which is performed (Metzler, 2011). In addition, this type of feedback seems to be even more important when offered during student practice (Rink & Hall, 2008). This is because during practice time students will have the opportunity to alter the way they perform the skill and practice the appropriate form before the emphasis or the complexity of the task alters.

Nevertheless, beyond the above teaching practices which were found to significantly impact student learning, some other generic and content-specific teaching practices were not found to do so. In particular, from the generic teaching practices, none of the dimensions of application and structuring were found to contribute to student learning. Contrary to the existing research findings, which support that students increase their psychomotor competence when their teachers offer a large number of instructional tasks (i.e., Constantinides et al., 2013; Rasmussen et al., 2014; Silverman et al., 1998), include easier and more complex tasks in their lessons (French et al., 1991; Hebert et al., 2000), and use structuring comments to facilitate student practice and learning (Rasmussen et al., 2014; Silverman et al., 1988), the results of this study did not show such significant relationships. However, this does not imply that these two generic teaching practices are not important for teaching PE. Rather, these results might be due to observers' insufficient content knowledge and unfamiliarity with the context of PE, both of which seem to be required to adequately capture these two generic teaching practices, as suggested by the examples described below.

As mentioned in the methodology section, five out of the six observers who used the *DMEE* instrument were master's or doctoral students in post-graduate degrees in Educational Administration and Evaluation, and therefore, they might not have been familiar with the context and the content of PE; whereas all four observers who used the *mTSS* instrument were master's or doctoral students in post-graduate PE and sport pedagogy degrees, which implies that they were more familiar with the PE context and they probably had sufficient knowledge of the content of PE. The absence of such knowledge and familiarity from the observers who used the *DMEE* instrument might lead to difficulties on how to code these two practices. For example, when coding the application time, observers who used the *DMEE* instrument did not focus on certain students but observed the class as a whole. As a consequence, in cases where students were waiting in lines (e.g., four lines) and only a few students (i.e., the first student of each line) were practicing the emphasized skills, these observers coded all the time period that the task took place as application time, without subtracting the waiting time. This fact is supported by the moderate correlation (i.e., r =.52) found between the '*application-frequency*'[37] variable of the *DMEE* and the '*total practice time*' of *mTSS*, which are supposed to capture the same construct (i.e., time on task). Similarly, due to their insufficient content-knowledge, these observers might

[37] As mentioned in the methodology chapter, however, for the frequency dimension of the application factor, the number of instructional activities has been used instead of the total application time of the lesson, because observers did not subtract the waiting time when capturing total application time, and as a result, this teaching aspect was not representative of the total practice time of students.

not be able to distinguish whether a subsequent activity was either more complex, of similar difficulty or easier from the previous (i.e., structuring); or even if an instructional activity was related to the lesson's goals, and as such, it should be counted as an application task or whether it was a warm-up activity, and thus, it should not be counted as an application task (i.e., application-frequency). The above observations/considerations seem to be related to Rink's (2014) recommendations that generic teaching practices need to be coded by observers who have some expertise in PE and a good understanding of the PE context. As a consequence, observer training and familiarity with the PE content and context seems to be imperative; a fact that is discuss below, when considering the methodological implications of this study.

From the content-specific teaching practices, the only practice not found to have a significant impact on student psychomotor development was task explicitness, a result that might probably be due to two reasons. The first reason, which is more methodological, pertains to the small variance of this teaching practice, which might obscure any effect that this practice might have on student learning had this variance been larger. In fact, from all the observed lessons, no implicit directions were recorded; in contrast, the vast majority of teachers mainly described the instructional tasks by giving partly explicit directions (see Table 4.2). Why this might be the case may be associated with teachers' conviction that providing students with the criterion/a against which they can judge their success or completion of a task represents redundant information, since students can have on-stated-task behavior even when presenting tasks in partially explicit terms. This conviction is supported by previous research works which found partially explicit instructional tasks to lead to a high rate of on-stated-task behavior (e.g., Tousignant & Siedentop, 1983; Silverman et al., 1995). However, this is just a speculation that warrants further investigation. The second reason relates to the significance of this teaching practice. As some scholars have supported (Silverman et al., 1998), task explicitness may not be as important as it was thought, since students' on-stated-task behavior can be controlled and adjusted by teacher's active supervision of student practice (Hastie & Siedentop, 1999).

Toward an Integration of Generic and Content-Specific Teaching Practices

Beyond examining the individual effects of each generic and content-specific teaching practice on student psychomotor learning, the present study also sought to investigate the joint contribution of those two types of practices on student psychomotor learning. Trying to bring together these two types of practices, this study followed three different approaches: the first one compared the explained variance which emerged when entering into a single multi-level model certain teaching aspects that pertained to both generic and content-specific teaching practices (see Model 3h in Table 4.7) to the variance explained when entering into a model individual factors; the second approach compared the variance that got explained when examining the contribution of the *mTSS* low-inference Rasch composite score (which involved both generic and content-specific teaching practices) to that explained by individual factors (see Model 3i in Table 4.7); and the third approach contrasted the contribution of the second-order factor of student ratings, which involved teaching aspects that pertained to both types of teaching practices to the contribution of the first-order factors that involved a single generic or content-specific teaching practice (see Model 3j in Table 4.7). Although the third approach did not confirm the study's hypothesis about the added value of combining generic and content-specific teaching practices, as the percentage explained when entering the second-order factor into the model was equal to that explained when examining the contribution of the time and classroom management first-order factor (see model 3k in Table 4.4), the other two approaches corroborated the study's assumption. In particular, a larger percentage of the remaining teacher-level variance was explained when combining certain generic and content-specific practices or when entering the *mTSS* low-inference Rasch composite score than that explained when considering only generic or content-specific teaching practices.

Certainly, one could argue that the first approach, and in turn the abovementioned finding about the added value of combining generic and content-specific practices might seem unsurprising and predictable, as it is reasonable to expect that by adding more explanatory variables into a single model more variance would be explained. However, the findings of this study suggest that this is not necessarily the case. Particularly, some teaching aspects that were found to have a significant individual (e.g., modeling, see model 3f in Table 4.3) or cumulative (e.g., task progression, see model 3f in Table 4.7) contribution did not continue to make a significant contribution to student learning when simultaneously entered in Model 3h (Table 4.7). In addition, if this was the case, it would be expected that both types of practices (Model 3h) would explain the sum of the variances explained by each of Models 3a and 3f (Table 4.7). Therefore, the assumption that adding more variables to the model leads to a straightforward increase in the variance explained cannot be supported.

Nevertheless, finding that when combining generic and content-specific practices in exploring student psychomotor learning leads to the explanation of a greater percentage of teacher-level variance as opposed to considering each type of practices in isolation, this study underlines the need for considering different approaches when it comes to studying the complex and multi-faceted phenomenon of teaching. Until now, PE researchers have mainly investigated the effect of certain generic or content-specific teaching practices to student psychomotor learning, thus, creating a list of teaching practices that were characterized as '*necessary but not sufficient'* (Rink, 2013) for student learning. Instead, by combining these two types of practices, this study made a first step toward the construction of a more comprehensive picture of effective teaching in PE. Importantly, this combination could explain the 59.54% of the teacher level variance that remained unexplained after controlling for certain student and teacher background characteristics; a percentage that is deemed quite large, given that it was emerged by only considering teaching practices. In particular, this study found that during a PE lesson the teachers who efficiently dealt with student misbehavior, provided and related to lesson's goals the importance for which a specific task or lesson was taking place, minimized waiting time, posed good questions and remediated students' misconceptions/wrong answers (generic teaching practices), demonstrated the emphasized skill and provided congruent and specific feedback during student practice (content-specific teaching practices) had students who had greater learning gains. Therefore, having in mind the percentage of the teacher-level variance that got explained (i.e., 59.54%), one could argue that the employment of the abovementioned generic and content-specific teaching practices within a PE lesson could ensure the basic conditions for an effective PE lesson. Certainly, the remaining unexplained percentage (about 40%) implies that other factors that relate to the teacher and/or teaching (e.g., teacher knowledge) are deemed important and should be put under the microscope of research.

Practices That Distinguished Most-Effective Teachers From Non-Most Effective

Another important finding of this study concerns the four generic and the three content-specific teaching practices which were found to be important predictors that helped distinguish among most-effective and non-most effective (i.e., typical or least-effective) teachers. This finding suggests that teachers whose students achieved the greatest psychomotor development employed these seven teaching practices on a more frequent basis or more qualitatively than the rest of their counterparts. Therefore, although these seven teaching practices had different discriminant loadings, this finding further corroborates the study's assumption that effective teachers are not those who effectively use either the generic or the content-specific teaching practices, but those who adequately employ both types of practices when teaching PE. This argument is further corroborated by the fact that one generic (i.e., the use of clear questions and the remediation of students' misconceptions/wrong answers) and one content-specific (i.e., the increase of on-stated-task and appropriate student practice trials) teaching practices were the two strongest predictors of teacher allocation in most-effective and non-most effective (i.e., typical or least-effective) categories.

However, a more careful examination of the discriminant loadings of each practice suggests that generic teaching practices were strong (i.e., questioning) to moderate (i.e., orientation, classroom disorder, time management) predictors of teachers' allocation, whereas the content-specific teaching practices were moderate (i.e., increasing student quality practice) to moderately weak (i.e., demonstration and congruent and specific feedback) predictors. This result might imply that generic teaching practices could form the basis for an effective PE lesson. In addition, the fact that these practices are generic and content-free (Ward, 2013), may imply that this could hold true for other subject matters as well. However, this is just a research hypothesis which awaits empirical validation from future studies.

Undoubtedly, all the above research findings, which should be interpreted in the light of the study's limitations discussed in the first chapter, have several theoretical, methodological and practical implications and suggestions for future research which are discussed in turn.

Implications of the Study

Theoretical Advancements of Educational Effectiveness Models: Toward the Development of a Comprehensive Conceptual Framework and its Associated Observation Instrument

The findings of this study supported the appropriateness of using comprehensive generic and content-specific frameworks to capture teaching quality in PE. Both the *DMEE* and *mTSS* were successfully employed and captured teaching quality in PE, something that is supported by the following four facts. First, five out of the seven (assessment was not explored) generic factors incorporated in the *DMEE* (i.e., orientation, questioning, modeling, time management, and classroom as a learning environment) and three out of the five content-specific factors included in the *mTSS* (i.e., task progression, demonstration, and congruent and specific feedback) have been shown to have a positive effect on student psychomotor learning in PE. Second, three generic teaching practices (i.e., questioning, orientation, and classroom disorder) incorporated in the *DMEE*, one generic practice (i.e., waiting time) and two content-specific practices (i.e., demonstration, and congruent and specific feedback) involved in the *mTSS* were found to be important predictors that helped distinguish among most-effective and non-most effective teachers. Third, four out of the five dimensions (except differentiation) involved in the *DMEE* and all three dimensions used in the *mTSS* were found to have a significant impact on student psychomotor learning through at least one teaching factor; something that outlines the important role that other dimensions than frequency and quality have on measuring instruction in PE. Fourth, teachers who had been assigned higher composite scores[38] were found to help their

[38] Using the Rasch lesson estimates of the *DMEE* and *mTSS* low-inference forms, two composite scores were calculated for each teacher by taking each teacher's average lesson estimate score across the three lessons. A corresponding score was also calculated for the *DMEE* high-inference form. However, because it was not possible to combine all the first-order teaching factors to create an overall model due to multicollinearity reasons, the score representing the quality of teaching was calculated by taking the average of each of the eight teaching factors of the high-inference form (i.e., second-order factor score, see Model 3c in Table 4.7).

students develop their psychomotor skills in a greater extent than those who scored lower in these observation forms. In particular, students who experienced superior teaching quality as captured with the two forms of *DMEE* and the low-inference form of *mTSS*, exhibited larger gains in their psychomotor learning compared to their counterparts whose teachers scored lower in these forms of the instrument. Collectively, then, the above four facts suggest that both frameworks could sufficiently measure teaching quality in PE.

Therefore, having in mind that one of the most important next steps in capturing teaching quality in PE is the development of comprehensive conceptual frameworks that would measure teacher effectiveness (Rink, 2013), the findings of this study can be used to draw suggestions for developing one such framework and its associated observation instrument. Particularly, the findings of the study support that a new comprehensive framework could be developed involving both generic and content-specific practices. Although, such efforts have recently been undertaken in PE by Chen and colleagues (2011) who developed and validated the *AQTR* and by PE stakeholders in Singapore who developed the *PELOT* framework by changing the descriptors of *FfT* (see Rink, 2013), these efforts seem to miss that some content-specific teaching practices can stand alone (i.e., increasing students' quality trials, task explicitness, and task progression in terms of the focus of the motor task: informing, refining, extending, applying), and cannot be generated by 'specializing' generic teaching practices. In essence, despite acknowledging that generic teaching practices may suffer from being too general and insensitive to the particularities of PE content and context, these scholars did not *integrate* generic and content-specific teaching practices, but rather, they attempted to 'specialize' generic teaching practices by considering their particular functioning in PE lessons, a process that might ignore important content-specific teaching practices recognized herein.

Instead, when calling for an *integration* of these two types of practices, this study suggests that the PE researchers should first investigate which of these teaching factors from either type of practices might stand alone, and which can be integrated and how. This suggestion derives from the findings of this study (see multicollinearity problems discussed above, during examining the joint contribution of the two types of practices) as well as Kane and Staiger's (2012) study, which showed that generic and content-specific practices are two different but not totally distinct worlds, since they have some common points of intersection. If generic and content-specific practices were totally distinct, then a comprehensive framework could have been developed by simply juxtaposing them. However, since this is not the case, simply juxtaposing the two types of practices to develop a comprehensive framework would not only lead to redundancies, since some practices would be represented more than once, but also to an associated observation instrument that would be difficult and impractical to use.

Therefore, before undertaking such a complex process, scholars should identify the areas of convergence and divergence among the two types of practices, and determine which of these teaching factors from either type of practices stand alone and which can be integrated and how. Thus, a deep consideration of each practice is required to develop a more comprehensive, but at the same time parsimonious and practical conceptual framework. Drawing on the 13 teaching practices (i.e., eight generic and five content-specific) involved in the frameworks employed in this study, what follows is a preliminary attempt to exemplify how this *integration* can be pursued. First, the generic and content-specific practices that can function complementary to each other are considered, followed by a discussion of one content-specific practice which at a surface level appears to have an overlap with a generic practice, but it can eventually stand alone in a comprehensive framework. The teaching practices that are totally distinct are not discussed, since they can be incorporated in a comprehensive framework as they are. Figure 5.1 provides an overview of this attempt.

