University of the Peloponnese
School of Economics and Technology
Department of Informatics and Telecommunications

ΓΑΒ LAB - Knowledge and Uncertainty Research Laboratory

Doctoral Thesis

Data analysis as a decision support tool in Greek education

Ilias K. Papadogiannis
St id: 287308785515

Supervisor:

Manolis Wallace
Associate Professor

Tripolis, October, 2023

Consulting Committee:

1. Manolis Wallace, Associate Professor
2. Costas Vassilakis, Professor
3. Georgios Lepouras, Professor

.

Approved by the selection board on 12-10-2023.

1. Manolis Wallace, Associate Professor
2. Costas Vassilakis, Professor
3. Georgios Lepouras, Professor
4. Vassilis Poulopoulos, Assistant Professor
5. Nikos Platis, Assistant Professor
6. Anastasios Theodoropoulos, Assistant Professor
7. Athanassios Katsis, Professor

....

...(Ilias K.Papadogiannis)
Department of Informatics and Telecommunications
University of Peloponnese

This work has been developed in collaboration with the Research Team Knowledge and Uncertainty(ΓAB LAB).

Acknowledgements

With this small text, I want to express my gratitude to all who played a crucial role in shaping and guiding my PhD journey. Despite the challenges and uncertainties, I am proud to have successfully crossed the finish line.

First and foremost, I express my sincere thanks to my supervisor, Associate Professor Manolis Wallace. His unwavering support, guidance, and encouragement have been instrumental in my research endeavor. I am truly indebted to his guidance, expertise and dedication. I would also like to extend my warm appreciation to Professors Costas Vassilakis and Georgios Lepouras for their valuable insights and constructive comments, which significantly contributed to the enhancement of my work. From the initial stages of my research to the final submission of my thesis, the unwavering presence and profound knowledge of the three-member consulting committee were invaluable in shaping my academic growth.

I would also like to express my gratitude to Assistant Professor Vassilis Poulopoulos for his invaluable support throughout this arduous journey and to Assistant Professor Nikos Platis for his assistance and improvements to the papers that accompany my thesis. Furthermore, I express my appreciation to Assistant Professor Anastasios Theodoropoulos for his support, as well as to Rector Athanassios Katsis for his valuable time dedicated to my dissertation.

Οφείλω να εκφράσω τη βαθιά ευγνωμοσύνη μου στην οικογένειά μου. Η πίστη τους στις ικανότητές μου και η υποστήριξή τους ήταν τα κλειδιά για αυτό το επίτευγμα. Θα ήθελα να ευχαριστήσω τη Δήμητρα για τους ρόλους που κλήθηκε να παίξει, προκειμένου να μπορέσω να συνεχίσω αυτή την προσπάθεια.

Πάνω απ' όλα, θέλω να αφιερώσω αυτή τη διατριβή στα παιδιά μου, τον Κώστα και τη Γεωργία, γνωρίζοντας ότι πλέον καταλαβαίνουν τι σημαίνει να κυνηγάς τα όνειρά σου, έστω και αργά.

Σας ευχαριστώ για όλα.

Σας αφιερώνω αυτή τη διατριβή.

Contents

List of Figures

List of Tables

Περίληψη

Η παρούσα διδακτορική διατριβή αναδεικνύει τις δυνατότητες που προσφέρει η Ανά λυση Εκπαιδευτικών Δεδομένων στην πρωτοβάθμια και δευτεροβάθμια εκπαίδευση. Ξεκινώντας από ένα κοινό ζήτημα στην εκπαιδευτική έρευνα - την ακαδημαϊκή επίδοση των μαθητών - δίνει έμφαση στις δυνατότητες αντικειμενικής αξιολόγησης και υποστήριξης αποφάσεων, που παρέχει η χρήση δεδομένων στα κεντρικά οργανωμένα εκπαιδευτικά συστήματα. Κεντρικό ρόλο διαδραματίζει η έννοια της εκπαίδευσης των ίσων ευκαιριών, η οποία διαπερνά ολόκληρη τη διατριβή.

Από την εποχή της βιομηχανικής επανάστασης, η εκπαίδευση θεωρήθηκε ως εργαλείο κοινωνικής κινητικότητας και παροχής ίσων ευκαιριών. Τα κράτη καθιέρωσαν την υποχρεωτική δημόσια εκπαίδευση και την χρηματοδότησαν μέσω της φορολογίας. Βασικός στόχος παραμένει η ισότητα των ευκαιριών στην εκπαίδευση για τους μαθητές, μέσω της δημόσιας παροχής και του κρατικού ελέγχου. Η έρευνα όμως έχει αναδείξει ότι οι διαχρονικές κοινωνικές ανισότητες αναπαράγονται και μέσω των εκπαιδευτικών συστημάτων, οδηγώντας σε προτάσεις για πιο συμπεριληπτικά εκπαιδευτικά συστήματα και εκπαιδευτικές παρεμβάσεις.

Η αξιολόγηση της αποτελεσματικότητας των εκπαιδευτικών συστημάτων και η έκφραση απόψεων στο δημόσιο διάλογο συχνότατα αντικατοπτρίζει προσωπικές θεωρήσεις, οι οποίες δεν στηρίζονται σε αντικειμενικά στοιχεία. Πρόσφατα στην Ελλάδα έγινε δυνατή η συστηματική συλλογή εκπαιδευτικών δεδομένων για τους εμπλεκόμενους στη εκπαιδευτική πολιτική, με την εισαγωγή ενός MIS για την πρωτοβάθμια και την δευτεροβάθμια εκπαίδευση, αλλά οι δυνατότητες εξόρυξης γνώσης από αυτό δεν έχουν ακόμη αξιοποιηθεί. Εξετάζοντας τις μαθητικές επιδόσεις των μαθητών στην Ελλάδα, η παρούσα διατριβή αναδεικνύει το δυνητικό όφελος της χρήσης ανάλυσης δεδομένων στη τεκμηριωμένη λήψη αποφάσεων και την εξαγωγή αντικειμενικών συμπερασμάτων. Η χρήση αυτών των εργαλείων παρέχει κρίσιμη γνώση για τη λήψη αποφάσεων από τους υπεύθυνους χάραξης πολιτικής και την εκπαιδευτική διοίκηση.

Η πλειονότητα των ερευνών της ανάλυσης εκπαιδευτικών δεδομένων σχετικά με τις επιδόσεις των μαθητών επικεντρώνεται στην τριτοβάθμια εκπαίδευση και την online μάθηση. Η παρούσα διατριβή έρχεται να προφέρει την ανάλυση του συνολικού μαθητικού πληθυσμού της χώρας, στατικά και διαχρονικά, και την εξαγωγή αντικειμενικών συμπερασμάτων για διαστάσεις του εκπαιδευτικού συστήματος συνολικά καθώς και επιμέρους εκπαιδευτικών παρεμβάσεων. Επιπλέον, διευρύνει το ερευνητικό πεδίο σε εκπαιδευτικές βαθμίδες με διαφορετικά χαρακτηριστικά από τριτοβάθμια εκπαίδευση, οι οποίες έχουν σημαντικό αντίκτυπο στους μαθητές και την κοινωνία.

Τα ερευνητικά ερωτήματα της διατριβής συνδέονται με την ακαδημαϊκή επίδοση των μαθητών. Το πρώτο ερευνητικό ερώτημα επικεντρώνεται στον αντικειμενικό εντοπισμό των διαφορετικών επιπέδων επίδοσής των μαθητών και αποτέλεσε τη βάση για τις περεταίρω αναλύσεις. Στο δεύτερο ερευνητικό ερώτημα εξετάστηκε η σταθερότητα των επιπέδων επιδόσεων που εντοπίστηκαν, στην πάροδο του χρόνου. Στο τρίτο ερευνητικό ερώτημα εξετάστηκε η λειτουργία του σχολείου ως θεσμού παροχής ίσων ευκαιριών, μέσα από την επίδραση δημογραφικών (μη - ακαδημαϊκών) χαρακτηριστικών, όπως το φύλο, το επάγγελμα του κηδεμόνα και η περιοχή διαμονής, στην ακαδημαϊκή επίδοση. Στο τέταρτο ερευνητικό ερώτημα μελετήθηκε η δυνατότητα αντικειμενικής αξιολόγησης μιας συγκεκριμένης εκπαιδευτικής παρέμβασης, αυτής της ενισχυτικής διδασκαλίας, υπό το πρίσμα των ίσων ευκαιριών για τους μαθητές. Τέλος, στο πέμπτο και τελευταίο ερευνητικό ερώτημα εξετάστηκε η προβλεπτική ικανότητα του GPA στην εκτίμηση των μελλοντικών επιδόσεων, έναντι εναλλακτικών - σταθμισμένων μετρικών, με διαφορετικές σταθμίσεις των μαθημάτων.

Για την κάλυψη της ερευνητικής μας προσέγγισης, χρησιμοποιήθηκαν δεδομένα (δημογραφικά και ακαδημαϊκά) των μαθητών της χώρας από το Υπουργείο Παιδείας. Λάβαμε δεδομένα του συνόλου του μαθητικού πληθυσμού, από την 5η Δημοτικού μέχρι την 3η Γυμνασίου. Τα δεδομένα αφορούσαν: a) Τους βαθμούς σε όλα τα μαθήματα b) Την τάξη κάθε μαθητή c) Το γενικό μέσο όρο βαθμολογίας d) Τις απουσίες των μαθητών e) Το φύλο των μαθητών f) Το επάγγελμα του κηδεμόνα g) Την Διεύθυνση Εκπαίδευσης που ανήκε κάθε μαθητής. Τα σχολικά έτη για τα οποία λάβαμε δεδομένα ήταν από 2016-17 έως και 2018-19.

Στην παρούσα διατριβή χρησιμοποιήθηκε μη εποπτευόμενη μάθηση για τον προσδιορισμό των επιπέδων επίδοσης των μαθητών, προκειμένου να ελαχιστοποιηθεί η παρέμβαση του ερευνητή. Από τον αλγόριθμο προέκυψε μια νέα μεταβλητή, αυτή του επιπέδου επίδοσης κάθε μαθητή, η οποία προστέθηκε στο σύνολο δεδομένων και ταξινόμησε τους μαθητές στα επίπεδα επιδόσεων. Η μεταβλητή χρησιμοποιήθηκε για να εξεταστούν τα υπόλοιπα ερευνητικά ερωτήματα όπως: οι διαφορές στις επιδόσεις με βάση δημογραφικά χαρακτηριστικά των μαθητών, όπως το φύλο, η περιοχή και το επάγγελμα του κηδεμόνα. Τέλος, έγινε διαχρονική ανάλυση του επιπέδου επίδοσης από τάξη σε τάξη, προκειμένου να μελετηθεί η σταθερότητα των επιδόσεων των μαθητών κατά τη διάρκεια του χρόνου.

Η διατριβή κατέδειξε τις δυνατότητες της ανάλυσης δεδομένων για την εξαγωγή ουσιαστικών και χρήσιμων συμπερασμάτων από εκπαιδευτικά δεδομένα, ακόμα και αν αυτά δεν έχουν συλλεχθεί για το συγκεκριμένο ερευνητικό σκοπό. Χρησιμοποίησε για πρώτη φορά, δεδομένα για όλους τους μαθητές σε εθνικό επίπεδο και τους κατέταξε σε τέσσερις, μαθηματικά υπολογισμένες, κατηγορίες, με βάση τις ακαδημαϊκές τους επιδόσεις. Η διαχρονική μελέτη των επιδόσεων των μαθητών έγινε με το συνδυασμό ομαδοποίησης με περιγραφικές στατιστικές μεθόδους και διαπίστωσε σταθερότητα των επιδόσεων διαχρονικά, με τους μαθητές με τις υψηλότερες και τις χαμηλότερες επιδόσεις να παρουσιάζουν εντονότερη σταθερότητα.

Επίσης, διαπιστώθηκε ότι, το επίπεδο επίδοσης του μαθητή επηρεάζεται από μη ακαδημαϊκούς παράγοντες όπως, το φύλο, η περιοχή κατοικίας και το επάγγελμα του κηδεμόνα. Η μη ανεξαρτησία των επιδόσεων από μη ακαδημαϊκά χαρακτηριστικά παρέχει σαφείς ενδείξεις υπέρ του επιχειρήματος ότι, το εκπαιδευτικό σύστημα δεν λειτουργεί ως ένα σύστημα παροχής ίσων ευκαιριών για τους μαθητές.

Η έρευνα διαπίστωσε ακόμη ότι η ενισχυτική διδασκαλία είχε βραχυπρόθεσμα και μακροπρόθεσμα αποτελέσματα στη βελτίωση των μαθητών συνολικά, πλην όμως η βελτίωση αυτή διαφέρει με βάση το επάγγελμα του κηδεμόνα, ευνοώντας πιο προνομιούχους μαθητές. Αυτό καταδεικνύει ένα αντίθετο αποτέλεσμα της ενισχυτικής διδασκαλίας από την στόχευσή της, που αφορά στην ενίσχυση των ίσων ευκαιριών για μαθητές που έχουν κοινωνικά, περιορισμένες δυνατότητες.

Σε επίπεδο συνεισφορών, πρόκειται για την πρώτη ερευνητική προσπάθεια με την χρήση ανάλυσης δεδομένων, σε επίπεδο χώρας. Τα αποτελέσματα των ερευνών μας αφορούν το σύνολο των μαθητών της χώρας, χωρίς την ανάγκη στατιστικής επαγωγής. Διαπιστώθηκε ότι η χρήση εκπαιδευτικών δεδομένων, τα οποία υπάρχουν ήδη στις βάσεις δεδομένων του Υπουργείου Παιδείας ακόμη και στην περίπτωση που δεν έχουν συλλεγεί για το συγκεκριμένο ερευνητικό σκοπό, μπορεί να οδηγήσει σε τεκμηριωμένες απόψεις για τη λειτουργία του εκπαιδευτικού συστήματος. Αυτό επιτρέπει στις υπηρεσίες του υπουργείου να ασχοληθούν σε βάθος με την ανάλυση εκπαιδευτικών δεδομένων για την εξαγωγή νέας γνώσης, που για την ώρα «κρύβεται» στο μεγάλο όγκο δεδομένων του MIS.

Αναπτύχθηκε μια προσέγγιση, αυτή του αντικειμενικού προσδιορισμού επίπεδων επίδοσης, μέσω ομαδοποίησης, η οποία μπορεί να χρησιμοποιηθεί και σε άλλες έρευνες σχετικά με τη μαθητική επίδοση, χωρίς να είναι απαραίτητη η μελέτη κατανομών της βαθμολογίας των μαθητών.

Διαπιστώθηκε ότι υπάρχει αντικειμενικός τρόπος χωρισμού των επιπέδων επίδοσης και χαρακτηρισμού των επιδόσεων των μαθητών, από τον οποίο προκύπτουν συγκεκριμένα και σταθερά σε αριθμό επίπεδα επίδοσης, αναδεικνύοντας αντίστοιχη σταθερότητα στο σύνολο των παραγόντων που επιδρούν στην επίδοση και θέτοντας παράλληλα προκλήσεις στην εκπαιδευτική πολιτική.

Η προσπάθεια για ένα σχολείο ίσων ευκαιριών θα πρέπει να συνεχιστεί, αφού η επίτευξη του στόχου δεν επιβεβαιώθηκε από τα δεδομένα. Η διαφοροποίηση των επιδόσεων μεταξύ μαθητών από διαφορετικά κοινωνικά και οικονομικά υπόβαθρα, δείχνει ότι πρέπει να υπάρξουν επιπλέον προσπάθειες, προκειμένου το σχολείο να λειτουργήσει σαν εργαλείο κοινωνικής κινητικότητας και παροχής ίσων ευκαιριών, μέσω των μαθητικών επιδόσεων.

Επιβεβαιώθηκε για πρώτη φορά και με χρήση συνολικών δεδομένων, η υπεραπόδοση των κοριτσιών σε σχέση με τα αγόρια. Οι αντίστοιχες έρευνες αφορούσαν δεδομένα μαθητικών διαγωνισμών, όπως το PISA, με περιορισμένο αριθμό μαθητών και μαθημάτων που εξετάζονται ή μικρά δείγματα. Η διατριβή επιβεβαίωσε τα ευρήματα για πρώτη φορά σε επίπεδο χώρας, χωρίς την ανάγκη επαγωγής των αποτελεσμάτων.

Συνολικά, μέσω της τεκμηριωμένης εκτίμησης για διαστάσεις του εκπαιδευτικού συστήματος και εκπαιδευτικές παρεμβάσεις, έγινε σαφές ότι η ανάλυση των εκπαιδευτικών δεδομένων της χώρας μας παρέχει τεράστιες δυνατότητες τεκμηρίωσης των αποφάσεων και αξιολόγησης των αποτελεσμάτων των εκπαιδευτικών πολιτικών. Τονίζεται έτσι, η ανάγκη ενσωμάτωσης των δεδομένων του πληροφοριακού συστήματος στη διαδικασία λήψης αποφάσεων, καθώς και η σημασία της προώθησης της λήψης αποφάσεων με βάση τα δεδομένα στην εκπαίδευση.

Λέξεις Κλειδιά: ανάλυση εκπαιδευτικών δεδομένων, επιδόσεις μαθητών, ίσες ευκαιρίες, εκπαιδευτική πολιτική

Abstract

This dissertation highlights the potential offered by the analysis of educational data in primary and secondary education. Starting from a common issue in educational research—the academic achievement of students—it emphasizes the potential for objective evaluation and decision support provided by the use of data in centralized educational systems. Central to this is the concept of equal opportunity education, which runs throughout the thesis.

Since the industrial revolution, education has been seen as a tool for social mobility and equal opportunities. States introduced compulsory public education and financed it through taxation. A main objective remains equality of educational opportunities for students through public provision and state control. But research has highlighted that long-standing social inequalities are also reproduced through education systems, leading to proposals for more inclusive education systems and educational interventions.

The evaluation of the effectiveness of education systems and the expression of views in the public debate often reflect personal perceptions, which are not based on objective evidence. Recently in Greece, systematic collection of educational data has become possible with the introduction of a MIS for primary and secondary education, but the potential for knowledge extraction from it has not yet been exploited. By examining student achievement in Greece, this thesis highlights the potential benefit of using data analysis in evidence-based decision-making and drawing objective conclusions. The use of these tools provides critical knowledge for decision-making by policymakers and educational administrators.

The majority of educational data analysis research on student achievement focuses on higher education and online learning. Additionally, studies often use small sample sizes, which may limit their generalizability. Longitudinal analyses, which could reveal the long-term effects of educational interventions, are scarce in the literature. This thesis analyzed the entire student population of the country, both statically and longitudinally, drawing objective conclusions on dimensions of the education system as a whole as well as individual educational interventions. In addition, it broadens the research scope to educa-

tional levels with different characteristics from higher education, which have a significant impact on students and society.

The research questions of the thesis are related to students' academic achievement. The first research question focuses on the objective detection of different levels of student achievement and forms the basis for further analyses. The second research question examined the stability of the identified achievement levels over time. The third research question examined the function of the school as an equal opportunity institution through the impact of demographic (non-academic) characteristics, such as gender, guardian occupation, and region of residence, on academic achievement. The fourth research question examined the potential for objective evaluation of a specific educational intervention, that of remedial teaching, in the light of equal opportunities for students. Finally, the fifth and last research question examined the predictive power of GPA in estimating future achievement against alternative, weighted metrics with different weights of courses.

To meet our research approach, we requested demographic and academic data of the country's students from the Ministry of Education. We obtained data of the entire student population, from 5 of primary school to grade 3 of Junior High School. The data were: a) Grades in all subjects b) The class of each student c) The overall Grade Point Average d) The students' absences e) The gender of the students (f) The profession of the guardian (g) The education directorate to which each pupil belonged. The school years for which we obtained data were from 2016-17 to 2018-19.

In this thesis, unsupervised learning was used to assess student achievement to reduce researcher intervention. The algorithm added each student's achievement level to the dataset and ranked them by achievement level. This variable was used to answer remaining research questions, such as student achievement differences by gender, region, and guardian occupation. Finally, a longitudinal analysis of achievement levels from grade to grade examined student achievement stability.

The thesis also showed that data analysis can yield meaningful conclusions from educational data, even if it was not collected for research. It used national student data for the first time to categorize them into four mathematically calculated academic achievement categories. The longitudinal study of student achievement found stability over time, with the highest and lowest performing students showing strong stability.

It was also found that the level of student achievement was influenced by non-academic factors such as gender, region of residence, and the guardian's occupation. The non-independence of achievement on non-academic characteristics provides clear evidence in favor of the argument that the education system does not function as a system of equal opportunities for students.

The research further found that remedial teaching had short- and long-term effects on the improvement of students overall, but the improvement differed by the profession of the guardian, favoring more privileged students. This demonstrates the opposite effect of remedial teaching from its objective, which is to enhance equal opportunities for pupils who have socially limited opportunities.

In terms of contributions, this is the first research effort using data analysis at the country level. Our study results relate to all students in the country without the need for statistical inference. It was found that the use of educational data, which already exists in the databases of the Ministry of Education even if not collected for a specific research purpose, can lead to informed opinions on the functioning of the educational system. This allows the ministry's services to engage in in-depth analysis of education data to extract new knowledge that is currently ”hidden” in the large volume of MIS data.

An approach has been developed, that of objective identification of achievement levels through clustering, which can be used in other researches on student achievement, without the need to study distributions of student grades.

It was found that there is an objective way of dividing achievement levels and characterizing student achievement. From this procedure, specific and numerically stable achievement levels emerge, highlighting corresponding stability in the set of factors affecting achievement while posing challenges to educational policy.

As the data do not support the achievement of the target, the pursuit of an equal opportunities school should continue. The differentiation in performance between students from different social and economic backgrounds shows that further efforts are needed in order for the school to function as a tool for social mobility and equal opportunities through pupil achievement.

For the first time, using total data, the overachievement of girls compared to boys has been confirmed. Similar studies have used data from student competitions, such as PISA, with a limited number of students and subjects tested or small samples. The thesis confirmed the findings for the first time at the country level, without the need to induce the results.

Overall, through the evidence-based assessment of dimensions of the education system and educational interventions, it became clear that the analysis of our country's educational data provides enormous potential for informing decision-making and evaluating the outcomes of educational policies. Thus, the need to integrate information system data into the decision-making process is emphasized, as well as the importance of promoting data-based decision-making in education.

Keywords: educational data analysis, students' achievement, equal opportunity, educational policy

Chapter 1

Introduction

The general topic of this doctoral dissertation pertains to educational data analytics in the Greek primary and secondary education system. Specifically, it aims to explore the potential of using educational data to support evidence-based decision-making and policy-making in education. Also, the idea of an equal opportunity education runs through the whole thesis.

Education systems have traditionally been viewed as instruments for social mobility and equal opportunity since the industrial revolution era. To this end, governments established compulsory public education systems and funded them through taxation. Ensuring equal opportunities for students through public provision and state control is mentioned as a key aim. However, research has also highlighted how social inequalities are often reproduced through education systems, leading to calls for holistic schooling approaches and targeted educational interventions.

Decision-making and public discourse around education frequently reflect personal viewpoints lacking objective evidence. Recently in Greece, systematic collection of educational data on stakeholders became possible with a new Management Information System, though knowledge extraction from this data remains underutilized. Examining Greek student academic performance, this dissertation aims to demonstrate the benefits of educational data analytics for empirically grounded decision-making and policy evaluation.

This thesis aims to highlight the significance and potential of data analysis in the educational system given that the current situation is characterized by subjectivity or a lack of support for the opinions expressed. By examining dimensions of the educational system as a whole as well as specific educational interventions through data, it expresses objective and data-based views. Starting from a common issue in educational research—the academic achievement of students—it emphasizes the potential for objective evaluation and decision support provided by the use of data in centrally organized educational systems. A key role is played by the concept of equal opportunity education, which runs throughout the thesis.