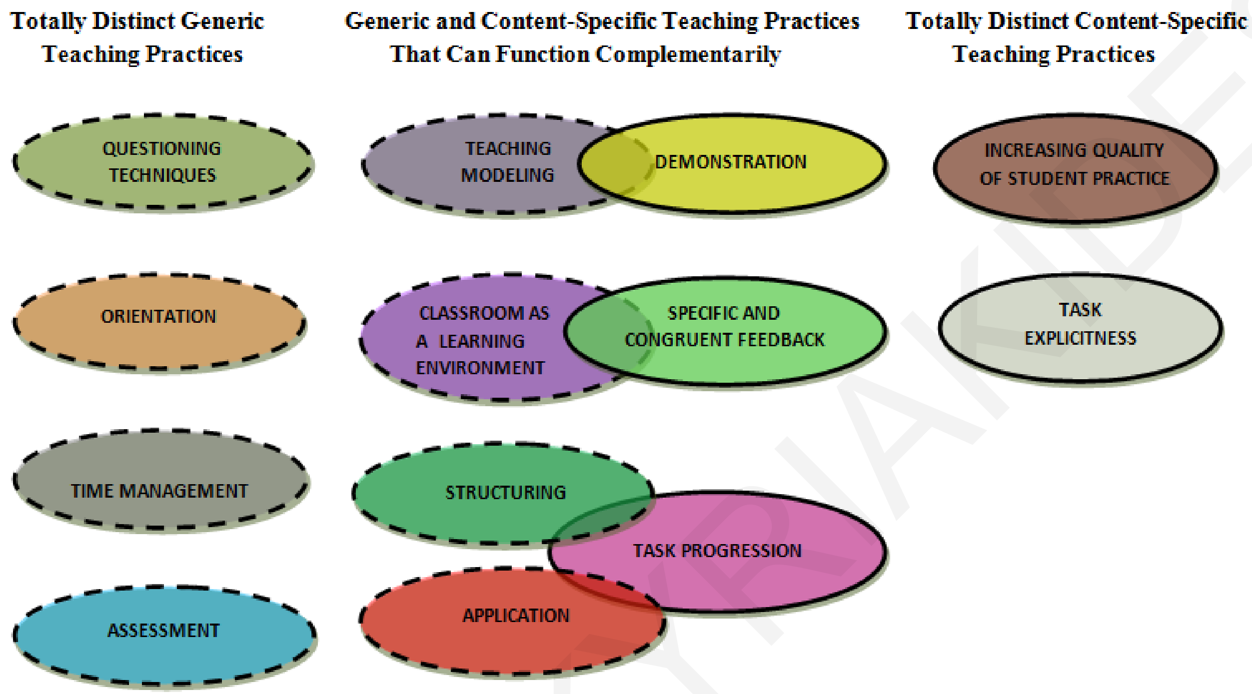


Figure 5.1. A preliminary attempt of integrating generic and content-specific teaching practices.
Notes. [a] Ovals with dashed lines represent generic teaching practices, while ovals with solid lines represent content-specific teaching practices.
[b] See Tables 2.1, 2.7, and 3.3 for a detailed description of each factor.

To begin with, consider the content-specific practice of providing students with specific and congruent feedback and the generic factor of classroom as a learning environment, and especially its aspect of providing generic feedback to students (teacher-student interactions). Both practices emphasize the need to provide students with feedback on their efforts. However, the one emphasizes the need for generic feedback, while the other underlines the need of directing students' attention toward the key learning cues of the emphasized skill. The findings of this study which supported that the simultaneous use of both aspects of feedback (i.e., reinforcement and congruent and specific feedback) can help students improve their psychomotor skills as well as the inconclusive findings of other studies on the effect of providing generic feedback (Lee et al., 1993) suggest that as a research community we might need to attend to both aspects of feedback—generic and content-specific—if we are to better capture the effect of this practice on student learning.

This complementarity can also be revealed when considering the content-specific practice of skill demonstration and the generic factor of modeling. Modeling is primarily concerned with presenting strategies to students or encouraging and supporting them to develop their own strategies to solve different types of problems (Creemers & Kyriakides, 2008). After students develop their own strategies, it is often common practice to encourage them to share these strategies with their classmates. It is at this juncture that the practice of demonstration could be productively employed to support student learning. For instance, a teacher could ask students to practice basketball dribble, trying to figure out different ways of dribbling a basketball without losing ball control. After students' experimentation, a whole-class discussion can ensue during which the ways discovered are presented; the correct execution of these different ways can be finally demonstrated (by students or the teacher), by emphasizing important learning cues that can support students to more effectively dribble the ball without losing it. Hence, modeling and skill demonstration seem to function complementary to each other when the teacher attempts to provide students with opportunities to not only discover effective ways of performing a motor task, but to also execute this task in appropriate ways that will maximize their performance.

A third example of practices that seem to function complementarily to each other concerns task progression and the generic factors of structuring and application. As far as the factors of task progression and structuring are concerned, both of them refer to the gradual increase of the difficulty level of instructional tasks given to students. However, the content-specific practice of task progression does not only involve sequencing the learning experiences from simple to complex or from easy to hard. Rather, it reflects a teacher's ability to combine the opportunity given to students for experimentation with the new skill (*informing* tasks), the progression of practice (*extending* tasks), the quality of performance (*refining* tasks) and the integration of application experiences (*applying* tasks) (see Rink, 2010). Likewise, the factor of structuring also involves teachers' attempts to give students clues as to how the lesson progresses (Creemers & Kyriakides, 2008), something that is not captured by the task progression. Collectively, then, these two practices seem to be capturing somewhat different aspects of instruction and hence can function complementary to each other. Turning to the factors of task progression and application, the four types of instructional tasks involved in task progression refer to different kinds of application opportunities that are provided to students to practice the emphasized skill(s). However, the dimensions of focus, quality, and differentiation of the application factor are not considered in task progression, something that suggests that these teaching factors can function complementarily to each other.

Considering the teaching practices that appear to have an overlap with each other, at a surface level, increasing students' quality trials seems to be identical to teachers' work to maximize the ALT (aspect of time management factor). However, time-on-task alone cannot adequately depict student engagement in PE (Siedentop, 2002). This is especially apparent in game-like situations, where students—particularly low-skill students—might appear to actively participate in the game, but their substantial engagement (e.g., appropriately practicing a skill like giving a catchable pass, or even touching the ball) is minimal, resulting in a negative relationship with their achievement (Rink, 1999). This fact was highlighted by a recent study which did not find any significant differences in the quantity and quality of student trials, among students practicing in either long lines (resulting in huge amounts of wait time) or in game-like activities involving the concurrent participation of the entire class (performing the same skills as in lines) (Hastie et al., 2011). Consequently, this teaching practice seems to capture a different teaching aspect than the ALT. As such, increasing student quality practice and time management should not be collapsed since they bring different perspectives, both needed to better understand the quantity and quality of student engagement with the content.

Turning to the development of the associated observation instrument of this new framework, this could have the structure of the low-inference form of the *mTSS* or the *DMEE*, and could be enhanced by the teaching aspects that are totally distinct or function complementarily to those of the other framework. For instance, take the low-inference form of *mTSS*, which already combines generic and content-specific teaching practices as the basis on which this new comprehensive instrument will be built. Beginning with the teaching practices that seem to significantly overlap, it would be redundant to add to this observation form the classroom disorder sub-factor of classroom management and the time management factor involved in the *DMEE* high-inference form (see Table 2.1), since it already captures the amount of time that students practice and the time that teachers devote to management issues.

On the contrary, the teaching factors of orientation and questioning and the teaching aspect of structuring that pertains to teacher's presentation of clues that guide students on how the lesson progresses seem to be totally distinct from the content-specific factors. Since all these teaching aspects engage students' cognitive domain, they could be embodied in the teaching aspect of *cognitive task* (see Table 2.7), which refers to tasks during which students attend to teacher's presentation of information about the lesson's objectives; pose/answer questions etc.

Thus, when cognitive tasks occur, the observer will have the opportunity to jot down information about the five dimensions of questioning, orientation and the aspect of structuring discussed above.

Turning to the factors that function complimentarily, the teaching modeling factor could be integrated with the demonstration factor. In particular, within the factor of demonstration a code capturing the three dimensions (i.e., focus, quality, and differentiation) of modeling that are not captured when demonstration occurs, could be added capturing whether the teacher encourages students to develop their own strategies for solving problematic situations. Likewise, the three dimensions (i.e., focus, quality, differentiation) of application and structuring (i.e., the aspect that pertains to the sequence of tasks form easier to more complex) could be added within the factor of task progression, since it does not captures these dimensions.

Collectively, then, having this preliminary attempt as a guide, future studies could invest in developing and validating more comprehensive frameworks and associated observation instruments that would ably incorporate an adequate set of generic and content-specific teaching practices. In addition, these works could also employ different generic (e.g., *CLASS*; *FfT*) and content-specific (e.g., *AQTR; QMTPS*) conceptual frameworks than the two employed in this study, to achieve this complicated task.

Practical Implications: Assessing and Improving Teaching Quality in PE

As a by-product of the above theoretical implications, such comprehensive frameworks can be used for both summative and formative purposes in teacher pre-service and in-service training. As far as the summative purpose is concerned, such instruments would provide a valid tool for teacher accountability, since they would capture an adequate set of teaching practices; whereas for formative purposes an observer would be able to identify the generic and/or content-specific areas that pre- or in-service teachers are in most need and offer specific feedback to support the continuous professional development of teachers. In addition, having in mind that generic teaching practices mainly require general pedagogical knowledge (Rink, 2013), whereas content-specific practices demand enhanced content knowledge to be effectively employed (Shulman, 1987), based on the identified areas that need to be addressed, the observer would be in place to recognize whether the teacher needs to improve his/her pedagogical knowledge (when the teacher does not employ certain generic teaching practices, or he/she does so inadequately) and/or his/her content knowledge (when certain content-specific teaching practices are absent from teaching or their use is not efficient).

A second practical implication that emerges from the results of the present study refers to the importance of providing pre- and in-service teachers with concrete examples and "hands-on" experiences of how to use certain generic (i.e., orientation, questioning, and modeling) and content-specific teaching practices (i.e., congruent and specific feedback, and task progression in terms of using all four instructional tasks, and especially refining tasks) in their work. Despite highlighted as important by many researchers (Rink, 2010; Siedentop & Tannehill, 2000), and found to improve student learning in the present and past studies, these teaching practices were infrequently used by the teachers participating in this study. For example, although refining tasks are considered the building blocks of successful skill development as they help students improve the quality of a skill (Rink, 2010), in this study only the 21% of teachers offered such tasks. Similarly, although modeling and constructivist teaching approaches have been underlined by PE researchers as crucial for student psychomotor and tactical learning (Chatzipanteli & Digelidis, 2011; Rovegno & Dolly, 2006), modeling was only captured in the 30% of the observed lessons. In the same way, while orientation is considered to help making learning more meaningful and instruction and student practice more focused on the emphasized skills (Siedentop & Tannehill, 2000), it was only observed in about the 30% of the lessons. Collectively, why this might be the case, has probably to do with teachers' content knowledge (as far as the content-specific teaching practices are concerned) and confidence of using such teaching practices during PE lessons. In addition, the infrequent use of some content-specific teaching practices cannot be dissociated from the fact that the participating teachers were generalist teachers, and they might not have the adequate knowledge or expertise to use those teaching practices. Therefore, it seems that teachers might need more ongoing, sustained, and meaningful professional development opportunities, which will be based on teachers' needs and interests, and they will provide teachers with concrete examples of how and under which circumstances each teaching practice could be employed during a PE lesson and how it could facilitate student learning (Patton, Parker, & Tannehill, 2015).

Beyond the teaching practices, this pattern was observed for some dimensions as well. Perhaps the most telling example of this phenomenon comes from the employment of the differentiation dimension. Despite that research evidence suggests that teachers should differentiate their tasks to help students of all skill levels to be successful (Silverman, 1993), this practice seems to have been very scarcely employed by the study participants (see Table 4.1). This infrequent use of differentiation is probably due to the fact that differentiation is a very demanding task for teachers, as among others they need to be flexible and adapt their instruction to suit their students' individual needs (Whipp, Taggart, & Jackson, 2014). Therefore, teachers would benefit from more practical examples of how to use differentiation during teaching PE (Rovegno & Bandhauer, 2016). In addition, given that this study reported significant differences on psychomotor learning between boys and girls and among students with different socioeconomic background, differentiation becomes even more necessary, as it could help minimize these differences.

Methodological Implications: Advancing Future Research Works

Turning to the methodological implications of this study's findings, first, more studies employing such a comprehensive conceptual framework as that described above seem to be needed in the field of PE as well as in other fields. By treating the findings of this study as initial indications about the added value of integrating generic and content-specific teaching practices in PE educational contexts, other (PE) researchers could replicate some aspects of this study in their own context and discipline, to shed more light into how the integration of both types of practices can help to better describe instructional quality and understand how it affects student learning in PE or other disciplines and learning domains as well. Certainly, these works could employ different conceptual frameworks than the two considered in herein. Although such efforts might be quite complex and difficult due to the inherent logistics, they could offer a deeper analysis of teaching, and they promise more powerful findings and implications.

In addition, moving the present work a step forward, these studies could also explore the correlations that exist among generic and content specific teaching practices, to help create a more rigorous understanding of the complex phenomenon of teaching. Despite, the present study made an effort to examine the correlations between generic and content-specific teaching practices on a theoretical basis, future studies could dig deeper and investigate more elaborated models in which generic and content-specific teaching practices interact with each other. This could lead to an even more advanced picture of how each generic and content-specific teaching practice functions in relation to the other teaching practices, and would help develop more comprehensive and parsimonious models to understand and measure teaching quality.

Second, the fact that some generic teaching practices were not found to have a significant impact on student learning probably due to observers' insufficient content-knowledge and unfamiliarity with the context and teaching of PE, calls for more careful and intensive training for all the observers as well as the scorers that participate in a study. Having in mind that '*good data are all you have*' (O' Sullivan, 2002), and that in the present study some factors were better captured when the observers had a relative to PE background, then, careful recruitment and PE-specified training of all the individuals who are involved in the data collection process is deemed essential. This suggestion seems to resonate with McKenzie and van der Mars' (2015) recommendations on reliably using systematic observation to assess physical activity or PE lessons.