The methodology used in the thesis focuses on the analysis of aggregate data in the context of the Greek educational system. The researcher sought to gather data on students' economic demographics and academic performance in order to explore the relationship between these factors and students' grades. To collect the necessary data, a request was submitted to the Greek Ministry of Education for student demographic and academic performance data for the years 2016-17 to 2018-19. This dataset included numerical grade data from the last two years of primary school (5^{th} and 6^{th} grade), as well as data from high school.

To complete the dataset, economic data related to the students' region of residence were used from the National Statistics Authority of Greece. These additional data included information on the gross product and gross per capita product of the region. It is important to note that, due to the broadness of the students' area of residence, there was heterogeneity in these characteristics. From the linear correlation analysis between students' grade point average (GPA) and economic variables, no significant correlation was observed, suggesting that these variables were not explanatory in predicting students' grades.

After the acquisition, data were imported into a new MySQL database for analysis. The first step of the analysis involved preprocessing the data, which may have included removing noise or outliers and performing the necessary transformations. Once the data were prepared, data analysis techniques were used to examine student achievement.

Initially, unsupervised learning was used to rank students in achievement levels. This procedure resulted in a new variable relating to the achievement level at which each student in the country was classified. In this way the students' scores are translated into achievement levels. In addition, a longitudinal study of students' achievement over time was carried out. This longitudinal analysis can reveal students' academic pathways and how specific policies or changes in the education system affect students' long-term students' outcomes.

Descriptive and inferential statistics were used to study differences based on students' non-academic characteristics. The use of both approaches is attributed to the fact that both approaches cover the entire population of the country, although some variables have a small proportion of missing values. This analysis facilitates the study of demographic characteristics and how they interact to either exacerbate or mitigate educational inequalities.

The research questions

This thesis concerns the academic performance and equal opportunities of Greek primary and secondary school students. Five research questions were developed for systematic research:

- The first research question pursuits to identify the different levels of academic performance objectively through clustering analysis of achievement metrics. This will classify students into discrete groups representing varying ability strata, establishing performance benchmarks for subsequent analyses.

- The second research question examines the stability of the performance levels longitudinally using three years of student data. Trend analysis will determine if students typically remain classified within the same level or move between them, illuminating the degree to which ability is predetermined.

- The third research question evaluates the education system's function as an equal opportunity provider by focusing on the influence of demographic characteristics like guardian occupation, gender and region of residence on performance level categorization. Disparities in representation across achievement groups would indicate inequitable access to achievement.

- The fourth research question considers the objective assessment of a specific educational intervention: remedial teaching programs. Remedial teaching is evaluated in terms of its unequal start and its impact on achievement level transitions, depending on the student's background.

- The final research question aims to evaluate the predictive power of the Grade Point Average (GPA) of the first grade of high school as a predictor of achievement levels and the GPA of the two subsequent grades. It also compares it with alternative measures that take into account the relative importance of individual course grades.

By systematically addressing these research questions, the overarching goal of empirically studying student performance through the lens of equal opportunity can be achieved. This approach can also be used in other related studies that use student performance as a metric.

This thesis explores key dimensions of the Greek educational system and policies through a quantitative analysis of student achievement data. The overarching aim was to evaluate whether the education system provides equal opportunities for students from different socioeconomic backgrounds. Several contributions emerge from the research.

The thesis highlighted the identification of stable levels of student achievement through data analysis techniques. Four categories emerged: ”very strong”, ”strong”, ”weak”, and ”very weak” students. This categorization proved stable and helped establish typical ”profiles” associated with each achievement level. The classification scheme captured underlying patterns in student scores and grades that effectively stratified the development of academic profiles. Grade Point Averages are sharply differentiating the levels. The stability in performance levels reflects a corresponding stability in the factors that influence performance. Therefore, the current situation in Greek primary and secondary education seems to have remained unchanged for the three school years studied.

One of the most substantial findings to emerge from the thesis's longitudinal analysis was the high degree of stability observed in student achievement levels over a three-year period. Using two distinct datasets spanning primary and secondary education, clear patterns of stable performance levels derived from initial assessments were tracked. Across both stages of schooling, a majority of pupils tended to remain situated within the same level of achievement that characterized their starting point.

Particularly notable was the enduring nature of the extremes: over three-fourths of students originally deemed ”very strong” or ”very weak” persevered within those designations and did not migrate to different achievement levels. This entrenched persistence at the upper and lower levels was remarkably enduring, which points to profoundly ingrained influences on achievement. For intermediate achievement levels deemed ”strong” and ”weak”, some fluctuation was evident, but categories still demonstrated stability across the tracking windows. Within these strata, trends of declining achievement into high school were discerned, implying a ratcheting-up difficulty curve.

For policymakers and educators aiming to mitigate the impacts of this stable situation, profound and sustained intervention seems merited. Stability also underlines the need for immediate attention. Pupils who consistently underperform need proactive, targeted help in the early stages of the educational pathway to prevent the consolidation of performance ceilings. For high-achieving students, maintaining achievement levels proves equally critical.

A core focus of the thesis was evaluating whether the Greek education system provided equal opportunities for students regardless of non-academic background characteristics. However, the quantitative analyses revealed several demographic factors that appeared to significantly influence measured achievement levels, indicating the system was not yet functioning as an institution providing equal opportunities. Guardian occupation emerged as a particularly potent indicator, with manual workers' children markedly underperforming relative to expectations. In contrast, students with guardianship professionals, freelancers, doctors, teachers, and civil servants perform disproportionately well. As occupation signals dimensions of social class, such as parental education, income, availability of resources and attitudes towards learning, this shows that socio-economic status is a determinant of achievement and a barrier to equal opportunities.

Gender differences also found, with girls consistently achieving above predicted levels while boys featured disproportionately in the lowest level. This thesis makes a valuable contribution to the research field by employing comprehensive data at the country level. The results of our study provide insights into the phenomenon of girls' over-achievement at the national level, specifically focusing on a three-year time frame including elementary and middle school education. On the other hand, the previous studies were constrained to samples or competitions' data, impacting the extent to which the conclusions can be generalized.

The region of residence seemed less deterministic due to the wide area covered. The regions can be divided into three performance groups, with the high-performing group of areas being those areas associated with overachieving students in the 'very strong' category. In contrast, the group of low-performing areas shows overperformance by very low-performing students. A linearity was also found in the rates of overachievement and underachievement in the high- and low-performing groups. Also, a consistent association between the disadvantaged region of West Attica and very low performance was identified. National studies suggest this likely relates to concentrations of Roma students confronting pervasive educational barriers.

Taken together, this thesis shows that the socioeconomic and sociodemographic linkages confirm that the education system currently reproduces rather than compensates for extant social stratification. Performance differences map closely to social and economic background attributes. In an ideal world, a student's achievement depends purely on merit and effort, independent of non-academic characteristics. An equal opportunities education remains a constant challenge.

The thesis conducted, through a black box approach, a detailed evaluation of remedial teaching as a specific educational policy aimed at promoting greater equality of opportunity within the Greek system by targeting disadvantaged students. Initially, some promising outcomes were observed from this widespread intervention. Seven of ten students demonstrated sufficient progress to climb achievement levels in the short term, and three of ten remained at these levels after two years, exceeding typical expectations.

However, digging deeper revealed that the policy may be reinforcing rather than reducing social disparities over the long run. Privileged students appeared to derive bigger benefits from remedial teaching, both in the short and middle term. Also, privileged students regressed to lower achievement bands at markedly lower rates than their non-privileged peers following program completion. This suggests remedial teaching in its current form risks entrenching relative advantages by augmenting already privileged students while struggling to overcome deep-rooted obstacles confronting non-privileged student populations.

The thesis also conducted a thorough study into the predictive capabilities of student grade point average (GPA), given its central role in this thesis and within the Greek educational system. Through quantitative analysis, several compelling findings emerged supporting GPA's validity as a metric. Firstly, GPA demonstrated remarkable consistency in forecasting future academic performance levels—only 1% of students saw significantly different levels from the baseline scenario between the first and third years. This affirms that GPA successfully captures the variation in developing achievement levels. Additionally, first-year GPA demonstrated excellent predictive power for later secondary school performance as measured by GPA in the following years as the prediction errors remained small. These associations substantiate that GPA tracks achievement in a predictable manner. This lends credence to the reliance on GPA regarding student achievement.

Finally, after experimenting with alternative GPA measures, none demonstrated clear superiority over conventional GPA in terms of predictive precision. While some debate exists around the optimization of the GPA, this national-scale analysis indicates the traditional GPA serves well as a holistic summary statistic of academic achievement and a predictor variable. Overall, the thorough testing provides strong evident-based backing for the ongoing use of GPA within the Greek educational system. This reinforces the reliability of the results and conclusions of our studies, which relied significantly on this indicator.

In summary, this thesis used quantitative analysis of aggregate student data to provide new knowledge about Greek education. Specific data-based conclusions were drawn at the level of the education system and educational policies. It also highlighted important challenges in pursuing equality of opportunity in education as a whole and at the level of educational interventions. Overall, a scientific approach to the assessment of student achievement in primary and secondary education was proposed. This approach can be used in other studies of student achievement. Finally, the current thesis highlights the potential of educational data analysis to document opinions and underpin decisions in education.

The introduction chapter of this thesis encompasses several essential elements. It provides the general background of the topic, clarifies the scope of the study, emphasizes the importance of the thesis, outlines the research questions and methodology, highlights the expected contributions, and provides an overview of the thesis structure. Through these components, the introduction establishes the context, relevance, and objectives of the research while also offering a roadmap for the reader to navigate the subsequent sections of the thesis.

The second chapter presents the literature review and highlights the extensive debate on higher education, while secondary and primary education lack similar research interest. The characteristics used in the studies are presented, and the results of their use are evaluated. The need to link academic research to educational reality is emphasized in order to provide evidence-based support for educational policy. But this objective remains largely unfulfilled. It also presents a theoretical framework for data-driven policymaking in the era of digital governance. Decision-making in the context of such a data-based public sector policy cycle and continuous evaluation within this policy cycle are described.

The third chapter begins by exploring the concept of educational data, along with its various types and sources. It then introduces a framework for using educational data and its role in educational decision-making. The chapter also delves into the decision-making cycle in the context of the data-based era, examining how data supports policy decisions. Furthermore, it discloses the dataset used in the thesis, along with information about the data sources. The chapter finally describes the data cleaning, making sure the data's quality and reliability are maintained for subsequent analysis.

The chapter three examines educational data, its different types and its application in educational contexts. It presents a framework for analyzing educational data. In addition, the chapter refers to the specific data used in this thesis, obtained from the Ministry of Education, which includes information on the coverage of the student population. The chapter also describes the data collection process and highlights the rules followed to clean the data, ensuring the accuracy and reliability of the data for subsequent analysis.

Chapter four describes the process of determining student achievement levels. It discusses the data collection and analysis, the technique and the algorithm used. Four fixed levels of achievement and the ranking of the levels were identified. The relative magnitudes of achievement levels and their stability in different grades of primary and middle school were studied.

Chapter five presents a longitudinal analysis of students' achievement levels using descriptive statistics. The analysis involved two data sets: from 5th grade to 1st grade and from 1st grade to 3rd grade. Performance stability was found based on the initial grouping of students in either 5th grade or 1st grade. The stability was primarily related to ”very strong” and ”very weak” students.

Chapter six presents the role of the school as an equal opportunity institution. Using a χ^2 test, the role of the guardian's occupation on student achievement was examined. The correlation of performance with socio economic characteristics was found, leading to the conclusion that the Greek school is not an equal opportunity institution. A key finding is also the overachievement of students from family backgrounds with intellectual-type occupations.

Chapter seven examined the effects of other demographic factors. Gender was found to separate student performance, with girls doing better than boys in general. The outperformance of girls grows over the years, particularly in high school. We also categorized the effects of residential areas into three levels, with no pattern identified within groups. The West Attica region appeared to systematically underperform in primary education, which is linked to the presence of minorities in the area.

Chapter eight evaluated remedial teaching; an educational policy aimed at improving the performance of underprivileged students. Using a 'black box' evaluation, it was possible to assess the short- and long-term effectiveness of remedial teaching. Overall positive results were found, but also the inability of remedial teaching to work to help more ’underprivileged’ students.

Chapter nine assessed the predictive power of the GPA. The predictive power was examined in relation to achievement levels and the GPA of subsequent classes. It was found that GPA was a strong predictor variable in both cases. GPA was studied and compared against alternative metrics, demonstrating that traditional GPA is the best alternative.

Chapter ten delves into an in-depth discussion of the primary findings and their implications. It addresses the limitations and validity concerns associated with the thesis. Moreover, the chapter offers policy recommendations grounded in the data and provides suggestions for future research directions in the field.

Finally, chapter eleven presents the main conclusions of the thesis. They relate to the finding that the education system and remedial teaching do not provide equal opportunities for all students, regardless of their social and economic background. Also, there is stability in student performance over time, at four levels of achievement. This stability reflects a corresponding stability of the variables that affect performance.

Chapter 2

Literature review

To assess the extent of the research in this study domain and delimit the scope of inquiry, a literature review of the studies spanning the past eight years, namely from 2015 onwards, was done. The subject of the studies reviewed was the assessment of student achievement. The extent to which the results of the studies were applied at a practical level was also examined. With this approach, the literature review conducted had three main objectives.

- To provide a summary of the existing discourse in the research field.
- To identify research spaces in the current research, highlighting the contribution of this thesis.
- To highlight significant studies, theories, methods or theoretical frameworks that may be applied in the assessment of student achievement research.

Also, in order to connect the results to everyday educational practice and educational management, the administrative context of using data for decision making and opinion making about the educational system was examined in detail.

2.1 Current situation of the academic field

In the last 15 years, the use of data mining techniques to examine educational data sets has increased significantly. This reality has been aided in part by the exponential rise of educational data that has occurred in recent years (Papamitsiou & Economides, 2014). The implementation of information systems in educational institutions permits the recording and storage of significant amounts of data over extended periods of time (Peña-Ayala, 2014). The growth of both synchronous and asynchronous forms of distance learning has made possible the collection more and different kind of data. Because of this, the circumstances for using data mining methods in education have been established, and the discipline of Educational Data Mining (EDM) has developed into its own multidisciplinary field (Romero & Ventura, 2007; Papamitsiou & Economides, 2014; Baker et al., 2016).

There has been a number of different attempts made to identify the sub-disciplines of this scientific domain. According to Agasisti and Bowers (Agasisti & Bowers, 2017) the use of data mining in the realm of education may be broken down into three separate approaches: a) Educational Data Mining (EDM), b) Learning Analytics (LA), and c) Academic Analytics (Ac An). The basic purpose of EDM is to understand patterns in educational interactions using data mining technologies. While maintaining its primary emphasis on teaching, LA incorporates a significant number of the EDM models. In this sense, its primary purpose is to improve the teaching methods. Ac An emphasizes the organizational level, as well as the potential to improve the educational process and its results. In their latest review (Romero & Ventura, 2020) made a more detailed split in the scientific field of Educational Data Mining:

- Learning Analytics (LA) is the study of teaching practices and student achievement data, and the planning, implementation, and evaluation of educational programs, as well as the creation and analysis of the efficacy of different educational activities (Prieto et al., 2014). It lays a strong focus on the teaching as evaluated from the perspective of the instructors.

- Academic Analytics (Ac An) emphasize to economic managerial side and concerned with the gathering, analysis, and visualization data from academic activities such as academic curricula,courses, fees etc to help management in decision making and strategy development (Campbell et al., 2007).

- Data-Driven Decision Making in Education (DDDM) is the methodical collecting and analyzing a wide variety of educational data to support decisions that lead to enhancing educational effectiveness and solving problems (Datnow & Hubbard, 2016).

- Big Data in Education (BDE) incorporate large amounts (in volume, diversity, value, and speed) of data into instructional settings (Daniel, 2019).

- Educational Data Science (EDS) include the use of data collected from educational environments for the sake of educational research, assessment, and systemic improvement addressing educational concerns (Romero & Ventura, 2017). It is a concept that unifies statistics, data analysis, and data management.

Because the tools, research methods, and the use of findings have a tendency to overlap, it can be challenging to differentiate between these approaches. As a result, many authors in the academic literature refer to all of them together as EDM. As the scientific subject as a whole is still in the process of quickly developing, this classification should in any event be considered ”temporary” or ”early” (Agasisti and Bowers, 2017). In this chapter, we present a comprehensive assessment of the literature on data analysis in education over the past few years, with an emphasis on studies conducted to measure and forecast student achievement. We discuss the tremendous potential of the future of this area of research and the intriguing paths that have yet to be explored after identifying existing trends.

Some techniques and methods of data mining have become common in educational research. The categorization that is now considered as common was done by Romero and Ventura (2010), who grouped the techniques used into the following categories:

Prediction: Romero and Ventura (2010) refers to ”prediction” as the process of developing a model with the intention of finding a variable's value using a mixture of other independent factors. Methods of classification or varied approaches of regression analysis may be utilized for prediction.

Distillation data for judgment: This technique sums up and presents a lot of educational data in a way that is useful, interactive, and aesthetically pleasing. This helps people understand the information and make better decisions. The aim is to distill the data so that humans can judge it. It facilitates user's comprehension of the content (teachers, education executives).

Relationship's mining: The purpose of this method is to determine how various variables are interconnected. Can consist of determining which variables are connected and to what degree. Relationship mining may generally be divided into the following categories: a) *Association rule mining* (a machine learning method for finding interesting relationships between variables in large databases). b) *Correlation mining* (its goal is to find interesting or unusual patterns of dependencies between several different variables (sequences, signals, images, videos). c) *Sequential pattern mining* (a type of data mining that looks for statistically important patterns between sets of data where the values come in a specified order) and d) *Causal data mining* (with the aim of discovering causal mechanisms).

Discovery through models: In this approach, a model that was initially created through machine learning is utilized in another study as a tool. The model that was developed makes possible further analysis between the estimates provided by the model and the observed values of the variables that are being investigated.

Text mining: It is the procedure of study massive collections of written texts in order to generate new data and convert text into structured data for the purpose of further analysis.

Clustering: Finding groups of instances that are similar to one another in some way is the objective of the clustering technique. Once an initial number of groups has been set, instances may be classified by finding the cluster they belong to which they are most similar. Clustering is a useful tool in EDM that may be used to group comparable course content or to categorize students according to the learning or interaction patterns.

Outlier detection: This is the method of detecting outliers involves the identification of data points within a dataset that do not fit in with the rest of the instances. In education, this data might reside in students who have trouble learning, teachers whose behavior is significantly different from the norm, and so on.

Social network analysis: The concept "Social Network Analysis" (SNA) refers to the modeling of linkages and transformations in interconnections among teaching-related actors (teachers, students etc.).

EDM has been the subject of a number of different literature reviews, concentrating substantially on the aims of the research, methods, and the variables considered for prediction. The reviews with the most citations and the greatest impact on the scientific field are the following:

- **Romero και Ventura, (2007)** studied papers up until 2005. The documents were presented in accordance with their aims. They found a rise in the number of papers published each year. The review also analyzes some possible future trends in the scientific field. It emphasizes the importance of implementing data mining into the educational environment, as well as the importance of EDM for academics and external users.

- **Baker and Yacef, (2009)** review analyzed the steadily escalating growth rates of studies that have been observed in the EDM sector. It recognized the rise in the amount of the available data as consequence of the growth of on-line platforms and the systematical collection of data from educational institutions. It outlined the changes made in the objectives and methods that were used. In this review, both the significance of having open educational data and the great opportunity for progress of EDM is mentioned.

- **Romero and Ventura, (2010)** classify the different types of EDM users. They organize techniques according to the strategies that were utilized and the objectives that were sought. They also express their beliefs about the need for (a) simplicity of results for non-specialists, (b) integrating in a basic setting, and (c) standardization of models and data used.

- **Papamitsiou and Economides, (2014)** carried out the very first systematic literature review in this field. The articles are organized into different categories according to the methodologies, objectives, and learning environments. They brought attention to the most important trends in the area. The added value that EDM and LA have also examined. The use game based learning and mobile learning was also offered as alternative sources of educational data. Finally, a SWOT analysis was carried out.

- **Peña-Ayala, (2014)** has carried out an in-depth presentation of the publications, categorizing them according to the methodology, algorithms, and aims pursued. In addition, statistical and clustering techniques were used, and the results of these methods revealed two distinct patterns of EDM methodology. A SWOT analysis was carried out, and a presentation of the enormous potential offered by the widespread implementation of information systems in educational settings was also made.

- **Sukhija, Jindal και Aggarwal, (2015)** presented the methods, the tools, and the findings of several papers spanning the years 2001 to 2015. It has been determined that EDM faces a few obstacles, the most significant of which are (a) the absence of data sets, (b) a lack of trust on the part of educational authorities in EDM's findings, and (c) a lack of coherence in the data sets that are available. In addition to this, they stressed that the majority of the research consisted of just limited-scale tests.

- The review of **del Rio et., al**., **(2016)** was a primary focus on the prediction of the academic success in tertiary education between year 2011 and 2016. They described the basic methodologies the predictor variables, and the practical application of the studies.

- In the critical review of **Papadogiannis, Poulopoulos και Wallace, (2020)** a plethora of methods as well as a prevailing emphasis on higher education were shown. The academic community was the most prevalent intended audience in the papers examined. Numerous algorithms have been found that enable researchers to assess the usefulness of their findings in gauging student achievement. Satisfactory levels of accuracy were reported without any statistically significant differences across the different methods. Moreover, the use of more complex algorithms did not lead to significant benefits. It was discovered that data mining techniques are used insufficiently to inform institutional and policy decisions in the field of education. Neither secondary nor primary schooling seemed to have piqued the attention of the scholars, since the great bulk of those publications concentrate on higher education.

- In the general literature review of the EDM field, **Romero and Ventura, (2020)** recorded the rapid growth of the field, the increasing number of books published, the conferences organized, the data sources available, the methods of analysis used and much more.

It is found that the majority of the reviews focused on the technical part of the research and noted in full the methodologies and the approaches used. The necessary link between research findings and educational policy has not yet been established. This link refers to any level, from the individual class to the national education system level. The Educational Data Mining may assist support decision-making at any of these levels. Agasisti and Bowers (2017) created a framework to bridge the gap between EDM and educational policy decisions at various levels.

2.2 Literature review method

The evaluation of student achievement is the focus of our research. This aim encompassed in a significant portion of EDM studies, as a result of the growing attention of researchers in recent years. Numerous studies have been published, most of which focus on higher education. (Bydžovská, 2016; Papadogiannis et al., 2020). Without a doubt, the value of early diagnosis of student achievement is important, as it can be used by educational institutions to develop actions and policies. To achieve the aim of the prediction of academic achievement a significant number of different types of data are used. These data refer to students' demographics as well as academic and behavioral data their activities on the internet, grades etc. (Rio et al., 2016). The primary objective is to summarize of the findings from studies conducted over the past seven years in the prediction of academic achievement.

This review it presents the articles published in the last eight years (2015–2023) that refer to the assessment of student achievement using EDM techniques, and on the other hand, it presents studies in which the results were applied to administrative or policy decision making.

Following criteria were used to compile this literature review, following the stages of the PICO approach (Pai et al., 2004), which was adapted form the medical research.

- **POPULATION:** Articles about the use of DM techniques that aim to predict or improve academic achievement in secondary schools, higher education institutions, and online courses are included in this population.
- **INTERVENTION:** Refers to methods, algorithms, and features used in the studies.
- **COMPARISON:** Comparative analysis was made between several types of data, education level studied and algorithms used.
- **OUTCOME:** It refers to the approaches, algorithms, and characteristics that are utilized the most frequently by authors and the implications of the findings for the field of education.

In order to better organize the structure of the literature review, the following research questions were posed.