Third, this study used student surveys as a supplementary measurement approach of instructional quality, as they functioned complementarily to and enhanced observation data (Kane & Staiger, 2012). Interestingly, in certain cases student ratings were found to function better than classroom observations in explaining student learning. For example, students were more capable of determining the quality of their practice, since, in contrast to observational data, student ratings on this practice were found to significantly contribute to student psychomotor learning. Why observational data did not yield similar results might be due to the inherent difficulties and limitations of coding student trials that have been explained above. Therefore, in studies where student practice trials are not of the same type (i.e., discrete or continuous) or cannot be directly observed, an alternative approach of doing so is to employ a student survey. Similarly, student ratings were found to explain a larger percentage of the unexplained teacher-level variance for the generic practices of time and classroom management as well as for the content-specific practices of skill demonstration and congruent and specific feedback, compared to observation data. This was not surprising, as student ratings aggregate the impressions of students who spend many hours with their teacher (MET, 2012; Peterson, 2000), and who observe his/her everyday teaching behaviors that might not be observed during the scheduled lesson observations. Therefore, although confirming the assumption that student surveys are difficult to complete with young students (Follman, 1995), and despite the difficulties encountered in validating student surveys (see methodology section), this study endorses the argument that when carefully constructed and validated, student surveys, even with third-graders, can offer a reliable tool for measuring teaching quality (Fauth et al., 2014), especially for studies with budget constraints, as this approach is considered cost- and time- efficient (Little et al., 2009; Peterson, 2000).

Additional Suggestions for Future Research

This study concludes with considering some additional suggestions for future research. First, the immediate next step of this study would be to adopt an experimental design and explore whether teachers who participate in professional development opportunities that target the use of the generic and content-specific teaching practices involved in the abovementioned comprehensive framework, could improve their teaching quality, and, in turn, their students' learning. Having in mind that the ultimate goal of teaching effectiveness research is to help the practitioners improve the quality of their teaching, and, in turn, the student learning, such studies are deemed imperative.

Second, the literature review of this study revealed a lack of PE studies that employ a multilevel model to study teaching quality. Such studies, which can follow an even more complex model (i.e., a three-level model, with students nested in teachers/classrooms, and teachers/classrooms nested in schools, see Kyriakides & Tsangaridou, 2008), than that employed herein, give the opportunity to the researchers to explore which factors at each level contribute to student psychomotor learning. Moreover, the literature review pointed to a number of studies that used only a post-test, and by doing so they only measured the impact of the emphasized variables on student *performance* at the time that the test was taken. Therefore, it is suggested for future studies to employ a pre-/post- (and even a retention-) test approach, to measure the effect of the explored variables on student *learning*. Collectively, if such attempts can be undertaken in conjunction with the abovementioned methodological suggestions, then a more authentic picture of the multidimensional nature of the teaching-learning process in PE will be provided.

Third, considering that quality PE programs should aim at developing all four learning domains (i.e., psychomotor, cognitive, affective, and social) future studies could investigate the effect of generic and content-specific teaching practices on students' cognitive, affective, and social development. By doing so, it would be examined which generic and content-specific teaching practices have a consistent effect on student all-round education.

Finally, the data analyses of this study revealed that the significant effect that certain teaching practices had on student learning, was followed by a plateau which implied that the further use of those teaching practices did not increase student learning (e.g., while engaging students in extending tasks was found to contribute to student learning, no differences were found among teachers who offered on average extending tasks that lasted more than eight minutes and those who did not offer this type of task). Since investigating why this might be the case was beyond the scope of this study, future studies could investigate the existence of curvilinear relationships among some teaching practices and student learning, and answer questions that pertain to the implementation of those teaching practices.

Conclusion

To come full circle, after recognizing that PE researchers have mainly pursued two parallel research strands in their effort to better understand what aspects of effective teaching relate to student learning in PE (Graham & Heimerer, 1981; Metzler, 2014), the present study attempted to combine these two perspectives to explore teaching quality in PE and its impact on student psychomotor learning. In doing so, this study first explored the individual contribution of certain generic and content-specific teaching practices and found that the majority of the explored teaching practices (10 out of 13 teaching factors) had a significant impact on student psychomotor learning. This finding resonated with several previous research works that underlined the contribution of certain generic and content-specific teaching practices to student psychomotor learning. In addition, this finding highlighted the important role that teaching has on student learning.

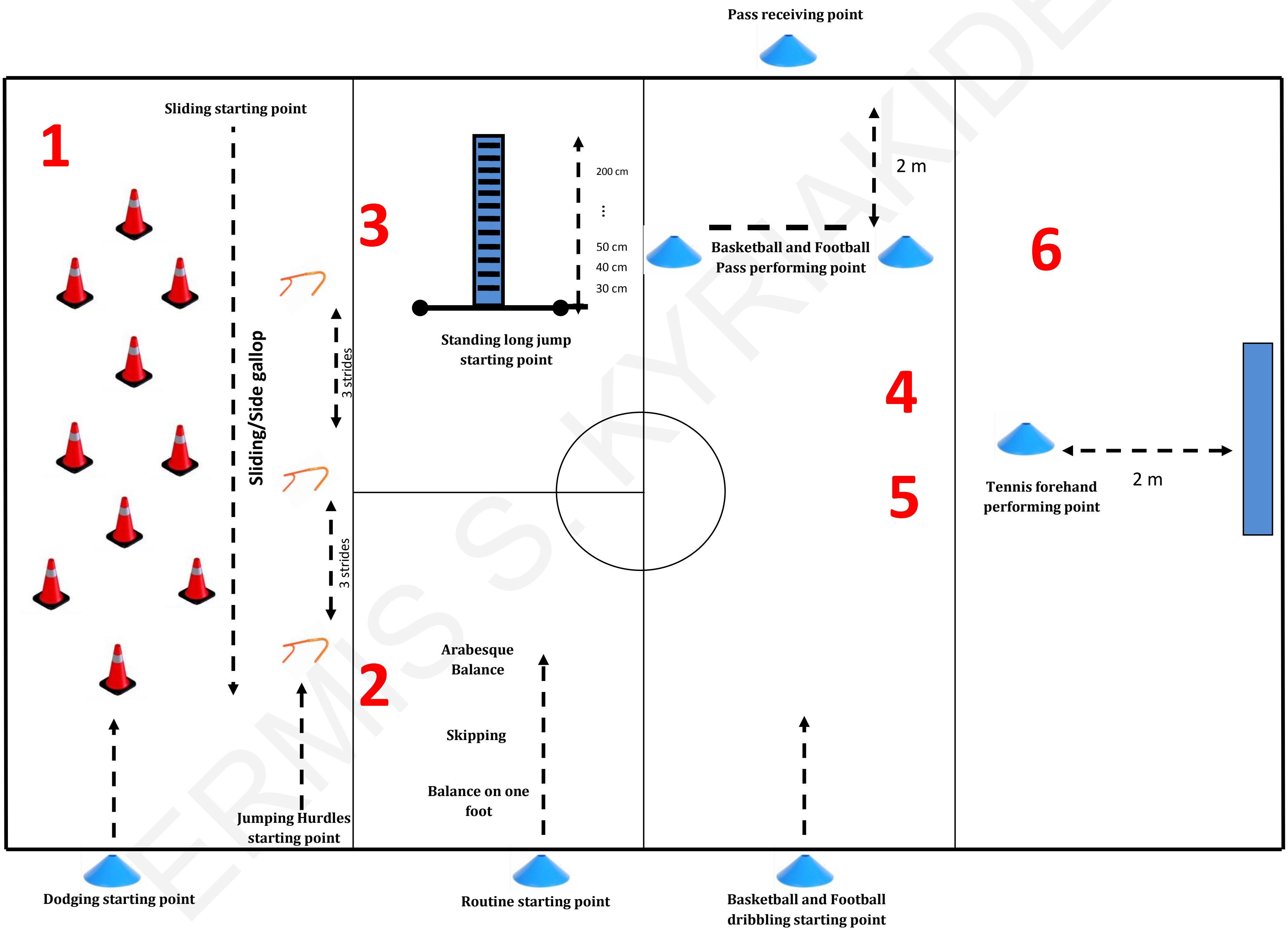
Moving the research works exploring PE teaching quality a step forward, the present study also sought to explore the joint contribution of generic and content-specific teaching practices to student psychomotor learning. The study's outcomes showed that the combination of generic and content-specific teaching practices could explain the 59.54% of the teacher level variance that remained unexplained after controlling for student and teacher background characteristics. This promising finding constitutes a preliminary indication that as a research community we can move toward the construction of a more comprehensive picture of effective teaching in PE, since the joint contribution of five teaching practices (four generic and one content-specific) could explain a significant percentage of the teacher unexplained variance. In addition, these teaching practices could discriminate among most-effective and non-most-effective teachers (something that constituted the third goal of the study).

Certainly, this research attempt does not aim to create another list of '*necessary but not sufficient'* (Rink, 2013) teaching practices. Rather, building on these preliminary findings, future research works could investigate more explanatory variables related to teachers and teaching (e.g., teacher knowledge), and try to explain an even larger percentage of the unexplained teacher-level variance; one that could render researchers confident enough to speak about '*necessary and sufficient*' teaching practices and conditions for teaching and student learning in PE.

Appendices

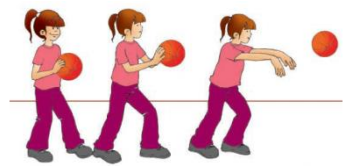
Appendix A: Psychomotor Test Organizing Diagram and Examples of Task Description, Assessment Scoring Rubric and Assessment Score Sheet

Psychomotor Test Organisizing Diagram



Example of Task Description

Task 4: Basketball Dribbling and Passing

Space: $\frac{1}{4}$ of basketball field (see task 4 of the psychomotor test organizing diagram).

Description: The scorer asks children to dribble the ball with their “strong” hand while jogging, starting from the starting point and stopping at the two cones which are 2 meters away from the passing point. When they reach the two cones, children catch the ball and perform a catchable chess-pass to their fellow-student who is standing at the pass receiving point.

Note: the student who is standing at the pass receiving point is the last numbered student of the class roster. This student receives all the passes and when his/her turn is shown up, then another student replaces him/her, so he/she can perform the task.

<u>Example of Assessment Scoring Rubric</u>

<u>Task 4</u>: Basketball Dribbling and Passing

a) Dribbling

<u>Selected critical cues:</u>

1. ***Pushing ball with upper finger pads***
2. ***Bouncing ball below waist height***
3. ***Maintaining ball in front of body and to the "dribble hand" side***
4. ***Keeping head up and eyes forward***

Level	Description of Performance
0	Looses ball control two or more times while performing the dribble
1	Dribbles very slowly or discontinuously (e.g., looses ball control for a moment or catches the ball while dribbling)
2	Dribbles the ball continuously, but he/she does not display all the selected critical cues (e.g., the student stares at ball instead of having his/her head up and eyes forward)
3	Dribbles the ball continuously by displaying all the selected critical cues with fluid motion

b) Chess-pass

<u>Selected critical cues:</u>

1. ***Holding the ball at chess height***
2. ***Making a step forward***
3. ***Pushing ball by stretching both hands***
4. ***Sending a catchable pass***

Level	Description of Performance
0	The pass is not performed correctly, and thus, the ball goes far beyond the fellow-student who waits to receive the ball
1	The pass is not performed correctly, but the ball goes somewhere near the fellow-student; yet, not in a catchable point (e.g., the ball hits the legs of the receiver)
2	The pass is performed correctly and goes to the receiver (in a catchable point), however, the student does not display all the selected critical cues (e.g., the student does not stretch his/her hands while performing the pass)
3	The pass is performed correctly by displaying all the selected critical cues with fluid motion and goes to the receiver (in a catchable point)

Example of Assessment Score Sheet

School Code:……………………………………………… Teacher Code: ………………………………………………

Scorer: ……………………………………………… Class:……………… Date:………………… Period:……………

Task 4: Basketball Dribbling and Passing

Student's Serial Number	Dribbling	Passing
1.	0 1 2 3	0 1 2 3
2.	0 1 2 3	0 1 2 3
3.	0 1 2 3	0 1 2 3
4.	0 1 2 3	0 1 2 3
5.	0 1 2 3	0 1 2 3
6.	0 1 2 3	0 1 2 3
7.	0 1 2 3	0 1 2 3
8.	0 1 2 3	0 1 2 3
9.	0 1 2 3	0 1 2 3
10.	0 1 2 3	0 1 2 3
11.	0 1 2 3	0 1 2 3
12.	0 1 2 3	0 1 2 3
13.	0 1 2 3	0 1 2 3
14.	0 1 2 3	0 1 2 3
15.	0 1 2 3	0 1 2 3
16.	0 1 2 3	0 1 2 3
17.	0 1 2 3	0 1 2 3
18.	0 1 2 3	0 1 2 3
19.	0 1 2 3	0 1 2 3
20.	0 1 2 3	0 1 2 3
21.	0 1 2 3	0 1 2 3
22.	0 1 2 3	0 1 2 3
23.	0 1 2 3	0 1 2 3
24.	0 1 2 3	0 1 2 3
25.	0 1 2 3	0 1 2 3

Appendix B: The Dynamic Model of Educational Effectiveness Observational Instruments
(Low- and High-Inference Forms)

DMEE LOW-INFERENCE OBSERVATION FORM

Observer:... Teacher Code:... .... School Code: .... Date:..................... Period:...........................

Class:............. Number of students:........ Subject:... Lesson: ....

(1) ORIENTATION																					
DIMENSIONS	**Instructions for coding**																				
Sequence of the activity	Ordinal number of the activity as observed during the lesson.																				
Duration	Duration in minutes.																				
Focus	*Relation with:* 1. an aim of the lesson 2. the day lesson 3. the unit/number of lessons.																				
Quality	1. typical 2. related to learning 3. students specify the aim(s).																				
Differentiation	Put down the sign √ for any type of differentiation you observe.																				

(2) STRUCTURING																					
DIMENSIONS	**Instructions for coding**																				
Sequence of the activity	Ordinal number of the activity as observed during the lesson.																				
Duration	Duration in minutes.																				
Focus	*Relation with:* 1. previous lessons 2. structure of the day lesson 3. the unit/number of lessons.																				
Quality-Clarity	1. clear for the students 2. not clear for the students																				
Differentiation	Put down the sign √ for any type of differentiation you observe.																				