Research questions are the following:

- **Research Question 1:** What methods and algorithms were applied in the process of predicting academic achievement?
- **Research Question 2:** What was average performance by method?
- **Research Question 3:** What characteristics are used to forecast students academic achievement?
- **Research Question 4:** In what extent findings utilized in the decision-making process?

Literature resources:

The following online libraries were searched: Springer Link, IEEE Xplore, Science Direct, ERIC, ACM and Journal of Educational Data Mining (JEDM). An initial search for articles was conducted using the Google Scholar search engine, which makes it possible to identify a variety of scientific publications. The online tools of each database were used to conduct additional searches, which have been completed in their entirety. Specific keywords are used in order to locate publications pertinent to this review. The following keywords were selected from the several alternatives available.

KEYWORDS:

The following criteria were utilized for the search: (Educational Data Mining) AND (Higher OR Secondary Education OR on-line) AND (student achievement OR student achievement OR student dropout).

To choose the most appropriate papers from a massive list generated by a search, it was required to set criteria for accepting or rejecting articles. In accordance with (Kitchenham et al., 2010) methodology, the following inclusion and exclusion criteria were used.

INCLUSION CRITERIA:

- Papers that attempt to forecast the academic achievement of students in secondary schools or tertiary institutions, or online education programs.
- Papers that use DM techniques for the purpose of predicting students' achievement or drop out of secondary, tertiary or online education.
- Publications that appeared in books, volumes, scientific journals, and conference proceedings.

- Papers that provide extensive information on the research approach that was used.

EXCLUSION CRITERIA:

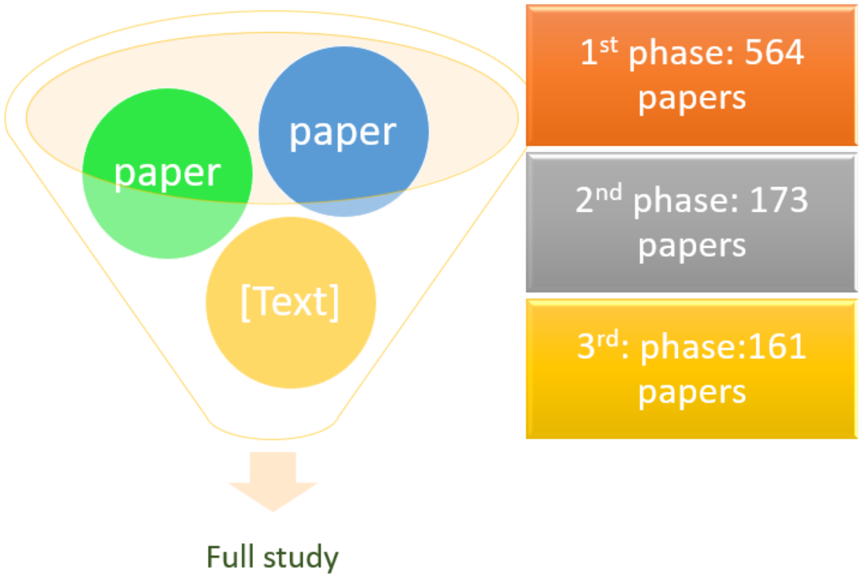
- Publications that do not use data analysis methods to predict student academic success or dropout.

- Papers that do not include specific details of the attributes applied.

- Papers that do not contain much detail about the algorithms used.

- Papers that do not provide in-depth information that can be utilized to evaluate the algorithms that were employed.

Figure 2.1. Studies selection

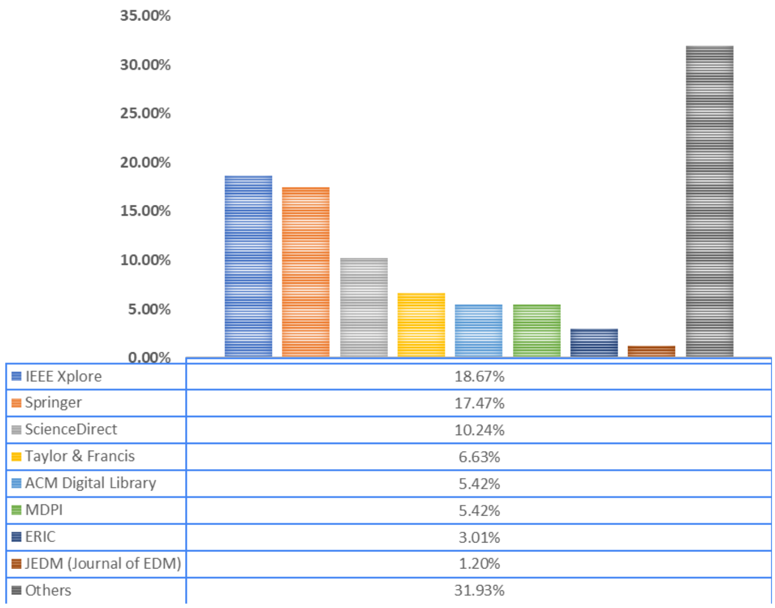


We found 564 publications during the discovery phase of the review.After the titles and abstracts of the publications were examined, their number was narrowed down using the inclusion criteria that had been established. This method resulted in the rejection of 391 papers, leaving 173 papers for additional study throughout the entire text. After reading all of the documents and carefully considering whether to include or exclude, we came up with a set of 161 articles. The critical literature review was built upon this foundation. The sources of the articles that we looked at are outlined in Figure 2.2 along with the percentage of articles that come from each source.

2.3 Findings

As it can be seen from Figure 2.2 from all the articles selected for the review, the main source was IEEE Xplore (18.67%). This was followed by Springer (17.47%) and Science Direct (10.24%). Smaller percentages were from other publishers, while a large proportion (31.39%) were individual papers, mainly from conference proceedings and edited volumes.

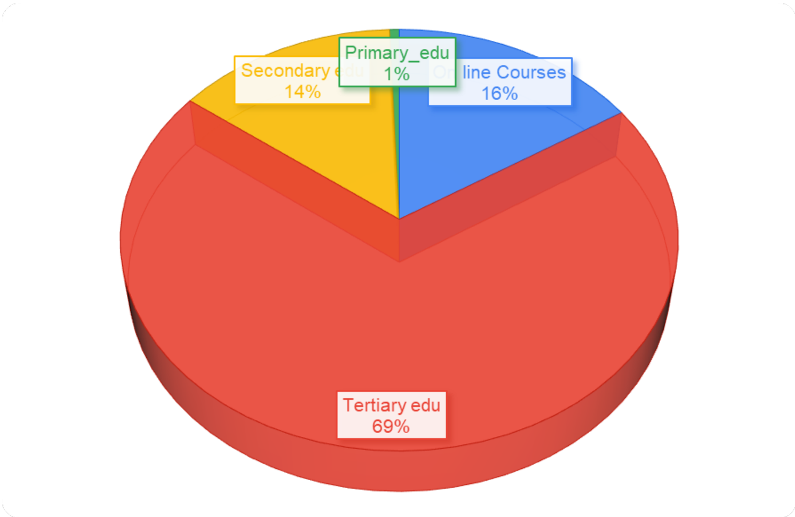
Figure 2.2. Review sources



The overwhelming majority of the articles were on higher education. This may be due to improved data access made possible by the use of learning management systems (LMS) in tertiary education, as well as the ease with which scientific experimentation can be conducted in higher education settings where the study's subjects—the students—are adults and able to consent to the study's conduct. As shown in Figure 2.3, the share of research undertaken in connection to higher education institutions reached 69%. Studies in secondary education came in second, accounting for 14% of the research, while online courses accounted for 16%. Additionally, we discovered that online data heavily dominated in the majority of studies conducted on distance learning platforms.

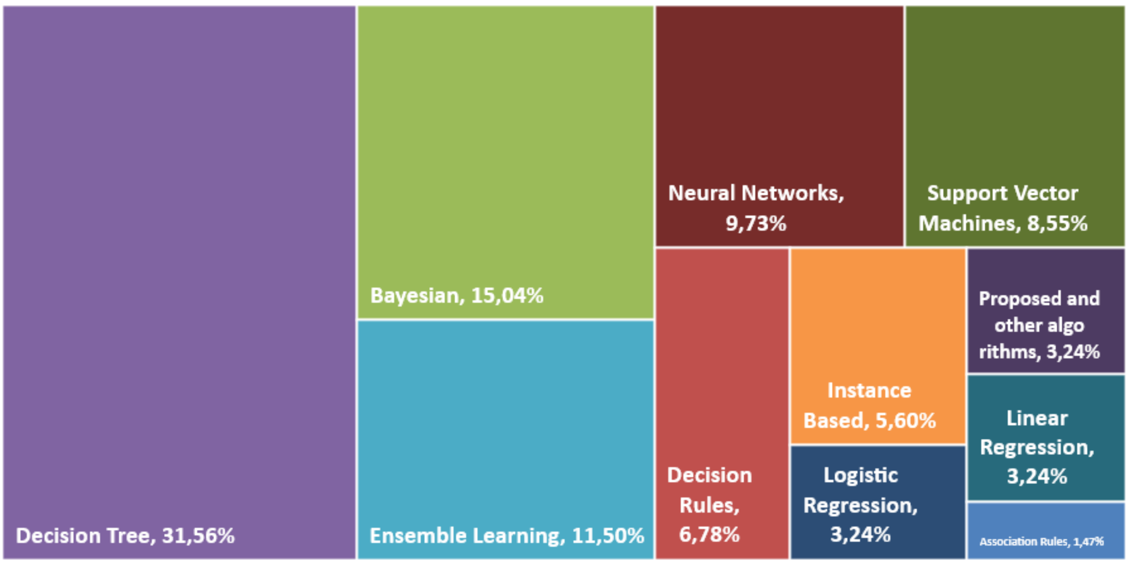
In the majority of the articles, more than one algorithm was utilized. In order to facilitate the completion of this review, the researchers' algorithms were separated into eleven categories in terms of the approach that they employed: Association or Decision Rules, Ensemble Learning, Decision Trees, Bayesian Methods, Instance Based Methods, Neural Networks, Logistic Regression, Linear Regression, Support Vector Machines, Linear Regression and Other Methods were used. the number of articles, per method and the corresponding algorithm evaluation scores were counted. Figure 2.4 demonstrates the

Figure 2.3. Studies per educational level



frequency per method used. The most commonly utilized evaluative criterion was accuracy.

Figure 2.4. Usage per method

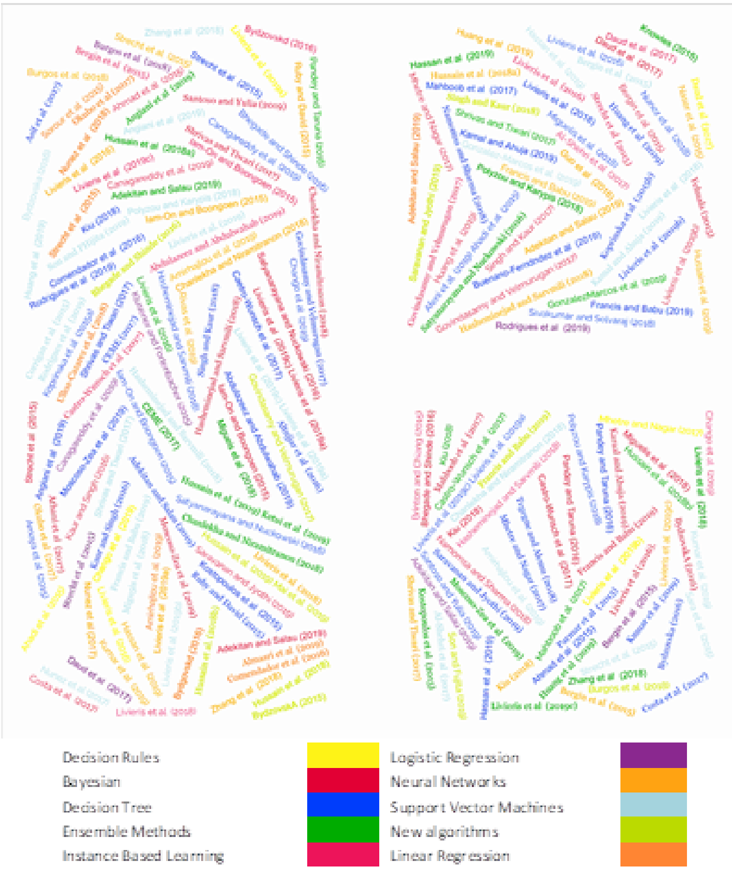


A significant number of articles using decision tree algorithms have been found. This high frequency may be linked to the extensive usage of the approach in general as well as the availability of Decision Trees algorithms in well-known DM programs such as WEKA, which is a popular tool in the studies. Algorithms based on Bayesian methodology were also utilized frequently. More advanced methods, such as neural networks, have a lower frequency. There has also been some application of ensemble learning methods and support vector machines, but to a smaller extent. Logistic Regression, despite the fact that has very accurate results, was only used in a few cases. The same is true for Instance-Based

Learning.

Next, a more in-depth presentation of the algorithms that are used the most frequently is provided. Because of the use of a variety of measures for assessment, there is a presentation of the statistics of studies who used the accuracy as evaluation tool.

Figure 2.5. Citations word cloud

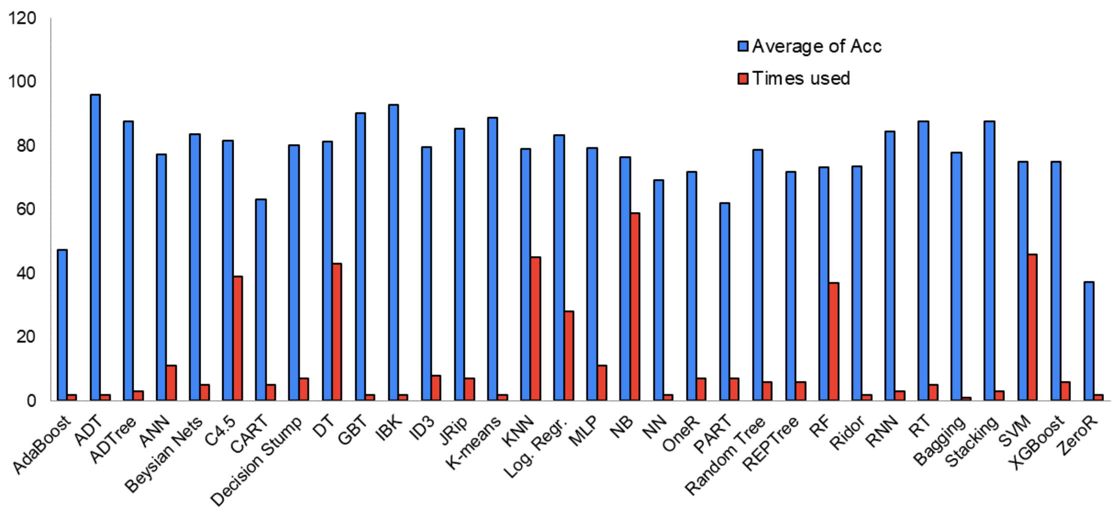


Bayesian methods were utilized in a significant portion of the research papers that we reviewed. The Naive Bayes algorithm was employed significantly more frequently (59 times). Bayesian Networks were utilized in a few studies as well. The Naive Bayes algorithm was widely used as a benchmark against which the accuracy of other, more complex, algorithms. It has been demonstrated that the Naive Bayes algorithm has an accuracy that is satisfactory, notwithstanding the speed at which calculations are performed and the tiny number of resources that are consumed. It was found that the Naive Bayes algorithm had the maximum accuracy in five different studies (Bergin et al., 2015; Zhou et al., 2015; Kamal and Ahuja 2019; Kumar et al. 2019; Santoso and Yulia 2019), while the algorithm's mean accuracy achieved 0.75.

Decision Tree algorithms comprised the majority of the used techniques and reached the greatest accuracy levels. Certainly, there are several algorithms within the category, such as ID3, CART, C4.5 and so on. As a direct consequence of this, researchers have the chance to select from a wide variety of methods, frequently within the same paper. In most cases, the algorithm C4.5 achieved the highest accuracy. The RepTree and ADTree had a high degree of accuracy, despite the fact that they were not used very often.

Logistic Regression, a sophisticated algorithm that has been utilized in only a few articles, showed an accuracy score comparable to Decision Trees. Instance-based learning techniques, support vector machines, and neural networks all demonstrated lower average accuracy. However, we must emphasize that comparing the performance of numerous techniques from different areas is a risky task. In any event, researchers didn't use these sophisticated algorithms as much as they might have.

Figure 2.6. Frequencies and average accuracy per algorithm



The Random Forest algorithm was used with a moderate accuracy. Esemble methods were used in a significant percentage. Stacking and AdaBoost algorithms exhibited a higher level of accuracy, despite the fact that they were used far less frequently. Another fact that attracted our attention was the incredibly low frequency with which unsupervised learning algorithms were used, which occurred just three times.

The use of regression (in many forms) was limited to a small number of studies. Eleven papers were included in this study, three of which were evaluated using the coefficient of determination (R^2). RMSE was applied four times. There were also four instances in which the MSE was used.

Given the limitations in the feasibility of comparing the accuracy of different algorithms in different domains, we study whether a particular method had a higher accuracy score. Assuming the normality of the distribution of accuracy score values, we used the One-Way Anova test. According to Table 2.1, was no statistically significant difference in the performance of the different algorithm categories. [F = 0.786, p = 0.778].

Table 2.1. ANOVA test comparing the accuracy of different methods

	Sum of Squares	df	Mean Square	F	p-value
Between Group	0,567	29	0.2	0.786	0.778
Within Group	6,245	251	0.25		
Total	6,812	280			

The frequency of each evaluation measure was used is outlined in Table 2.2. As a result of the widespread usage of classification algorithms, the associated measures emerged more often. Accuracy was the most often used metric (79.48%), followed by F1 (12.50%), while AUC and precision were seldom used. Tools evaluating regression such as R^2, MSE, and RMSE are displayed when this method used.

Table 2.2. Evaluation tools

Measure	Frequency
Accuracy	79,48%
F1 measure	12,50%
Error Rate	1,18%
AUC	1,18%
Precision	0,24%
R^2	0,71%
Mean Squared Error (MSE)	2,59%
Root Mean Square Error (RMSE)	2,12%

Student grades were the most often utilized attribute in the studies (32.8%). Student demographic characteristics were also used very often (24.87%). Typical examples of demographic characteristics were: gender, age, parent's occupation, geographical location, and other similar factors. In addition to the demographic data, socioeconomic data was used. This data concerned the students' and their families' financial or social conditions and it was used to better understand the student population (Miguéis et al., 2018). Academic data, except grades, was also commonly used as attributes (23.28%). These included class attendance, activity involvement, number of projects involved etc. Fewer studies used student activity data in a digital or physical setting (13.76%). The use of behavioral, social, economic and other data was marginal.

To evaluate the accuracy of these data categories, the average accuracy for the three most popular data categories (grades, demographics, and academic features) was determined. The percentage frequencies of the data categories are shown in Figure 2.7. The majority of the papers reviewed (32.80%) used students' grades in analysis. To evaluate students' final academic achievement, most researchers studied grades from earlier classes

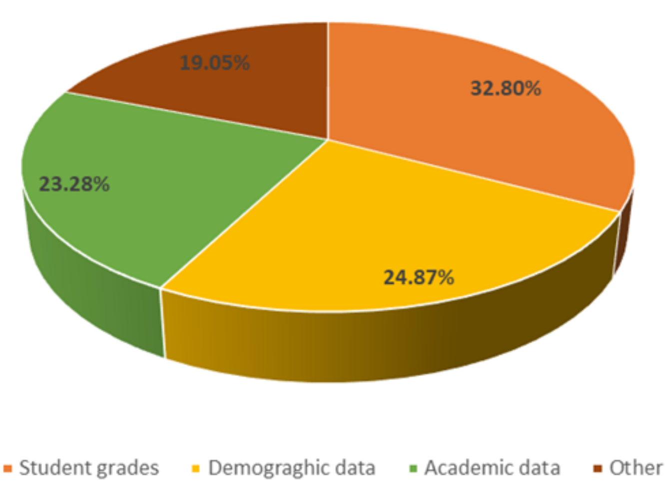
Table 2.3. Utilized attributes

Attribute's type	Frequency
Student grades	32.80%
Demographic data	24.87%
Academic data	23.28%
Student on line activity data	13.76%
Behavioral data	3.17%
Socio-economic	1.06%
Text records	0.53%
Other	0.53%

or from the beginning of the semester. Demographic data was used in a sizable proportion of the studies (24.87%) and other academic information was used at 23.28%.

The studies on student achievement on online learning was dominated by data pertaining to student on-line activities in different platforms. The platforms included massive open online courses (MOOCs) and distance learning platforms provided by universities (Amrieh et al., 2015; Tang et al., 2015; Mahboob et al., 2017; Hassan et al., 2019). In addition to this, the students' demographic data and scores they received on a variety of classes were used. The use of text data was found in only to two studies conducted by Zhou et al., (2015; 2021) although behavioral data, internet records (logs), and motivational data were used to a significantly lower extent than those mentioned above.

Figure 2.7. Characteristics used in studies

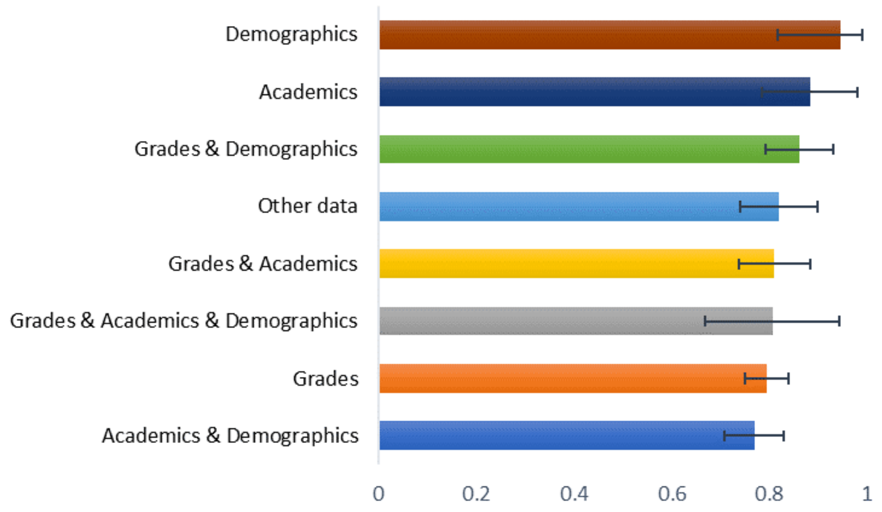


We also analyzed the impact of using academic,demographic data, and grades as features had on the variation of the average accuracy. The three categories of features were used because they that featured most often in the analyzed articles. In Table 2.4, we can see that the utilization of the grades as explanatory variables offers a high level of accuracy (from 0.7494 to 0.8401) and also used at 26.60% of the total papers. In contrast, accuracy declined when only demographic and other academic data are used (0.7085 to 0.8295). The use of demographic data on its own appears to be the most accurate method, but it can only be applied to one study (3.19%). Using confidence intervals, we can see that none of the possible feature combinations has a statistically significant better accuracy. In terms of variation, we can see that the use of tree categories together shown the biggest variation. In contrary, using only grades express the lowest variation. It is generally found that the accuracy levels of the categorization algorithms are close to 0.80 for most studies. Although some studies showed extremely high accuracy (I. Livieris et al., 2016; Burgos et al., 2018), this was not true for the majority of them. Besides, the small samples in the case studies in which these algorithms were used do not allow general conclusions to be drawn.