(3) APPLICATION																					
DIMENSIONS	**Instructions for coding**																				
Sequence of the activity	Ordinal number of the activity as observed during the lesson.																				
Duration	Duration in minutes.																				
Focus	*Relation with:* 1. only a part of the lesson 2. the whole lesson 3. the unit/a number of lessons.																				
Quality	1. use of the same activity to find a specific result, 2. activation of certain cognitive processes for the solution of more complex activities.																				
Differentiation	Put down the sign √ for any type of differentiation you observe.																				

(4) TEACHING MODELLING-NEW LEARNING																				
DIMENSIONS	**Instructions for coding**																			
Sequence of the activity	Ordinal number of the activity as observed during the lesson.																			
Duration	Duration in minutes.																			
Focus	1. can be used only in the lesson 2. can be used in the unit 3. can be used across units.																			
Quality: teacher 's role	1. given by the teacher 2. guided discovery 3. discovery																			
Quality: appropria-teness of the model	1. successful. 2. not successful.																			
Quality: lesson's stage observed	1. After a problematic situation. 2. Before a problematic situation.																			
Differentiation	Put down the sign √ for any type of differentiation you observe.																			

(5) QUESTIONING TECHNIQUES

DIMENSIONS	Instructions for coding																			
Sequence of the activity	Ordinal number of the activity as observed during the lesson.																			
Waiting time	Time given before answering																			
Focus	*<u>Relation with</u>*: 1. only a specific task 2. the whole lesson 3. the unit/a number of lessons.																			
Quality: type of question	1. product 2. process.																			
Quality: reaction if no answer from pupils (in case there is an answer put an X).	1. restates (easier words) 2. poses an easier question 3. moves to another question or answers the question him/herself.																			
Quality: type of feedback/reaction to students' answers	1. negative comments to incorrect and partly correct answers. 2. positive comments to correct answer only. 3. positive comments to correct answer and constructive comments to incorrect and to partly correct answers. 4. no comments.																			
Quality: reaction about the answer	1. teacher ignores the answer. 2. teacher indicates that the answer is correct or partly correct or incorrect. 3. students are invited to give comments on the answer.																			
Differentiation:	Put down the sign √ for any type of differentiation you observe.																			

DMEE HIGH-INFERENCE OBSERVATION FORM

Observer's Name: ....

Teacher's Name: ....

School: .... Date:.................... Time: ....

Class: .... Number of Students:......... Subject: ....

Lesson:..

DIRECTIONS: Use the scale to note the extent to which you agree with the following statements. (<u>*Scale:*</u> *1:Minimum point ….. 5: Maximum point*).

	STATEMENT	MINIMUM POINT				MAXIMUM POINT
1.	The orientation activities that were organized during the lesson helped students understand the new content.	1	2	3	4	5
2.	The teacher explained the structure of the lesson in a way that was clear for the pupils.	1	2	3	4	5
3.	The lesson transited from easier to more complex activities.	1	2	3	4	5
4.	The observed application activities referred (were linked) to the whole lesson.	1	2	3	4	5
5.	The observed application activities referred (were linked) to certain parts of the lesson.	1	2	3	4	5
6.	The observed application activities referred (were linked) to previous lessons as well.	1	2	3	4	5
7.	The application activities were nothing else but a replication of the activities that were organized during the presentation of the new content.	1	2	3	4	5
8.	The teacher spent the teaching time on learning activities.	1	2	3	4	5
9.	During the lesson, the teacher gave only to some pupils the opportunity to participate in the lesson.	1	2	3	4	5
10.	The teacher was interacting with pupils for the whole of the lesson.	1	2	3	4	5
11.	During the lesson, some pupils were co-operating with each other while others did not.	1	2	3	4	5
12.	Pupils interacted with each other during the whole of the lesson.	1	2	3	4	5
13.	Interaction between pupils contributed in achieving the lessons goals.	1	2	3	4	5
14.	The teacher discouraged the negative aspects of competition.	1	2	3	4	5

15.	There was pupil misbehavior in the form of verbal intimidation during the lesson.	1	2	3	4	5
16.	There was pupil misbehavior in the form of serious verbal intimidation during the lesson.	1	2	3	4	5
17.	There was pupil misbehavior in the form of bodily intimidation without putting others in danger during the lesson.	1	2	3	4	5
18.	There was pupil misbehavior in the form of bodily intimidation putting others in danger during the lesson.	1	2	3	4	5
19.	The lesson was interrupted by the misbehavior of some pupils.	1	2	3	4	5
20.	In the case of misbehavior in the classroom, the teacher spent enough teaching time to deal with it.	1	2	3	4	5
21.	The teacher was forced to make remarks to some students because they were talking to each other.	1	2	3	4	5
22.	In the case of misbehavior in the classroom, the teacher ignored it deliberately.	1	2	3	4	5
23.	In the case of misbehavior in the classroom, the teacher reacted and temporarily solved the problem.	1	2	3	4	5
24.	In the case of misbehavior in the classroom, the teacher reacted and managed to solve the problem.	1	2	3	4	5
25.	In the case of misbehavior in the classroom, the teacher reacted but did not manage to solve the problem.	1	2	3	4	5
26.	The lesson was interrupted by external factors.	1	2	3	4	5
27.	The aims that the teacher had set before the lesson were met during the 40-minute period of the lesson.	1	2	3	4	5
28.	The activities that were organized during the lesson helped each pupil to advance conceptually, according to his/her abilities.	1	2	3	4	5
29.	The majority of pupils were engaged in activities that were provided by their teacher.	1	2	3	4	5
30.	During the lesson the majority of the pupils were on task.	1	2	3	4	5
31.	Less able pupils considered the lesson activities as very difficult.	1	2	3	4	5
32.	More able pupils considered the lesson activities as very easy.	1	2	3	4	5
33.	The teacher used to pose questions that were clear for the pupils in terms of their content.	1	2	3	4	5
34.	The teacher used to correct pupils' misconceptions using their wrong answers.	1	2	3	4	5
35.	Pupils were puzzled by the procedures or strategies that the teacher presented to them for overcoming problematic situations.	1	2	3	4	5
36.	The teacher expanded on students' thinking, to help them discover a procedure or strategy for overcoming a problematic situation.	1	2	3	4	5
37.	The procedures or strategies that the teacher presented to the pupils to help them overcome the problematic situations they faced can be used in other lessons as well.	1	2	3	4	5

38.	The teacher used to explain the procedures and strategies to the pupils and then she/he requested using them.	1	2	3	4	5
39.	Pupils understood the procedures and strategies that were presented by the teacher.	1	2	3	4	5
40.	Pupils used on their own initiative, ways or strategies presented by the teacher, to solve similar problems.	1	2	3	4	5

If you have any further comments, please use the space provided below:

....

....

....

....

....

....

Thank you for your assistance

Appendix C: The Modified Task Structure System Observational Instruments
(Low- and High-Inference Forms)

mTSS LOW-INFERENCE OBSERVATION FORM [a]

Observation Sheet

DATE: ………………… OBSERVER: ……………………………………… LESSON'S SUBJECT: ……………………

UNIT: …………………………. NUM. OF LESSON WITHIN UNIT: …….. FROM ……….. NUM.OF STUDENTS: ……

EPISODE 1		COMMENTS
TIME		
EPISODE TYPE		
TASK TYPE		
SKILL DEMONSTRATION		CONQRUENCY & QUALITY OF RESPONSE
TASK PRESENTATION		ST:
EXPLICITNESS		M+:
ACCOUNTABILITY		M-:
STUDENT CODE		OT:

EPISODE 2		COMMENTS
TIME		
EPISODE TYPE		
TASK TYPE		
SKILL DEMONSTRATION		CONQRUENCY & QUALITY OF RESPONSE
TASK PRESENTATION		ST:
EXPLICITNESS		M+:
ACCOUNTABILITY		M-:
STUDENT CODE		OT:

EPISODE 3		COMMENTS
TIME		
EPISODE TYPE		
TASK TYPE		
SKILL DEMONSTRATION		CONQRUENCY & QUALITY OF RESPONSE
TASK PRESENTATION		ST:
EXPLICITNESS		M+:
ACCOUNTABILITY		M-:
STUDENT CODE		OT:

EPISODE 4		COMMENTS
TIME		
EPISODE TYPE		
TASK TYPE		
SKILL DEMONSTRATION		CONQRUENCY & QUALITY OF RESPONSE
TASK PRESENTATION		ST:
EXPLICITNESS		M+:
ACCOUNTABILITY		M-:
STUDENT CODE		OT:

CONGRUENCY OF STUDENT'S RESPONSE
1. Stated (S) 2. Modified Up (T+)
3. Modified Down (T-) 4. Off Task (OT)

EPISODE TYPE
1. Management (M)
2. Transition (T)
3. Waiting (W)
4. Warm up-Recovery
   4a) Relevant to lesson's objective (R1)
   4b) Irrelevant to lesson's objective (R2)
5. Instructional (I)

TASK TYPE
1. Cognitive (C)
2. Informing (I)
3. Refining-Qualitative improvement (R)
4. Extending-Differ.Conditions-Complexity (E)
5. Applying-Game-Performance conditions
   5a) Relevant to lesson's objective (A1)
   5b) Irrelevant to lesson's objective (A2)

SKILL DEMONSTRATION
1. Did not occur (NO)
2. Verbally
   2a) Partial/Brief reference to the critical cues (VP)
   2b) Full reference to the critical cues (VF)
3. Practically
   3a) Partial/Brief practical demonstration (PP)
   3b) Full practical demonstration (PF)
4. Combination
   4a) Partial-Brief demonstration (CP)
   4b) Full Demonstration (CF)

TASK PRESENTATION
1. YES (Y) 2. NO (N)

TASK EXPLICITNESS
1. Implicit (I)
2. Partially explicit (PE)
3. Fully explicit (FE)

ACCOUNTABILITY
1. No supervision (NS)
2. Monitoring (M)
3. Reinforcement (R)
4. Specific&Congruent Feedback (F)
5. Reinforcement&Spec.&Congr. Feed. (RF)

STUDENT CODE
1. Male – High Level (M1)
2. Female – Low Level (F2)
3. Male – Middle Level (M3)
4. Female – High Level (F4)
5. Male – Low Level (M5)
6. Female – Middle Level (F6)

QUALITY OF STUDENT'S RESPONSE
1. Appropriate (A)
2. Inappropriate (I)

Note. [a] The mTSS low-inference form consists of eight identical sheets, each including four episodes. Here only the first sheet is presented. The teacher's and school's codes are filled in on the last sheet.

mTSS HIGH-INFERENCE OBSERVATION INSTRUMENT

Observer: ....

Teacher Code:...

School Code: .... Date:.............................. Time: ....

Class: .... Number of students: .... Unit: ....

Focus of the Lesson: ....

Instructions: Circle the appropriate number of the scale situated next to each statement, to indicate the degree to which the incidents described by the statements below occurred in the lesson you observed.

	Statement	**DID NOT OCCUR**	**OCCURRED VERY LITTLE**	**OCCURRED MODERATELY**	**OCCURRED TO A GREAT EXTENT**
1	The description of a task was accompanied by a visual presentation from the teacher or a student, to help students understand what exactly is required from them to do.	0	1	2	3
2	The teacher or a mediating agent (e.g., a student) practically demonstrated the correct form of the lesson's main psychomotor skill(s)/strategy(ies).	0	1	2	3
3	The teacher verbally presented the key learning cues of the lesson's main psychomotor skill(s)/strategy(ies) in a simple and accurate manner.	0	1	2	3
4	The teacher used efficient ways to place, distribute or collect the equipment needed for the lesson.	0	1	2	3
5	The teacher used effective routines to assign students to pairs/groups.	0	1	2	3
6	Teacher's clear and complete directions minimized the transitioning time from one task to the next.	0	1	2	3
7	Students were not waiting in lines during practice.	0	1	2	3
8	The teacher efficiently solved any disciplinary problems.	0	1	2	3
9	The teacher held students accountable for performing to the specified performance level[39].	0	1	2	3

[39] For this statement, the scale is used as follows: If for the most of the lesson: 0 = the teacher did not monitor or supervise students' involvement; 1 = the teacher was watching students to practice; 2 = the teacher reinforced students' practice; 3 = the teacher provided specific and congruent feedback to each practicing student.

	Statement	DID NOT OCCUR	OCCURRED VERY LITTLE	OCCURRED MODERATELY	OCCURRED TO A GREAT EXTENT
10	After students' practice, the teacher was giving specific and congruent (in terms of the learning cues) feedback to the whole class.	0	1	2	3
11	The instructional tasks were developmentally appropriate for the students' skill level.	0	1	2	3
12	The instructional tasks gave students the opportunity to actively participate in the lesson (both the organization and structure of the activities—e.g. the number of students in each group--as well as students' waiting time in lines should be considered).	0	1	2	3
13	The instructional tasks were coherent and helped students develop the lesson's main psychomotor skill(s)/strategy(ies).	0	1	2	3
14	The instructional tasks were pertinent to the psychomotor goal(s) of the lesson.	0	1	2	3
15	There was a variety of instructional tasks (informing, refining, extending, and applying).	0	1	2	3
16	The instructional tasks had a clear progression.	0	1	2	3
17	The congruency of students' practice trials was on stated task.	0	1	2	3
18	The quality of students' practice trials was appropriate.	0	1	2	3
19	The opportunities offered to students to practice lesson's main psychomotor skill(s)/strategy(ies) were adequate.	0	1	2	3

Appendix D: Student Survey

Student Questionnaire

Name:………………………………………………… Surname:………………………………………………………

School: ………

Class:…………………………………………………………… Date: ……………………………………………………………

Dear student,

This questionnaire consists of two parts and involves questions about you and your family (Part A′) and how often do you do some things in your physical education lessons (Part B′). You have 40 minutes to complete this questionnaire. Please, answer all the questions as accurately and honestly as possible, as there is no correct or wrong answer.

Thank you for your help!

PART A′

In this part you should put a √ in the "Yes" or "No" boxes, to show what applies for your family and yourself.

Question 1: For each statement below, put a √ in the "Yes" box, if you have the described things at your home. If you do not have them, put a √ in the "No" column.