Table 2.4. Accuracy per characteristics used

Grades	Academics	Demographics	%	Mean	Lower	Upper
NO	NO	NO	12.77%	0,8200	0,7407	0,8992
		YES	3.19%	0,9456	0,8164	10,748
	YES	NO	5.32%	0,8828	0,7848	0,9808
		YES	21.28%	0,7690	0,7085	0,8295
YES	NO	NO	26.60%	0,7947	0,7494	0,8401
		YES	13.83%	0,8615	0,7927	0,9302
	YES	NO	8.51%	0,8096	0,7366	0,8825
		YES	8.51%	0,8057	0,6682	0,9432

Figure 2.8. Average accuracy and Standard deviation per attribute combination



The results of EDM studies may be used for academic or research purposes, but they may also have direct uses in school administration. According to Romero and Ventura

(2007), some potential users of the findings of EDM researches are presented in the Table 2.5.

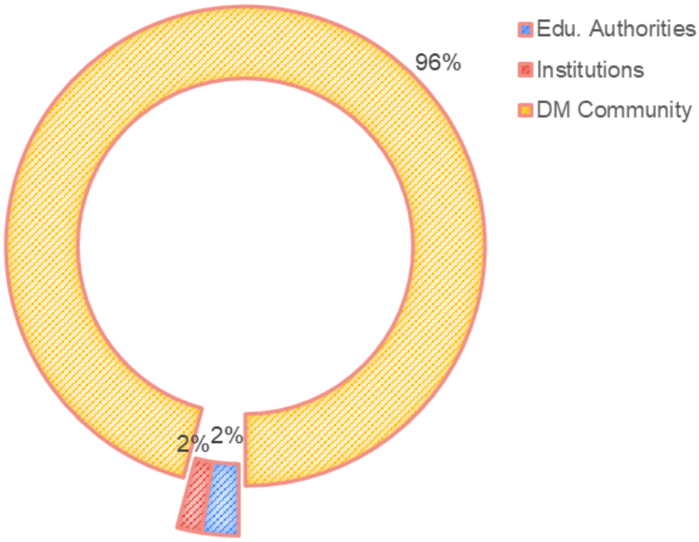
Table 2.5. Data analysis potential users.

Students	In order for students to increase their overall achievement, they frequently need additional learning options and activities. Students are provided with feedback as well as ideas, allowing them to modify their method of study or select alternative classes or sessions.
Teachers	The EDM can assist teachers and trainers in the development of differentiated teaching methods through the categorization of students, as well as in the development of novel teaching approaches. In addition to this, it is helpful in understanding learning from a cognitive, behavioral, and social perspective.
Researchers	The scope of the researchers is to evaluate the effectiveness of various techniques and methods in resolving various educational issues. They are also interested in developing new methods and instruments.
Education authorities	The goal of the administrators is to develop new educational means, improve curricula, and make more efficient use of the institution's existing resources. Evidence-based decision making (EBDM) is a useful tool that can assist with all of these goals by providing evidence to back up decisions.

The use of studies findings to the process of decision making can be of great assistance to educational authorities. Making decisions based on evidence can provide powerful decision support at any level, including class, school, local, regional, national, and even international. Our study focused on examples in which the results of research were incorporated in some way into decision-making or policy formation.We discovered that the great majority of them are research papers based on a tiny sample size. The researchers' primary focus was not on utilizing their findings to support decision-making or inform actors in educational policy but rather on determining the efficacy of various algorithms and methods. We believe that it is just as vital to emphasize and use the findings for decision and policy-making as it is to perform research to improve algorithms.

There were, however, a few instances in our research when web-tools were intended for educational institutions such as schools and universities. In one instance, we discovered that the outcomes, recommended algorithms, and complete methodology presented in the report were still being used in the decision-making process (Knowles, 2015). Five articles referred the development of tools that teachers can use. We also contacted to the authors to find out more about how their results were used. The feedback that was obtained is shown down below. Knowles's (2015) work is the most in-depth analysis. It

Figure 2.9. Research audience



pertains to the state of Wisconsin, which has one of the highest graduating rates in the country of United States, it also has a significant achievement gap between students. The development of an early warning system (DEWS) for the potential for student dropout was the main objective of this study. The program referred to 225,000 students across the state of Wisconsin. It analyses demographic, socioeconomic, and academic data by utilizing a wide variety of algorithmic methods. Using a package called ”caretEnsemble,” it educating a number of algorithms and selects the most effective one. The model's AUC ranged from 0.86 to 0.91. This article provides a detailed demonstration of a fully operational tool. The tool and method that was developed is currently being used in the state of Wisconsin, according to researcher.

In addition, the results were applied in two studies to minimize the failure rate, with concrete and quantitative results being achieved. The aim of the study that was published by Corrigan et al., (2015) aimed to predict students' achievement at the end of the semester. They accessed the information using the Virtual Learning Environment of Dublin City University (Moodle, 1,200 students). Utilized data included the number of logins per student, the mean time spent online, weekend use, etc. They use SVM and their findings were provided to the students as a weekly feedback in order to assist them in improving their efforts, which ultimately resulted in an increase of 2.67% in their final grade.

In the study conducted by Burgos et al., (2018), historical student achievement data was analyzed using logistic regression to predict students' success. Data for this study came from students enrolled at the Open University of Madrid. The data referred to scores on different tests and were obtained from the distance learning platform of the university. The findings shown that the logistic regression achieves an accuracy of more than 90%. This research is significant since it relates to several historical periods. Utilization of the study results in following school year resulted to a fourteen percent (14%) decrease in the number of students who dropped out of university.

The findings presented in the next two articles served as basis for the development of a Java application that may be used by educational institutions such as schools. this decision support tool was presented by Livieris et al., (2016). This tool was meant to predict students' success on year-end examinations. 14- to 15-year-old high school students enrolled in a private school in Greece were evaluated on their proficiency in mathematics. The implemented algorithms were: RIPPER, Naive Bayes, KNN, C4.5, BP and SMO. The accuracy varied from 51.6% to 70.3%. Following the implementation of ensemble strategies, the accuracy increased to 90.3% (using voting). This was a preliminary study with a limited number of participants. According to the authors, a Java-based tool to estimate students' achievement was developed but not used by the school.

Another Java application developed by I. E. Livieris, Drakopoulou, et al. (2019). This tool was based on secondary education data. Researcher explored the perspective of identifying students who performed poorly and in their overall GPA. They have access to data from 2,260 students who had completed the first two years of high school. The algorithms SMO, NB, RBF, JRip, C4.5, MLP, KNN, PART were used. The SMO achieved a higher accuracy, 91.14%. On the basis of the findings, a Java-based forecasting system was developed in order to help students take the necessary actions to improve achievement. We don't have any information about how these studies affect how well students do in school. According to researcher, the tool was never used by the school.

Finally, the academic achievement of students in Brazilian public schools is studied in Fernandes et al. (2019). Both demographic and academic data were used in this study. They used two datasets. The first included data taken in the week prior to the start of the course, and the second included data collected the next month. In the year 2015, a total of 238,575 entries were gathered and in the year 2016, this number increased to 247,297 entries. The Gradient Boosting Machines was the algorithm that was implemented. The study achieved a high level of accuracy after identifying the most important characteristics. According to the information provided to us by the first from the authors, the researchers state that one of their objective is the use of their methodology by the regional education authorities in Brazil. However, this is not currently possible.

As is evident from the small number of studies with a practical focus and from the previous findings of this review, there is a wide research area in relation to the use of data in education. This is the research area on which this thesis focuses.

2.4 Data-based educational policy

The use of data for policy-making is not new, but how and what data can be used for changes the way decisions are made on both a theoretical and practical level (Mandinach, 2012). Decisions based on evidence reduce uncertainty. However, as various sources of information compete in the policy-making process, the challenges posed by the use of data analytics techniques must be addressed, increasing the requirements of decision-makers to properly understand and use the data and conclusions derived from the analysis. The literature review showed a clear lack of connection between the academic - research community and educational administration on this issue.

It has been mentioned above that the use of data and analyses based on them have the potential recipients of educational institutions. The various administrative structures of the educational pyramid, such as schools, administrative heads, school advisors, ministries, or universities, are recognized as such. All of the above structures have one thing in common: they all need to make decisions in order to do their jobs and follow a certain educational policy. The way in which data is used and influences decision-making differs between management and policy bodies. To understand how the public sector uses the data collected, various approaches have been proposed, such as the ”Data Readiness Concept” and the ”Digital-era Governance” (DEG). These approaches are supported by e-Government and the introduction of technologies that transform public administration in a direction of greater responsiveness and accountability (Jetzek, 2016).

The concept of Digital-era Governance succeeds the concept of New Public Management and is a new example of public administration (Margetts & Dunleavy, 2013). The implementation of Digital-era Governance in the public sector is at a slower pace than in the private sector, which is linked to higher levels of digital literacy in new technologies and computers in general. It thus appears as a challenge to the acquisition of new skills by public sector education managers and their training. Alternatively, outsourcing to collaborators with the appropriate expertise is proposed. It is also necessary that educational authorities are able to process data and reach conclusions, which requires corresponding knowledge and skills.

The increased use of technology led to a change in organizations' learning patterns as it increased the quantity and quality of information. This leads to new challenges in the development of public institutions as ”learning organizations” in the handling, understanding, and exploitation of the resulting conclusions (Margetts & Dunleavy, 2013). Data readiness assesses their ability to be organizationally aligned with data use. The use of appropriate data for the specific organizational structure; the maturity of the organization for e-Government initiatives; its ability to use data; expertise in data science; and compliance with legislation are all examples of organizational alignment. Such readiness allows for the improvement of the entire value chain of data use, including the collection, processing, analysis, and use of data in decision-making or policy (Klievink et al., 2016).

The use of data has been seen as a driver of innovation in the public sector. Janssen et al., (2017) describe how private operators and citizens are pushing governments to renovate administrative structures and to develop new forms of data that will then be used in policy-making. Other factors in this direction may be strategic or organizational and be linked to data governance or be of a purely technical nature (Williamson, 2016). The important distinction between public innovation in general and data-driven innovation has to do with the source of the process. Data-based innovation does not necessarily come from public administration bodies but can be initiated by private organizations as well as citizens and lead to new organizational forms (Janssen et al., 2017). On the contrary, previous innovation efforts were directed from inside to outside. Janssen et al., (2017) note that a prerequisite for such a development is trust between the actors/interested parties. Policies that motivating cooperation between public and private actors can stimulate innovation.

The use of EDM in educational policy and management requires determining that will access, analyze, or review the data and for what purpose. People at all levels of the educational pyramid use the analysis, but for different reasons, as we can see in the table below.

• Educators. Most of the time, teachers use data to evaluate their students' needs, skills, progress, and achievement in the classroom. In this way, they adapt teaching. Teachers can also use data to reflect on their own strengths and weaknesses.

• Principals. A duty of school leaders is to evaluate how well the school is doing. This includes looking at how well the students and staff are doing as a whole. Achieving the school's goals requires managers to define and follow practices, programs, and actions. To this end, data and analyses are needed for teachers' contributions to students' achievements; for teachers' practices in the classroom; and for the school performance as a whole.

• Managers at regional level, use data for activities undertaken, which often involve assessing the achievements of students and staff. In a centralized system, these people are asked to use data to carry out education policy at the local level.

• Education managers. A main duty of employees of state education organizations is to carry out analyses that monitor the levels of achievement at the national level. Often, their work also includes references to federal agencies, regions, or other agencies. Also, specific qualitative and quantitative criteria must be used to confirm the quality of the educational work. Lastly, education executives are in charge of measuring and evaluating how pilot programs and activities have been developed in some regions or schools.