		Yes	No
1.	One car		
2.	Two or more cars		
3.	Your own bedroom		
4.	Your own books (**apart from** the school books)		
5.	Musical Instrument/s **apart from flute** (e.g., piano, violin, guitar)		
6.	PC (computer, laptop, tablet)		
7.	Internet		
8.	Your own mobile phone		

Question 2: Sometimes, some parents have people helping them in house chores. In your house, what applies? *(For each statement put a* √ in either the "Yes" or "No" column*).*

		Yes	No
1.	A maid usually comes at my house		
2.	A maid lives in my home		
3.	A gardener usually comes at my house to care the garden		
4.	The house chores are done only by the members of my family		

Question 3: This month **your mother:** ***(Put only one √ in the boxes of this question)***

Works	
Does not work, but she is looking for a job	
Does not work (e.g., she takes care of the family)	
None of the above	

If your mother works, please write her job: ______________________________

Question 4: This month **your father:**
(Put only one √ in the boxes of this question)

Works	
Does not work, but he is looking for a job	
Does not work (e.g., he takes care of the family)	
None of the above	

If your father works, please write his job: ______________________________

Question 5: Put **as many √ as applicable** in mother's and father's columns to show what applies for your parents education.

	Mother	**Father**
She/he has graduated from gymnasium		
She/he has graduated from technical school		
She/he has graduated from lyceum		
She/he has graduated from university/college		
She/he has a master's or a doctoral degree		

Question 6: Put a √ in each statement below to show where were you and your parents born.

	Cyprus	**Greece**	**Other Country**
I was born			
My mother was born			
My father was born			

Question 7: What language do you speak at home? *(Put a √ only in one box)*

Greek	
Other language	

If you have chosen "Other language", please write what language do you speak: ____________

Question 8: Do you participate in out-of-school athletic activities/sports (e.g., soccer, swimming, dance, gymnastics etc)? (*Please circle one answer)*

1. Yes 2. No

If you have answered "Yes", please answer the question 9. Otherwise, move on to the Part B.

Question 9: Circle the out-of-school activities that you participate:

- Soccer or Futsal
- Basketball
- Tennis
- Handball
- Volleyball
- Swimming
- Dance (ballet, Latin)
- Traditional dances
- Track and field (e.g., sprints, long jump)
- Gymnastics or Rhythmic gymnastics (e.g., ribbon, tumbling)
- Martial Arts (e.g., karate, taekwondo)

- Other (write the activity): ………………………………………………

PART B′

This part asks you to indicate how often you do the following things in your Physical Education lessons.

	How often do you do the following things in your **Physical Education lessons?**	**Never**	**A few lessons**	**About half the lessons**	**Many lessons**	**Every lesson**
1	At the beginning of the lesson we are doing warm-up activities (e.g. jogging, stretching etc.).	1	2	3	4	5
2	We discuss how we can apply the skills/knowledge we are learning in the lesson to our daily lives.	1	2	3	4	5
3	Before we start practicing a motor skill (e.g., dribble), my teacher or a fellow-student demonstrates how to perform this skill.	1	2	3	4	5
4	At the beginning of the lesson we are connecting what we have learned in previous lessons with what we will learn in the day's lesson.	1	2	3	4	5
5	The equipment (e.g., balls) to be used during the lesson is set up, before the lesson starts.	1	2	3	4	5
6	When my teacher asks a question that we do not understand, then he/she restates the question with simpler words, to help us understand what he/she was asking.	1	2	3	4	5
7	We are doing/playing the activities/games that we like/prefer for the majority of the lesson's time. (e.g., we are playing soccer).	1	2	3	4	5
8	We discuss about the optimal way of performing correctly a given skill (e.g., dribble).	1	2	3	4	5
9	While we are performing a skill, my teacher tells to each student how to perform correctly the practicing skill.	1	2	3	4	5
10	We are performing the emphasized skill by doing several, different activities.	1	2	3	4	5
11	We are standing in lines and we are practicing one-by-one.	1	2	3	4	5
12	The teacher interrupts the lesson to scold ill-disciplined students.	1	2	3	4	5
13	Before we start practicing a motor skill, we discuss the learning cues that we have to attend in order to perform the skill appropriately (e.g., when dribbling, we push the ball with our finger tips instead of hitting it with the palm).	1	2	3	4	5
14	We have ample time to exercise the emphasized performance skills.	1	2	3	4	5
15	When my teacher explains the activities, he/she tells us what we should achieve to be successful (e.g., when we are dribbling in basketball, my teacher tells us that we have to dribble for a whole minute without losing ball control).	1	2	3	4	5
16	We are doing activities in which we are practicing only a technical feature of the skill (e.g., when we are dribbling in basketball, my teacher asks us to give emphasis only in pushing the ball with our upper finger pads).	1	2	3	4	5
17	We discuss why it is important for us to be engaged in certain activities.	1	2	3	4	5
18	Before engaging in an activity, my teacher or a fellow-student presents what is expected from us to do.	1	2	3	4	5
19	At the end of the lesson, we sum up the main points of the lesson.	1	2	3	4	5

	How often do you do the following things in your **Physical Education lessons?**	**Never**	**A few lessons**	**About half the lessons**	**Many lessons**	**Every lesson**
20	At the beginning of the lesson we play a game during which we use the skill that we are going to perform/learn in the lesson.	1	2	3	4	5
21	When my teacher asks questions, he/she let us enough time to thing and come up with an answer before answer the questions.	1	2	3	4	5
22	We have the opportunity to perform a lot of trials for the emphasized skill.	1	2	3	4	5
23	We discuss about strategies that we can use in different games (e.g., to exchange many passes in a basketball game, the unmarked player needs to move in the free space).	1	2	3	4	5
24	We are divided into groups and each group is practicing a different skill (e.g., a group is practicing in tennis, while the other two groups are playing a basketball game).	1	2	3	4	5
25	We are practicing first in some easy activities and then we are going to difficult ones.	1	2	3	4	5
26	When we are working in groups or pairs, we form these groups quickly with losing time.	1	2	3	4	5
27	When we have to collaborate among ourselves, there are students who don't do so.	1	2	3	4	5
28	The majority of my fellow-students perform successfully the lesson's activities.	1	2	3	4	5
29	The Physical Education lesson starts on time, without losing any time.	1	2	3	4	5
30	We are doing competitive activities/games during which we have to apply the emphasized skill (e.g., during the lesson that we have learned the dribble in basketball, we had played a modified basketball game).	1	2	3	4	5
31	During a lesson, we are practicing in activities with different level of difficulty; in some easy and some difficult activities.	1	2	3	4	5
32	We are losing our Physical Education lesson to do a rehearsal for an upcoming event or to do other activities or lessons.	1	2	3	4	5
33	When we are learning a new skill, we are practicing first in an easy activity.	1	2	3	4	5
34	The majority of my fellow-students perform the emphasized skills as have been stated by the teacher.	1	2	3	4	5
35	My teacher keeps reminding us of the important learning cues that we have to keep in mind in order to perform a skill accurately.	1	2	3	4	5
36	When we are practicing a skill, my teacher tells to the students who are not successfully performing the skill, what they should focus on to correctly perform the skill.	1	2	3	4	5

Thank you for your effort in completing this questionnaire!!!

References

Alison, S. & Thorpe, R. (1997). A comparison of the effectiveness of two approaches to teaching games within physical education. A skills approach versus a games for understanding approach. *British Journal of Physical Education, 28*(3), 9-13.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system — Secondary. *School Psychology Review, 42*, 76–97.

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75. doi: 10.3102/0013189X08316420.

Anderson, W. G., & Barrette, G. T. (1978). Teacher behavior. In W. G. Anderson & G. T. Barrette (Eds.), What's going on in gym: Descriptive studies of physical education classes (pp. 25-38). Newtown, CT: Motor Skills: Theory into Practice.

Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, *1*(4), 363-378. doi: 10.1207/s15324818ame0104_7.

Ashford, D., Bennett, S. J., & Davids, K. (2006). Observational modeling effects for movement dynamics and movement outcome measures across differing task constraints: A meta-analysis. *Journal of Motor Behavior, 38*(3), 185-205. doi: 10.3200/JMBR.38.3.185-205.

Ashy, M., Lee, A., & Landin, D. (1988). Relationship of practice using correct technique to achievement in a motor skill. *Journal of Teaching in Physical Education, 7*(2), 115-120. doi: 10.1123/jtpe.7.2.115.

Ayvazo, S., & Ward, P. (2011). Pedagogical content knowledge of experienced teachers in physical education: Functional analysis of adaptations. *Research Quarterly for Exercise and Sport, 82*(4), 675-684. doi: 10.1080/02701367.2011.10599804.

Bailey, R. (2006). Physical education and sport in schools: A review of benefits and outcomes. *Journal of School Health, 76*(8), 397-401. doi: 10.1111/j.1746-1561.2006.00132.x.

Bailey, R., Armour, K., Kirk, D., Jess, M., Pickup, I., Sandford, R., and the BERA Physical Education and Sport Pedagogy Special Interest Group. (2009). The educational benefits claimed for physical education and school sport: An academic review. *Research Papers in Education*, *24*(1), 1-27. doi: 10.1080/02671520701809817.

Barnett, L. M., van Beurden, E., Morgan, P. J., Brooks, L. O., & Beard, J. R. (2010). Gender differences in motor skill proficiency from childhood to adolescence. *Research Quarterly for Exercise and Sport, 81*(2), 162-170. doi: 10.1080/02701367.2010.10599663.

Belka, D. (2002). A strategy for improvement of learning-task presentations. *Journal of Physical Education, Recreation and Dance, 73*(6), 32-35. doi: 10.1080/07303084.2002.10607825.

Bentler, P. M. (1995). *EQS: Structural equations program manual.* California: Multivariate software inc.

Bevans, K. B., Fitzpatrick, L., Sanchez, B. M., Riley, A. W., & Forrest, C. (2010). Physical education resources, class management, and student physical activity levels: A structure-process-outcome approach to evaluating physical education effectiveness. *The Journal of School Health*, *80*(12), 573-580. doi:10.1111/j.1746-1561.2010.00544.x.

Bond, T. G. & Fox, C. M. (2012). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York, NY: Routledge.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

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

Buck, M., Harrison, J., & Bryce, G. (1991). An analysis of learning trials and their relationship to achievement in volleyball. *Journal of Teaching in Physical Education, 10*(2), 134-152. doi: 10.1123/jtpe.10.2.134.

Burns, R., & Burns, R. (2008). Discriminant Analysis. In *Business Research Methods and Statistics using SPSS* (Chapter 25). Retrieved from https://studysites.uk.sagepub.com/burns/website%20material/Chapter%2025%20-%20Discriminant%20Analysis.pdf

Burzycka-Wilk, D. (2010). Effectiveness of visual information in the process of teaching swimming motor activities. *Human Movement*, *11*(2), 184-190. doi: 10.2478/v 10038-010-0019-x.

Butterfield , S. A., Angell, R. M., & Mason, C. A. (2012). Age and sex differences in object control skills by children ages 5 to 14. *Perceptual and Motor Skills, 114*(1), 261-274. doi: 10.2466/10.11.25.PMS.114.1.261-274.

Cantrell, S., & Kane, T. J. (2013). *Ensuring fair and reliable measures of effective teaching*. Seattle: Bill & Melinda Gates Foundation. Retrieved from Measures of Effective Teaching Project website: http://www.metproject.org/reports.php

Castelli, D. & Rink, J. (2003). A comparison of high and low performing secondary physical education programs. *Journal of Teaching in Physical Education, 22*(5), 512-532. doi: 10.1123/jtpe.22.5.512.

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. doi: 10.3102/0013189X018001026.

Charalambous, C. Y., Komitis, A., Papacharalambous, M., Stefanou, A. (2014). Using generic and content-specific teaching practices in teacher evaluation: An exploratory study of teachers' perceptions. *Teaching and Teacher Education*, *41*, 22-33. doi: 10.1016/j.tate.2014.03.001.

Chatoupis, C., & Vagenas, G. (2011). An analysis of published process-product research on physical education teaching methods. *International Journal of Applied Sports Sciences, 23*(1), 271-289.

Chatzipanteli, A. & Digelidis, N. (2011). The influence of metacognitive prompting on students' performance in motor skills test in physical education. *International Journal of Sports Science and Engineering, 5*(2), 93-98,

Chatzipanteli, A., Digelidis, N., Karatzoglidis, C., & Dean, R. (2014). A tactical-game approach and enhancement of metacognitive behavior in elementary school students. *Physical Education and Sport Pedagogy,* (ahead-of-print), 1-16. doi: 10.1080/17408989.2014.931366.

Cheffers, J. T. F., & Mancini, V. H. (1989). Cheffers' Adaptation of Flanders' Interaction Analysis System (CAFIAS). In P.W. Darst, D. B. Zakrajsek, & V. H. Mancini (Eds). Analyzing physical education and sportinstruction (2nd ed., pp. 119-135). Champaign, IL: Human Kinetics.

Chen, W., Hammond-Bennett, A., Upton, A., & Mason, S. (2014). Accomplished teachers' implementation of quality teaching practices. *The Physical Educator, 71*(2), 320-343.

Chen, W., Hendricks, K., & Archibald, K. (2011). Assessing pre-service teachers' quality teaching practices. *Educational Research and Evaluation, 17*(1), 13–32. doi: 10.1080/13803611.2011.578008.

Chen, W., Hypnar, A. J., Mason, S. A., Zlamout, S. Hammond-Bennett, A. (2014). Students' daily physical activity behaviors: The role of quality physical education in a comprehensive school physical activity program. *Journal of Teaching in Physical Education,33*(4), 592-610. doi: 10.1123/jtpe.2014-0060.

Chen, W., Mason, S., Hammond-Bennett, A., & Zlamout, S. (2014). Development and validation of assessing quality teaching rubrics. *Journal of Education and Training Studies, 2*(3), 52-61. doi: 10.11114/jets.v2i3.409.

Chen, W., Mason, S., Staniszewski, C., Upton, A., & Valley, M. (2012). Assessing the quality of teachers' teaching practices. *Educational Assessment, Evaluation and Accountability, 24*, 25–41. doi: 10.1007/s11092-011-9134-2.

Chen, W., Zhu, W., Mason, S., Hammond-Bennett, A., & Colombo-Dougovito, A. (2016). Effectiveness of quality physical education in improving students' manipulative skill competency. *Journal of Sport and Health Science, 5*(2), 231-238. doi: 10.1016/j.jshs.2015.04.005.

Christodoulou, D. (2010). The status of physical education in Cypriot schools. *Physical Culture and Sport Studies and Research, 50*(1), 110-117. doi: 10.2478/v10141-010-0028-7.

Ni Chroinin, D., & Cosgrave, C. (2013). Implementing formative assessment in primary physical education: teacher perspectives and experiences. *Physical Education and Sport Pedagogy, 18*(2), 219-233. doi: 10.1080/17408989.2012.666787.

Cochran-Smith, M. (2003). Teaching quality matters. *Journal of Teacher Education, 54*(2), 95-98. doi: 10.1177/0022487102250283.

Coe, R., Aloisi, C., Higgins, S., & Major, L. E. (2014). *What makes great teaching? Review of the underpinning research.* London: The Sutton Trust. Retrieved from http://www.suttontrust.com/wp-content/uploads/2014/10/What-Makes-Great-Teaching-REPORT.pdf

Cohen, D. K. (2011). *Teaching and its predicaments*. Cambridge, MA: Harvard University Press.

Cohen, D., Raudenbush, S., & Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119-142. doi: 10.3102/01623737025002119.

Constantinides, P., Montalvo, R., & Silverman, S. (2013). Teaching processes in elementary physical education classes taught by specialists and nonspecialists. *Teaching and Teacher Education*, *36*, 68-76. doi: 10.1016/j.tate.2013.07.004.

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. doi: 10.1080/09243450902850176.

Core Practice Consortium (2014, April). *Enriching research and innovation through the specification of professional practice.* Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA.

Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools.* London: Routledge.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). *Methodological advances in educational effectiveness research*. London: Taylor & Francis.

Daniels, S. R. (2006). The consequences of childhood overweight and obesity. *The Future of Children*, *16*(1), 47-67. doi: 10.1353/foc.2006.0004.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson Group. (2013). *The Framework for Teaching evaluation instrument* (*FfT*). Princeton, NJ: Author.

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. doi: 10.1076/sesi.15.1.3.27490.

DeCorby, K., Halas, J., Dixon, S., Wintrup, L. & Janzen, H. (2005). Classroom teachers and the challenges of delivering quality physical education. *The Journal of Educational Research*, *98*(4), 208-220. doi: 10.3200/JOER.98.4.208-221.

De Knop, P. (1986). Relationship of specified instructional teacher behaviors to student gain on tennis. *Journal of Teaching in Physical Education*, *5*(2), 71-78. doi: 10.1123/jtpe.5.2.71.

Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher*, *38*(7), 518-521. doi:10.3102/0013189X09350881.

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 392-431). New York, NY: Macmillan.

Dunkin, M., & Biddle, B. (1974). *The study of teaching*. New York, NY: Holt Rinehart & Winston.

Dyson, B., Linehan, N., & Hastie, P. (2010). The ecology of cooperative learning in elementary physical education classes. *Journal of Teaching in Physical Education*, *29*(2), 113-130. doi: 10.1123/jtpe.29.2.113.

Dyson, B. & Strachan, K. (2004). The ecology of cooperative learning in a high school physical education programme. *Waikato Journal of Education*, *10*, 117-139.

Ennis, C. (2006). Curriculum: Forming and reshaping the vision of physical education in a high need, low demand world of schools. *Quest*, *58*(1), 41-59. doi: 10.1080/00336297.2006.10491871.

Ennis, C. D. (2014). The role of students and content in teacher effectiveness. *Research Quarterly for Exercise and Sport*, *85*(1), 6-13. doi: 10.1080/02701367.2014.872979.

Ennis, C. D., & Chen, A. (2011). Learning motor skill in physical education. In M. E. Richard, & A. A. Patricia (Eds.), *Handbook of research on learning and instruction* (pp. 148-165). New York, NY: Routledge.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1-9. doi:10.1016/j.learninstruc.2013.07.001.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record*, *107*(1), 186-213. doi: 10.1111/j.1467-9620.2005.00462.x.

Fink, J., & Siedentop, D. (1989). The development of routines, rules, and expectations at the start of the school year. *Journal of Teaching in Physical Education*, *8*(3), 198-212. doi: 10.1123/jtpe.8.3.198.

Fletcher, T., & Mandigo, J. (2012). The primary schoolteacher and physical education: A review of research and implications for Irish physical education. *Irish Education Studies*, *31*(3), 363-376. doi: 10.1080/03323315.2012.710063.

Fogo, B. (2014). Core practices for teaching history: The results of a delphi panel survey. *Theory & Research in Social Education*, *42*(2), 151-196. doi: 10.1080/00933104.2014.902781.

Follman, J. (1995). Elementary public school pupil rating of teacher effectiveness. *Child Study Journal*, *25*(1), 57-78.

French, K., Rink, J., Rikard, L., Mays, A., Lynn, S., & Werner, P. (1991). The effects of practice progressions on learning two volleyball skills. *Journal of Teaching in Physical Education, 10*(3), 261-274. doi: 10.1123/jtpe.10.3.261.

Gallahue, D. L. & Donnelly, F. C. (2003). *Developmental physical education for all children* (4th ed.). Leeds: Human Kinetics.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1), 134-150. doi: 10.1162/rest.89.1.134.

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

Graber, K. C., Locke, L. F., Lambdin, D. & Solmon, M. A. (2008). The landscape of elementary school physical education. *The Elementary School Journal*, *108*(3), 151-159. doi: 10.1086/529098.

Graham, G., & Holt/Hale, S. A., & Parker, M. (2013). *Children moving: A reflective approach to teaching physical education* (9th ed.). Boston: McGraw-Hill Higher Education.

Graham, G., & Heimerer, E. (1981). Research on teacher effectiveness: A summary with implications for teaching. *Quest*, *33*(1), 14-25. doi: 10.1080/00336297.1981.10483718.

Green, K. (2012). Mission impossible? Reflecting upon the relationship between physical education, youth sport and lifelong participation. *Sport, Education and Society*, *19*(4), 357-375. doi: 10.1080/13573322.2012.683781.

Griggs, G. (2007). Physical education: Primary matters, secondary importance. *Education 3-13: International Journal of Primary, Elementary and Early Years Education, 35(1), 59-69. doi:* 10.1080/03004270601103350.

Griggs, G. (2012). Surveying the landscape of primary physical education. In G. Griggs (Ed.), *An introduction to primary physical education* (pp. 3-12). New York, NY: Routledge.

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores* (National Bureau of Economic Research Working Paper No. 16015). Cambridge, MA: National Bureau of Economic Research.

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*, *119*(3), 445-470. doi: 10.1086/669901.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*(1), 184-205. doi: 10.3102/0002831207312906.

Gubacs-Collin, K. (2007). Implementing a tactical approach through action research. *Physical Education and Sport Pedagogy, 12*(2), 105-126. doi: 10.1080/17408980701281987.

Gusthart, L., Kelly, I., & Rink, J. (1997). The validity of the Qualitative Measures of Teaching Performance Scale as a measure of teacher effectiveness. *Journal of Teaching in Physical Education, 16*(2), 196-210. doi: 10.1123/jtpe.16.2.196.

Gusthart, L., & Sprigings, E. (1989). Student learning as a measure of teacher effectiveness in physical education. *Journal of Teaching in Physical Education, 8*(4), 298-311. doi: 10.1123/jtpe.8.4.298.

Hambleton, R. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R., & Rogers, J. (1991). Advances in criterion-referenced measurement. In R. Hambleton, & J. Zaal (Eds.), *Advances in educational and psychological testing* (pp. 3-44). New York: Springer Science+Business Media.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., … Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*(4), 461-487. doi: 10.1086/669616.

Hardman, K. (2008). The situation of physical education in schools: A European perspective. *Human Movement, 9*(1), 5-18. doi: 10.2478/v10038-008-0001-z.

Harrison, J. M., Blakemore, C. L., Richards, R. P., Oliver, J., Wilkinson, C., & Fellingham, G. (2004). The effects of two instructional models-tactical and skill teaching-on skill development and game play, knowledge, self-efficacy, and student-perceptions in volleyball. *Physical Educator, 61*(4), 186-199.

Hastie, P. A. (1994). Selected teacher behaviors and student ALT-PE in secondary school physical education classes. *Journal of Teaching in Physical Education, 13*(3), 242-259. doi: 10.1123/jtpe.13.3.242.

Hastie, P. A. (2000). An ecological analysis of sport education season. *Journal of Teaching in Physical Education, 19*(3), 355-373. doi: 10.1123/jtpe.19.3.355.

Hastie, P. A., Calderon, A., Palao, J., & Ortega, E. (2011). Quantity and quality of practice. *Research Quarterly for Exercise and Sport, 82*(4), 784-787. doi: 10.1080/02701367.2011.10599815.

Hastie, P., & Siedentop, D. (1999). An ecological perspective on physical education. *European Physical Education Review, 5*(1), 9-29. doi: 10.1177/1356336X990051002.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hay, P. J. (2006). Assessment for learning in physical education. In D. Kirk, D. MacDonald, & M. O'Sullivan (Eds.), *The handbook of physical education* (pp. 312-325). London: Sage Publications.

Hebert, E., Landin, D., & Solmon, M. (2000). The impact of task progressions on students' practice quality and task-related thoughts. *Journal of Teaching in Physical Education, 19*(3), 338-354. doi: 10.1123/jtpe.19.3.338.

Hickson, C., & Fishburne, G. (2004). *What is effective physical education teaching and can it be promoted with generalist trained elementary school teachers?* Paper presented at the AARE conference, Melbourne. Retrieved from: http://www.aare.edu.au/data/publications/2004/hic04158.pdf

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511. doi: 10.1080/07370000802177235.

Hodges, N. J., & Franks, I. M. (2002). Modelling coaching practice: the role of instruction and demonstration. *Journal of Sports Sciences, 20*(10), 793-811. doi: 10.1080/026404102320675648.

Horn, R. R., Williams, A. M., Hayes, S. J., Hodges, N. J., & Scott, M. A. (2007). Demonstration as a rate enhancer to changes in coordination during early skill acquisition. *Journal of Sport Sciences, 25*(5), 599-614. doi: 10.1080/02640410600947165.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equations Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi: 10.1080/10705519909540118.

Hunt, G., Wiseman, D., & Touzel, T. (2009). *Effective teaching: Preparation and implementation* (4th ed.). Illinois: Charles Thomas Publisher.

Hushman, G., Hushman, C. & Carbonneau, K. (2015). Addressing educational reform: Exploring PE metrics as a system to measure student achievement in physical education. *The Physical Educator, 72*, 179-191.

Iserbyt, P., Ward, P., & Li, W. (2015). Effects of improved content-knowledge on pedagogical content-knowledge and student performance in PE. *Physical Education and Sport Pedagogy*. doi: 10.1080/17408989.2015.1095868.

Janelle, C. M., Champenoy, J. D., Coombes, S. A., & Mousseau, M. B. (2003). Mechanisms of attentional cueing during observational learning to facilitate motor skill acquisition. *Journal of Sports Sciences, 21*(10), 825-838. doi: 10.1080/0264041031000140310.

Jess, M., Dewar, K., & Fraser, G. (2004). Basic moves: Developing a foundation for lifelong physical activity. *The British Journal of Teaching Physical Education, 35*(2), 24-27.

Joe, J. N., Tocci, C. M., Holtzman, S. L., & Williams, J. C. (2013). *Foundations of observation: Considerations for developing a classroom observation system that helps districts achieve consistent and accurate scores*. Princeton, NJ: Bill & Melinda Gates Foundation. Retrieved from Measures of Effective Teaching Project website: http://www.metproject.org/downloads/MET-ETS_Foundations_of_Observation.pdf

Jones, D. L. (1992). Analysis of task systems in elementary physical education classes. *Journal of Teaching in Physical Education, 11*(4), 411-425. doi: 10.1123/jtpe.11.4.411.

Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with students surveys and achievement gains.* Seattle: Bill & Melinda Gates Foundation. Retrieved from Measures of Effective Teaching Project website: http://www.metproject.org/reports.php.

Kelder, S. H., Mitchell, P. D., McKenzie, T. L., Derby, C., Strikmiller, P. K., Luepker, R. V., & Stone, E. J. (2003). Long-term implementation of the CATCH physical education program. *Health Education & Behavior, 30*(4), 463-475. doi: 10.1177/1090198103253538.

Kennedy, M. M. (2010). Approaches to annual performance assessment. In M. M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 225-250). San Francisco, CA: Jossey-Bass.

Kirk, D. (2005). Physical education, youth sport and lifelong participation: The importance of early learning experiences. *European Physical Education Review*, *11*(3), 239-255. doi: 10.1177/1356336X05056649.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.

Kloser, M. (2014). Identifying a core set of science teaching practices: A delphi expert panel approach. *Journal of Research in Science Teaching, 51*(9), 1185-1217. doi: 10.1002/tea.21171.

Kniffin, K. M., & Baert, H. (2015). Maximizing learning through assessment in middle and high school physical education. *Journal of Physical Education, Recreation & Dance, 86*(4), 7-16. doi: 10.1080/07303084.2015.1009203.

Ko, B. (2008). *An examination of teaching practices of elementary physical educators* (Unpublished doctoral dissertation). Ohio State University, Ohio.

Konstantopoulos, S. (2012). Teacher effects: Past, present and future. In S. Kelly (Ed.), *Assessing teacher quality: Understanding teacher effects on instruction and achievement* (pp. 33-48). New York, NY: Teachers College Press.

Krause, J.V., Meyer, D., & Meyer, J. (2008). Basketball skills and drills (3rd ed.). Champaign, IL: Human Kinetics.

Kulinna, P. H., Scrabis-Fletcher, K., Kodish, S., Phillips, S., & Silverman, S. (2009). A decade of research literature in physical education pedagogy. *Journal of Teaching in Physical Education, 28*(2), 119-140. doi: 10.1123/jtpe.28.2.119.

Kwak, E. C. (2005). The immediate effects of various task presentation types on middle school students' skill learning. *International Journal of Applied Sports Sciences, 17*(1), 7-17.

Kyriakides, E. (2011). *Challenges affecting delivery of quality physical education in Cyprus primary schools* (Unpublished master's thesis). Leeds Metropolitan University, Leeds.

Kyriakides, L. (2005). Extending the Comprehensive Model of Educational Effectiveness by an empirical investigation. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 16*(2), 103-152. doi: 10.1080/09243450500113936.

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, 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. doi: 10.1016/j.tate.2013.07.010.

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. doi: 10.1076/sesi.11.4.501.3560.

Kyriakides, L., & Creemers, B. P. M. (2008a). 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: An International Journal of Research, Policy and Practice, 19*(2), 183-205. doi: 10.1080/09243450802047873.

Kyriakides, L., & Creemers, B.P.M. (2008b). A longitudinal study on the stability over time of school and teacher effects on student learning outcomes. *Oxford Review of Education, 34*(5), 521-545. doi: 10.1080/03054980701782064.

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. doi: 10.1080/19415530903043680.

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. doi: 10.1016/j.tate.2008.06.001.

Kyriakides, L., Creemers, B. P. M., 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. doi: 10.1080/01411920903165603.

Kyriakides, L., & Tsangaridou, N. (2008). Towards the development of generic and differentiated models of educational effectiveness: A study on school and teacher effectiveness in physical education. *British Educational Research Journal, 34*(6), 807-838. doi: 10.1080/01411920802041467.

La Paro, K.M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, *104*(5), 409-426. doi: 10.1086/499760.

Landin, D. (1994). The role of verbal cues in skill learning. *Quest*, *46*(3), 299-313. doi: 10.1080/00336297.1994.10484128.

Lee, A., Keh, N., & Magill, R. (1993). Instructional effects of teacher feedback in physical education. *Journal of Teaching in Physical Education, 12*(3), 228-243. doi: 10.1123/jtpe.12.3.228.

Li, W. (2015). Strategies for creating a caring learning climate in physical education. *Journal of Physical Education, Recreation & Dance, 86*(4), 34-41. doi: 10.1080/07303084.2015.1009204.

Li, W., & Lee, A. (2004). A review of conceptions of ability and related motivational constructs in achievement motivation. *Quest, 56*(4), 439-461. doi: 10.1080/00336297.2004.10491836.

Lidor, R. (2004). Developing metacognitive behavior in physical education classes: The use of task-pertinent learning strategies. *Physical Education and Sport Pedagogy, 9*(1), 55-71. doi: 10.1080/1740898042000124.

Lindsay, E. L. (2014). Effective teaching in physical education: The view from a variety of trenches. *Research Quarterly for Exercise and Sport, 85*(1), 31-37. doi: 10.1080/02701367.2014.873330.

Little, O., Goe, L., & Bell, C. (2009, April). *A practical guide to evaluating teacher effectiveness*. Retrieved from http://files.eric.ed.gov/fulltext/ED543776.pdf

Luke, D. A. (2004). *Multi-level modeling: Quantitative applications in the social sciences*. Newbury Park: Sage Publications.

MacPhail, A., & Halbert, J. (2010). 'We had to do intelligent thinking during recent PE': Students' and teachers' experiences of assessment for learning in post-primary physical education. *Assessment in Education: Principles, Policy & Practice, 17*(1), 23–39. doi: 10.1080/09695940903565412.

MacNamara, A., Collins, D., Bailey, R., Toms, M., Ford, P., & Pearce, G. (2011). Promoting lifelong physical activity and high level performance: Realising an achievable aim for physical education. *Physical Education and Sport Pedagogy, 16*(3), 265-278. doi: 10.1080/17408989.2010.535200.

Magill, R. A. (2010). *Motor learning and control: Concepts and applications.* (9th ed.). New York, NY: McGraw-Hill.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., … Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732-749. doi: 10.1111/j.1467-8624.2008.01154.x.

Masser, L. (1987). The effect of refinement on student achievement in a fundamental motor skill in grades K through 6. *Journal of Teaching in Physical Education*, *6*(2), 174-182. doi: 10.1123/jtpe.6.2.174.

Masser, L. (1993). Critical cues help first-grade students' achievement in handstands and forward rolls. *Journal of Teaching in Physical Education*, *12*(3), 301-312. doi: 10.1123/jtpe.12.3.301.

Matanin, M, & Tannehill, D. (2004). Assessment and grading in physical education. *Journal of Teaching in Physical Education, 13*(4), 395-405. doi: 10.1123/jtpe.13.4.395.

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale". *Educational Assessment*, *13*(4), 267-300. doi: 10.1080/10627190802602541.

Matsumura, L. C., Slater, S.C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006, May). *Measuring reading comprehension and mathematics instruction in urban* *middle schools: A pilot study of the Instructional Quality Assessment*. Retrieved from https://www.cse.ucla.edu/products/reports/R681.pdf

McKenzie, T. L. (2007). The preparation of physical educators: A public health perspective. *Quest*, *59*(4), 345-357. doi: 10.1080/00336297.2007.10483557.

McKenzie, T. L., Alcaraz, J. E., Sallis, J. M., & Faucette, F. N. (1998). Effects of a physical education program on children's manipulative skills. *Journal of Teaching in Physical Education, 17*(3), 327-341.

McKenzie, T. L., & van der Mars, H. (2015). Top 10 research questions related to assessing activity and its context using systematic observation, *Research Quarterly for Exercise and Sport, 86*(1), 13-29. doi: 10.1080/02701367.2015.991264.

McNeil, M. C., Fry, J. M., Wright, S. C., Tan, C. W. K., & Rossi, T. (2008). Structuring time and questioning to achieve tactical awareness in games lessons. *Physical Education and Sport Pedagogy, 13*(3), 231-249. doi: 10.1080/17408980701345766.

Measures of Effective Teaching [MET] (2012). *Asking students about teaching: Student perception surveys and their implementation.* Retrieved from Measures of Effective Teaching Project website: http://www.metproject.org/downloads/Asking_Students_Practitioner_Brief.pdf

Mercier, K. & Doolittle, S. (2013). Assessing student achievement in physical education for teacher evaluation. *Journal of Physical Education, Recreation and Dance, 84*(3), 38-42. doi: 10.1080/07303084.2013.767721.

Metzler, M. (2011). *Instructional models for physical education* (3rd ed.). Boston: Allyn and Bacon.

Metzler, M. (2014). Teacher effectiveness research in physical education: The future isn't what it used to be. *Research Quarterly for Exercise and Sport*, *85*(1), 14-19. doi: 10.1080/02701367.2014.872932.

Ministry of Education and Culture (2010). *Physical education curriculum*. Retrieved from: http://archeia.moec.gov.cy/mc/2/fysiki_agogi.pdf

Ministry of Education and Culture (2013). *Outline of the content of physical education curriculum*. Retrieved from: http://archeia.moec.gov.cy/mc/1/dee_dim_fysiki_agogi.pdf

Mintah, J. K. (2003). Authentic assessment in physical education: Prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science*, *7*(3), 161-174. doi: 10.1207/S15327841MPEE0703_03.

Morgan, P., & Hansen, V. (2007). Recommendations to improve primary school physical education: Classroom teachers' perspective. *The Journal of Educational Research, 101*(2), 99-111. doi: 10.3200/JOER.101.2.99-112.

Morgan, P., & Hansen, V. (2008a). Classroom teachers' perceptions of the impact of barriers to teaching physical education on the quality of physical education programs. *Research Quarterly for Exercise and Sport*, *79*(4), 506-516. doi: 10.1080/02701367.2008.10599517.

Morgan, P., & Hansen, V. (2008b). The relationship between PE biographies and PE teaching practices of classroom teachers. *Sport, Education and Society*, *13*(4), 373-391. doi: 10.1080/13573320802444994.

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. doi: 10.1080/09243453.2014.885451.

Muijs, D., & Reynolds, D. (2001). *Effective teaching: Evidence and practice.* London: Paul Chapman Publishing.

Muijs, D., & Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics. *Educational Research and Evaluation, 9*(3), 289–314. doi: 10.1076/edre.9.3.289.15571.

National Association for Sport and Physical Education (NASPE). (2010). *PE metrics: Assessing national standards 1-6 in elementary school.* Reston, VA: NASPE.

Nevett, M., Rovegno, I., & Babiarz, M. (2001). Fourth-grade children's knowledge of cutting, passing and tactics in invasion games after a 12-lesson unit of instruction. *Journal of Teaching in Physical Education, 20*(4), 389-401. doi: 10.1123/jtpe.20.4.389.

Nevett, M., Rovegno, I., Babiarz, M., McCaughtry, N. (2001). Changes in basic tactics and motor skills in an invasion-type game after a 12-lesson unit of instruction. *Journal of Teaching in Physical Education, 20*(4), 352-369. doi: 10.1123/jtpe.20.4.352.

Nye, S., Dubay, C., Gilbert, L., & Wajciechowski, M. (2009). The importance of assessment within physical education. *Virginia Association for Health, Physical Education, and Dance, 30*(1), 8-9.

Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237-257. doi: 10.3102/01623737026003237.

Oslin, J. L., Stroot, S., & Siedentop, D. (1997). Use of component-specific instruction to promote development of the overarm throw. *Journal of Teaching in Physical Education, 16*(3), 340-356. doi: 10.1123/jtpe.16.3.340.

O' Sullivan, M. (2002). Good data are all you have: The legacy of a PETE scholar. *Journal of Teaching in Physical Education, 21*(4), 441-453. doi: 10.1123/jtpe.21.4.441.

O' Sullivan, M., & Dyson, B. (1994). Rules, routines and expectations of 11 high school physical education teachers. *Journal of Teaching in Physical Education, 13*(4), 361-374. doi: 10.1123/jtpe.13.4.361.

Panayiotou, A., Kyriakides, L., Creemers, B. P. M., McMahon, L., Vanlaar, G., Pfeifer, M., … Bren. M. (2014). Teacher behavior and student outcomes: Results of a European study. *Educational Assessment, Evaluation and Accountability, 26*(1), 73-93. doi: 10.1007/s11092-013-9182-x.

Patton, K., Parker, M., & Tannehill, D. (2015). Helping teachers help themselves: professional development that makes a difference. *National Association of Secondary School Principals* (*NASSP) Bulletin*, *99*(1), 26-42. doi: 10.1177/0192636515576040.

Paye, B., & Paye, P. (2013). *Youth basketball drills* (2nd ed.). Champaign, IL: Human Kinetics.

Pearson, P. J., & Webb, P. (2008). *Developing effective questioning in Teaching Games for Understanding (TGfU)*. Paper presented at the 1st Asia Pacific Sport in Education conference, Adelaide.

Pellett, T. L., & Harrison, J. M. (1995a). The influence of refinement on female junior high school students' volleyball practice success and achievement. *Journal of Teaching in Physical Education*, *15*(1), 41-52. doi: 10.1123/jtpe.15.1.41.

Pellett, T. L., & Harrison, J. M. (1995b). The influence of a teacher's specific, congruent, and corrective feedback on female junior high school students' immediate volleyball practice success. *Journal of Teaching in Physical Education*, *15*(1), 53-63. doi: 10.1123/jtpe.15.1.53.

Peterson, K. (2000). *Teacher evaluation: A comprehensive guide to new directions and practice.* Thousand Oaks, CA: Corwin Press.

Phillips, D. A., & Carlisle, C. (1983). A comparison of physical education teachers categorized as most and least effective. *Journal of Teaching in Physical Education, 2*(3), 55-67. doi: 10.1123/jtpe.2.3.55.

Pianta, R., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109-119. doi: 10.3102/0013189X09332374.

Pickup, I. (2012). The importance of primary physical education. In G. Griggs (Ed.), *An introduction to primary physical education* (pp. 13-24). New York, NY: Routledge.

Portman, P. (1995). Who is having fun in physical education classes? Experiences of sxth-grade students in elementary and middle schools. *Journal of Teaching in Physical Education, 14*(4), 445-453. doi: 10.1123/jtpe.14.4.445.

Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2012). *A user's guide to MLwiN: Version 2.26.* Bristol, UK: Centre for Multilevel Modeling, University of Bristol.

Ram, N., Riggs, S. M., Skaling, S., Landers, D. M., & McCullagh, P. (2007). A comparison of modeling and imagery in the acquisition and retention of motor skills. *Journal of Sport Sciences, 25*(5), 587-597. doi: 10.1080/02640410600947132.

Rasmussen, J. F., Scrabis-Fletcher, K., & Silverman, S. (2014). Relationships among tasks, time, and student practice in elementary physical education. *The Physical Educator*, *71*, 114-131.

Rikard, G. L. (1991). The short term relationship of teacher feedback and student practice. *Journal of Teaching in Physical Education, 10*(3), 275-285. doi: 10.1123/jtpe.10.3.275.

Rikard, G. L. (1992). The relationship of teacher task refinement and feedback to student practice success. *Journal of Teaching in Physical Education, 11*(4), 349-357. doi: 10.1123/jtpe.11.4.349.

Rink, J. E. (1999). *What do students learn in physical education and how do they learn?* Lecture presented at the International AIESEP Congress, Besancon. Retrieved from: http://ruc.udc.es/bitstream/2183/10997/1/CC%2059%20art%2016.pdf

Rink, J. E. (2003). Effective instruction in physical education. In S. J. Silverman, & C. D. Ennis (Eds.), *Student Learning in Physical Education* (2nd ed., pp. 165-186). Champaign, IL: Human Kinetics.

Rink, J. E. (2009). Necessary but not sufficient. In L. Housner, M. Metzler, P. Schempp, & T. Templin (Eds.), *Historic traditions and future directions of research on teaching and teacher education in physical education* (pp. 23-31). West Virginia: West Virginia University.

Rink, J. E. (2010). *Teaching physical education for learning* (6th ed.). New York, NY: McGraw-Hill.

Rink, J. E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport, 84*(4), 407-418. doi: 10.1080/02701367.2013.844018.

Rink, J. E. (2014). Teacher effectiveness in physical education-Consensus? *Research Quarterly for Exercise and Sport, 85*(3), 282-286. doi: 10.1080/02701367.2014.932656.

Rink, J. E., French, K., Werner, P., Lynn, S., & Mays, A. (1991). The influence of content development on the effectiveness of instruction. *Journal of Teaching in Physical Education, 11*(2), 139-149. doi: 10.1123/jtpe.11.2.139.

Rink, J. E., & Hall, T. J. (2008). Research on effective teaching in elementary school physical education. *The Elementary School Journal, 108*(3), 207-218. doi: 10.1086/529103.

Rink, J. E., & Werner, P. (1989). Qualitative Measures of Teaching Performance Scale. In P. Dart, D. Zakrajsek, & P. Mancini (Eds.), *Analyzing physical and sport instruction* (2nd ed., pp. 269-276). Champaign, IL: Human Kinetics.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417-458. doi: 10.1111/j.1468-0262.2005.00584.x.

Roberts, G. (2014). Time to demonstrate how we impact student learning. *Research Quarterly for Exercise and Sport*, *85*(1), 27–30. doi: 10.1080/02701367.2013.872490.

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, *94*(2), 247-252. doi: 10.1257/0002828041302244.

Rosenshine, B. V. (1983). Teaching functions in instructional programs. *Elementary School Journal, 83*(4), 335-351. doi: 10.1086/461321.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York, NY: Simon & Schuster.

Rovegno, I., & Bandhauer, D. (2016). *Elementary physical education: Curriculum and instruction* (2nd ed.). Burlington, MA: Jones & Bartlett Learning.

Rovegno, I., & Dolly, J. P. (2006). Constructivist perspectives on learning. In D. Kirk, D. MacDonald, & M. O'Sullivan (Eds.), *The Handbook of physical education* (pp. 242-261). London: Sage Publications.

Safrit, M. J., Cohen, A. S., & Costa, M. G. (1989). Item response theory and the measurement of motor behavior. *Research Quarterly for Exercise and Sport, 60*(4), 325-335. doi: 10.1080/02701367.1989.10607459.

Sariscsany, M. J., Darst, P. W., & van der Mars, H. (1995). The effects of three teacher supervision patterns on student on-task and skill performance in secondary physical education. *Journal of Teaching in Physical Education, 14*(2), 179-197. doi: 10.1123/jtpe.14.2.179.

Schacter, J., & Thum, Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review, 23*(4), 411-430. doi: 10.1016/j.econedurev.2003.08.002.

Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement, 24*(1), 1-38. doi: 10.1080/09243453.2012.691100.

Scheerens, J., & Bosker, R.J. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon.

Schuldheisz, J. M., & van der Mars, H. (2001). Active supervision and students' physical activity in middle school physical education. *Journal of Teaching in Physical Education, 21*(1), 75-90. doi: 10.1123/jtpe.21.1.75.

Schwab, J. J. (1978). Education and the structure of the disciplines. In: I. Westbury & N. J. Wilfolk (Eds.) *Science, curriculum, and liberal education: Selected essays* (pp. 229-272). Chicago: The University of Chicago Press.

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. doi: 10.3102/0034654307310317.

Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences*. (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. doi: 10.3102/0013189X015002004.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-21. doi: 10.17763/haer.57.1.j463w79r56455411.

Sidaway, B., Fairweather, M., Powell, J., & Hall, G. (1992). The acquisition and retention of a timing task: Effects of summary KR and movement time. *Research Quarterly for Exercise and Sport, 63*(3), 328-334. doi: 10.1080/02701367.1992.10608750.

Siedentop, D. (2002). Ecological perspectives in teaching research. *Journal of Teaching in Physical Education, 21*(4), 427-440. doi: 10.1123/jtpe.21.4.427.

Siedentop, D. (2009). Research on teaching physical education: Celebrating our past and focusing on our future. In L. Housner, M. Metzler, P. Schempp, & T. Templin (Eds.), *Historic traditions and future directions of research on teaching and teacher education in physical education* (pp. 3-14). West Virginia: West Virginia University.

Siedentop, D., Doutis, P., Tsangaridou, N., Ward, P., & Rauschenbach, J. (1994). Don‘t sweat gym: An analysis of curriculum and instruction. *Journal of Teaching in Physical Education, 13*(4), 375-394. doi: 10.1123/jtpe.13.4.375.

Siedentop, D., & Tannehill, D. (2000). *Developing teaching skills in physical education* (4th ed.). Mountain View, CA: Mayfield.

Siedentop, D., Tousignant, M., & Parker, M. (1982). Academic learning time—physical education: coding manual. Columbus, OH: The Ohio State University.

Silverman, S. (1985). Relationship of engagement and practice trials to student achievement. *Journal of Teaching in Physical Education, 5*(1), 13-21. doi: 10.1123/jtpe.5.1.13.

Silverman, S. (1990). Linear and curvelinear relationships between student practice and achievement in physical education. *Teaching and Teacher Education, 6*(4), 305-314. doi: 10.1016/0742-051X(90)90023-X.

Silverman, S. (1993). Student characteristics, practice and achievement in physical education. *The Journal of Educational Research, 87*(1), 54-61. doi: 10.1080/00220671.1993.9941166.

Silverman, S. (1994). Communication and motor skill learning: What we learn from research in the gymnasium. *Quest, 46*(3), 345-355. doi: 10.1080/00336297.1994.10484131.

Silverman, S. (2009). Learning in physical education: The complex nature of relationships in the gymnasium. In L. Housner, M. Metzler, P. Schempp, & T. Templin (Eds.), *Historic traditions and future directions of research on teaching and teacher education in physical education* (pp. 41-49). West Virginia: West Virginia University.

Silverman, S. (2011). Teaching for student learning in physical education. *Journal of Physical Education, Recreation, & Dance, 82*(6), 29-34. doi: 10.1080/07303084.2011.10598642.

Silverman, S., & Ennis, C. (2003). Enhancing learning: An introduction. In S., Silverman & C. D. Ennis (Eds.), *Student learning in physical education: Applying research to enhance instruction* (2nd ed., pp.3-7). Champaign, IL: Human Kinetics.

Silverman, S., Kulinna, P. H., & Crull, G. (1995). Skill-related task structures, explicitness, and accountability: relationships with student achievement. *Research Quarterly for Exercise and Sport, 66*(1), 32-40. doi: 10.1080/02701367.1995.10607653.

Silverman, S., & Mercier, K. (2015). Teaching for physical literacy: Implications to instructional design and PETE. *Journal of Sport and Health Science, 4*(2), 150-155. doi: 10.1016/j.jshs.2015.03.003.

Silverman, S., & Skonie, R. (1997). Research on teaching in physical education: An analysis of published research. *Journal of Teaching in Physical Education, 16*(3), 300-311. doi: 10.1123/jtpe.16.3.300.

Silverman, S., Subramaniam, P. R., & Woods, A. M. (1998). Task structures, student practice, and student skill level in physical education. *Journal of Educational Research, 91*(5), 298-307. doi: 10.1080/00220679809597557.

Silverman, S., Tyson, L., & Krampitz, J. (1992). Teacher feedback and achievement in physical education: interaction with student practice. *Teaching and Teacher Education, 8*(4), 333-344. doi: 10.1016/0742-051X(92)90060-G.

Silverman, S., Tyson, L. S., & Morford, L. M. (1988). Relationships of organization, time, and student achievement in physical education. *Teaching and Teacher Education, 4*(3), 247-257. doi: 10.1016/0742-051X(88)90004-2.

Silverman, S., Woods, A. M., & Subramaniam, P. R. (1999). Feedback and practice in physical education: Interrelationships with task structures and student skill level. *Journal of Human Movement Studies, 36*, 203-224.

Simons-Morton, B. G., Taylor, W. C., Snider, S. A., & Huang, I. W. (1993). The physical activity of the fifth-grade students during physical education classes. *American Journal of Public Health, 83*(2), 262-264. doi: 10.2105/AJPH.83.2.262.

Singh, R., & Sarkar, S. (2015). Does teaching quality matter? Students learning outcome related to teaching quality in public and private primary schools in India. *International Journal of Educational Development, 41*, 153-163. doi: 10.1016/j.ijedudev.2015.02.009.

Smith, B. O. (1983). Some comments on educational research in the twentieth century. *Elementary School Journal, 83*(4), 488-492. doi: 10.1086/461328.

Standage, M., Duda, J., & Ntoumanis, N. (2006). Students' motivational processes and their relationship to the teacher ratings in school physical education: A self-determination theory approach. *Research Quarterly for Exercise and Sport, 77*(1), 100-110. doi: 10.5641/027013606X13080769704046.

Stodden, D. F., Goodway, J. D., Langendorfer, S. J., Roberton, M. A., Rudisill, M. E., Garcia, C., & Garcia, L. E. (2008). A developmental perspective on the role of motor skill competence in physical activity: An emergent relationship. *Quest, 60*(2), 290-306. doi: 10.1080/00336297.2008.10483582.

Stodolsky,S. S. (1988). *The subject matters: Classroom activity in Math and Social studies*. Illinois: University of Chicago Press.

Stodolsky, S. S., & Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subjects. *American Educational Research Journal, 32*(2), 227-249. doi: 10.3102/00028312032002227.

Strong, M. (2011). *The highly qualified teacher: What is teacher quality and how do we measure it?* New York, NY: Teachers College Press.

Tannehill, D., & Lund, J. (2010). Building a quality physical education program. In J. Lund, & D. Tannehill (Eds.), *Standards-based physical education curriculum development* (pp. 23-55). Sudbury, MA: Jones and Bartlett Publishers.

Tannehill, D., van der Mars, H., & MacPhail, A. (2015). *Building effective physical education programs*. Burlington, MA: Jones & Bartlett learning.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer Press.

Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2015). *Research methods in physical activity* (7th ed.). Champaign, IL: Human Kinetics.

Thorburn, M., & Seatter, K. (2015). Asking better questions! A review of the pedagogical strategies used in one senior level award in Scotland. *Journal of Pedagogy*, *6*(1), 123-149. doi: 10.1515/jped-2015-0007.

Tinning, R. (2005). Physical education in Australia: an interpretative account. In: U. Puhse & M. Gerber (Eds), *International comparison of physical education: concepts, problems, prospects* (pp. 50-65). Oxford: Meyer & Meyer Sport.

Tousignant, M., & Siedentop, D. (1983). A qualitative analysis of task structures in required secondary physical education classes. *Journal of Teaching in Physical Education, 3*(1), 47-57. doi: 10.1123/jtpe.3.1.47.

Trudeau, F., Laurencelle, L., Tremblay, J., Rajic, M., & Shephard, R. (1999). Daily primary school physical education: effects on physical activity during adult life. *Medicine and Science in Sports and Exercise, 31*(1), 111-117. doi: 10.1097/00005768-199901000-00018.

Tsangaridou, N. (2006). Teachers' knowledge. In: D. Kirk, D. Macdonald, & M. O'Sullivan, (Eds.), *Handbook of Research in Physical Education* (pp. 502-515). London: Sage Publications.

Tsangaridou, N., & Yiallourides, G. (2008). Physical education and education through sport in Cyprus. In G. Klein, & K. Hardman, (Eds.), *Physical Education and Sport Education in European Union* (pp. 69-83). Paris: Editions Revue EP.S.

Turner, A. P., & Martinek, T. J. (1999). An investigation into Teaching Games for Understanding: Effects on skill, knowledge, and game play. *Research Quarterly for Exercise and Sport, 70*(3), 286-296. doi: 10.1080/02701367.1999.10608047.

Valentini, N. (2004). Visual cues, verbal cues and child development. *Strategies: A Journal for Physical and Sport Educators*, *17*(3), 21-23. doi: 10.1080/08924562.2004.10591081.

van der Mars, H. (2006). Time and learning in physical education. In D. Kirk, D. MacDonald, & M. O'Sullivan (Eds.), *The Handbook of physical education* (pp. 191-213). London: Sage Publications.

Veal, M. L., & Anderson, W. G. (2011). *Analysis of teaching and learning in physical education.* Sudbury, MA: Jones and Bartlett Learning.

Venetsanou, F., & Kambas, A. (2010). Environmental factors affecting preschoolers' motor development. *Early Childhood Education Journal, 37*(4), 319-327. doi: 10.1007/s10643-009-0350-z.

Walling, M. D., & Martinek, T. J. (1995). Learned helplessness: A case study of a middle school student. *Journal of Teaching in Physical Education, 14*(4), 454-466. doi: 10.1123/jtpe.14.4.454.

Wang, Y., & Lobstein, T. (2006). Worldwide trends in childhood overweight and obesity. *International Journal of Pediatric Obesity, 1*(1), 11-25. doi: 10.1080/17477160600586747.

Ward, P. (2009). Content matters: Knowledge that alters teaching. In L. Housner, M. Metzler, P. Schempp, & T. Templin (Eds.), *Historic traditions and future directions of research on teaching and teacher education in physical education* (pp. 345-356). West Virginia: West Virginia University.

Ward, P. (2013). The role of content knowledge in conceptions of teaching effectiveness in physical education. *Research Quarterly for Exercise and Sport, 84*(4), 431-440. doi: 10.1080/02701367.2013.844045.

Ward, P., Kim, I., Ko, B., & Li, W. (2015). Effects of improving teachers' content knowledge on teaching and student learning in physical education. *Research Quarterly for Exercise and Sport, 86*(2), 130-139. doi: 10.1080/02701367.2014.987908.

Werner, P., & Rink, J. (1989). Case studies of teacher effectiveness in second grade physical education. *Journal of Teaching in Physical Education, 8*(4), 280-297. doi: 10.1123/jtpe.8.4.280.

Whipp, P., Taggart, A., & Jackson, B. (2014). Differentiation in outcome-focused physical education: pedagogical rhetoric and reality. *Physical Education and Sport Pedagogy, 19*(4), 370-382. doi: 0.1080/17408989.2012.754001.

Wood, J., Tocci, C. M., Joe, J. N., Holtzman, S. L., Cantrell, S., & Archer, J. (2014). *Building trust in observations: A blueprint for improving systems to support great teaching*. Princeton, NJ: Bill & Melinda Gates Foundation. Retrieved from Measures of Effective Teaching Project website: http://www.metproject.org/downloads/MET_Observation_Blueprint.pdf

World Health Organization (WHO). (2012). *Population-based approaches to childhood obesity prevention*. Geneva: World Health Organization. Retrieved from World Health Organisation website: http://www.who.int/dietphysicalactivity/SPF-en-2008.pdf.

Wright, M. T., & van der Mars, H. (2004). Blending assessment into instruction: Practical applications and meaningful results. *Journal of Physical Education, Recreation, & Dance, 75*(9), 29-34. doi: 10.1080/07303084.2004.10607297.

Wright, S., Horn, S., & Sanders, W. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57-67.

Yiallourides, G. (1998). *Factors influencing the attitudes of 9 to 14 year old Cypriot pupils towards Physical Education* (Unpublished doctoral dissertation). University of Manchester, Manchester.

Zeng, H. Z., Leung, R., Liu, W., & Hipscher, M. (2009). Physical education in urban high school class settings: Features and correlations between teaching behaviors and learning activities. *Physical Educator, 66*(4), 186-196.

Zhu, W., Fox, C., Park. Y., Fisette, J. L., Dyson, B., Graber, K. C., . . . Raynes, D. (2011). Development and calibration of an item bank for PE metrics assessments: Standard 1. *Measurement in Physical Education and Exercise Science*, *15*(2), 119-137. doi: 10.1080/1091367X.2011.568366.