It is vital to describe the method followed by public entities in order to make decisions in order to integrate the use of data into the decision-making process by the public admin-

Table 2.6. Educational data users

Users	Students' data	Teachers' data	Curriculum	
Classroom teachers	Prior achievement of students and classes, Specific academic strengths and weaknesses	Added value. Teaching practices. Students' perceptions. Ways to improve teaching practice.	N/A	Specific
School principals	Dropout risk, specific academic needs of students	Teachers' achievement and professional needs. Needs for hiring teachers. School performance and improving.	Quality of curriculum implementation	
School boards etc.	Data for achievement and attainment. Satisfaction of students and parents. Enrollment rates in regional level.	Effectiveness of teaching, Principal performance and improving	Quality of within-school curriculum implementation. School value-added, program costs per student and school in regional level.	
State officials	Data for achievement and attainment. Satisfaction of students and parents. Enrollment rates in national level.	Effectiveness of schools in different objectives.	Quality of within-school curriculum implementation. School value-added, program costs per student and school in national level.	Aggregated

istration. This discusses the concept of a policy cycle, which is a broad term for the life cycle of a policy from its conception to its implementation. Lasswell (1956) established the concept of modeling the policy process in seven stages, known as: (1) intelligence, (2) promotion, (3) prescription, (4) invocation, (5) application, (6) termination, and (7) appraisal. An alternative concept often used has six steps, which are as follows: (1) agenda setting, that acknowledges the problem; (2) policy formulation, in which an option is introduced; (3) decision-making, in which the solution is selected and institutionalized; (4) implementation, in which the solution is enacted; (5) evaluation, and monitoring of the outcomes; and, in some cases, (6) the decision to keep, replace, or dismiss the plan. As a basic foundation for policy analysis, this model has proven quite successful. It serves as the foundation for a variety of policy process taxonomies.

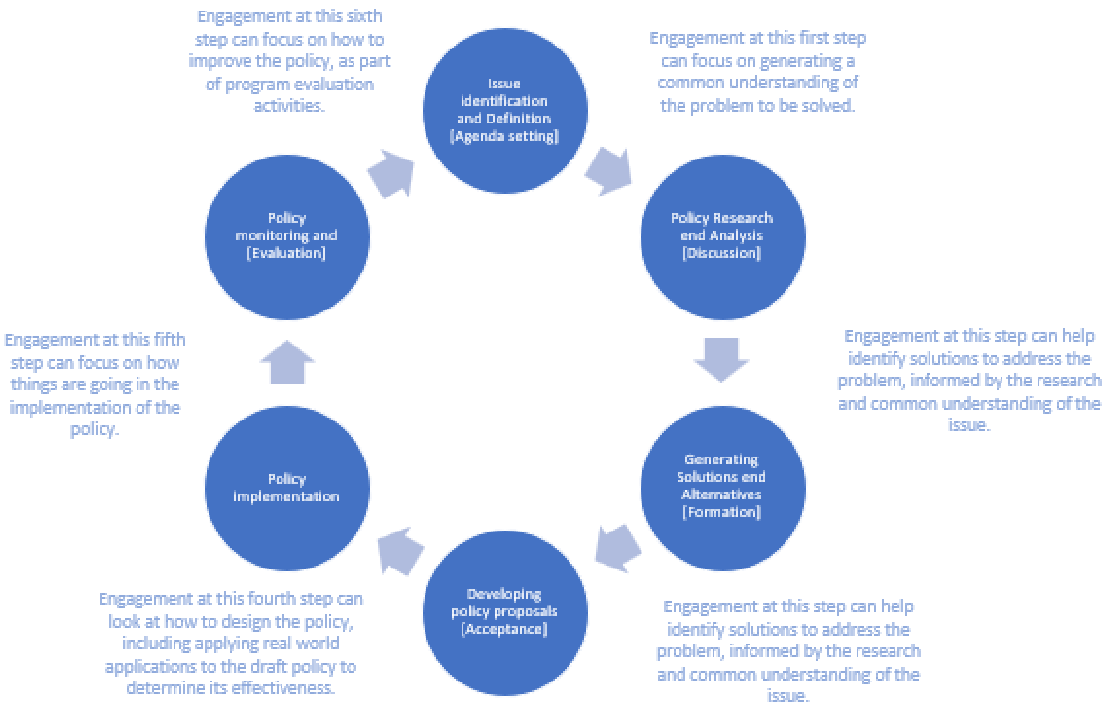
In the 1960s and 1970s, the expansion of political science resulted in the organization and systematization of a growing number of literature and research on the six-stage model. Following that, some typology adjustments were presented by Anderson and other scholars (B. E. Anderson & Wallace, 1975; Jenkins, 1978; Brewer & DeLeon, 1983). Policy scientists believe that there is a distinction to be made between creating an agenda, developing a policy, making a decision, implementing the policy, and reviewing (or terminating) the policy. But as mentioned, "The model is quite old and describes with great precision the depths of the political decision-making process" (B. E. Anderson & Wallace, 1975).

This model incorporates something that is common, if not a rule, in the decision-making of the public sector: the interaction between its stages. This interaction led to the development of criticism of the model as it was considered somewhat arbitrary to separate phases into a policy cycle. Also, on a theoretical level, criticism focused on the fact that theory overemphasizes the process rather than quality (Badie et al., 2011). Advocates of the approach argue that the focus on the process allows good policies to be developed while at the same time reducing the inherent complexity of the political decision-making process. The policy cycle is thus a theoretical depiction of policy implementation that is integrated into the governance process (Colebatch, 2005). The difference between the different stages gives meaning to the complicated policy-making process and makes it possible to use ICT and data in different ways during each stage.

The increased complexity and availability of data helps to implement evidence-based policies. Of course, the use of data is not a panacea. The application of such tools and subsequent scientific evidence through data does not lead to an automatic improvement in the quality of political decisions (Sanderson, 2002). Kogan, (1999) found that, although governments welcome the notion of evidence-based decision-making and identify it with the concept of legality, they often use evidence as they see fit and only if it supports specific and pre-decided objectives and policies (Kogan, 1999). Another problem is the question of the re-election of policies. If a politician's main goal is to get re-elected, he or she may use recommendations based on all kinds of evidence (Sanderson, 2002).

It is a common belief in the literature that the use of evidence in the decision-making process is combined with a rational policy model (Sanderson, 2002). Rationality means that the data or the results of their scientific processing form the basis of the decisions taken. This is evident by including the use of data in the feedback and evaluation process within the policy cycle. The new policy cycle model comprises the stages, together with an additional external feedback cycle. The first stage is the setting of the agenda, where problems are identified and the need for action is expressed. The next stage concerns the policy debate and aims to determine the correct way to address the problem identified at

Figure 2.10. Traditional policy cycle

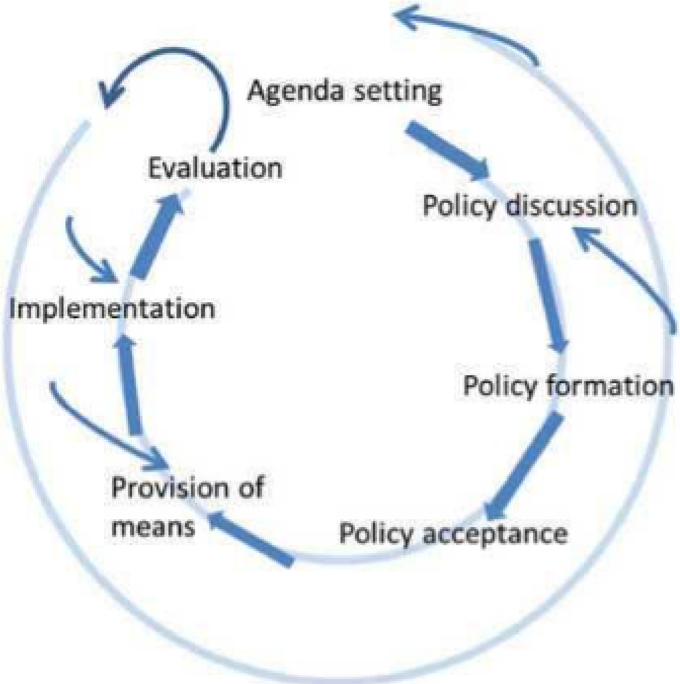


the previous stage. A specific policy option will be born out of a policy debate and the definition of how to address the issue. To carry out policies requires the provision of the means necessary for their implementation. It then follows the stage of implementation and taking the necessary measures to achieve the desired results. After completion and during implementation, it is important to evaluate the results. The evaluation will depend on whether the policy has been successful.

However, corresponding feedback and evaluation occur at each stage of the policy cycle. This evaluates the entire process, from the first to the last stage. This process leads to a cycle of improvement. Through an ongoing process of feedback, evaluation, and improvement, organizations are driven by changes as data is used to support and document positive or negative assessments and influence future behavior (Tresch et al., 2011)

The generated information and data are used as evidence in the decision-making process, which helps to improve the overall quality of the decisions that are made. On the other hand, this procedure is not carried out automatically. It has been demonstrated in the past that merely increasing the amount of information available might result in a decline in the overall quality of decisions (Keller & Staelin, 1987). This could be the result of an excessive amount of information, the failure of filters designed to decrease noise, or the development of complicated decision-making models that, in the end, are less effective. In order to eliminate this noise and uncover the crucial information beneath it, it will be necessary to employ advanced algorithms and techniques, particularly when dealing with

Figure 2.11. Data analysis and policy cycle



massive amounts of data. This section will investigate both the how and the why of data using throughout the cycle.

The key issue in setting the agenda is to identify the issues that will attract the interest of those responsible. In today's world, the choice of the agenda is becoming increasingly complicated. Online news sources have a different way of influencing the agenda, which is no longer necessarily mechanically linked to broadcast media news topics (Russell Neuman et al., 2014). Social media is now playing an important role, which has led to a reversal of the way the theme is shaped (McCombs et al., 2013; Berkovich, 2021). Public opinion is influenced by ideological, demographic, or cultural factors. The media and social networks now act as a link between the public and policy-makers. In fact, they provide a large quantity and variety of content, such as text, audio, video, and images.

The processing of such data enables governments to directly identify emerging issues and shape the agenda accordingly. The collection of data from social networks, characterized by a high degree of participation, allows governments to identify citizens' preferences and issues that they recognize as important and to take them into account when determining the issues to be addressed. But there is also a dangerous side to this practice. The attempt to influence and manipulate citizens' views through social media has also been documented in order to change their behavior (Lazer et al., 2014; Enli, 2017). Also, there is a use of technology to control public order in countries such as China and Singapore (Creemers, 2019). In these countries, there is a wide range of observation and recording of information on citizens' preferences in order to develop early warning systems for political unrest (G. King et al., 2013). A recent U.S. study confirmed that lawmakers are more likely to follow than to lead the debate on public issues, as they are more likely to respond to their supporters than to the general public (Barberá et al., 2019)

Policy debate is the next stage of the policy cycle where there is a discussion among stakeholders about the various options—actions. At this stage, data can play an important role in illuminating the details of the problems that need to be addressed. Specific priorities are set at this stage in relation to the order in which the various problems will be addressed. Their final ranking is the political priority set. At this stage, the various stakeholders are pressing policy-makers to prioritize their own claims (Lehmbruch, 2003). So, the role of data is very important because it can show which areas should be at the top of the list, using some objective criteria (Shum & Luckin, 2019).

However, the question arises of obtaining useful information from multiple data sources. Nowadays, public organizations have access to structured data from public administration databases as well as to a wealth of unstructured information from social media, blog forums, etc. Non-automated or ad-hoc monitoring of unstructured data sources is essentially unprofitable. Thus, the development of automated methods and tools, especially in relation to text mining, is necessary in order to retrieve and integrate useful information in the policy-making process. Over the past years, Alfaro et al., (2016) have shown that it is important to use automated methods. Sentiment mining and analysis, clustering, and other machine learning algorithms could help shape political discussions and tell policy-makers how public opinion changes in response to proposed changes (Kamateri et al., 2015).

The formulation of a policy requires a description of the actions to be carried out during the implementation phase. When a policy moves to the stage of framing, its credibility and legal basis must be obvious. In order for social groups and other stakeholders to support this strategy, it is necessary to supply them with evidence in the form of data. During the policy formation and acceptance stages, data, analysis methods, and scenario techniques can all help with evidence-based policy making (Harris & Williams, 2019). The decision-making process in government and administration is often marked by a high number of factors, variables, and differing target processes. The large number of independent variables and their different correlations pose challenges to the regression algorithms that have to operate beyond the least squares (OLS). Ridge, lasso, or elastic net regressions are only some of the techniques that can be used to develop models that support political decisions.

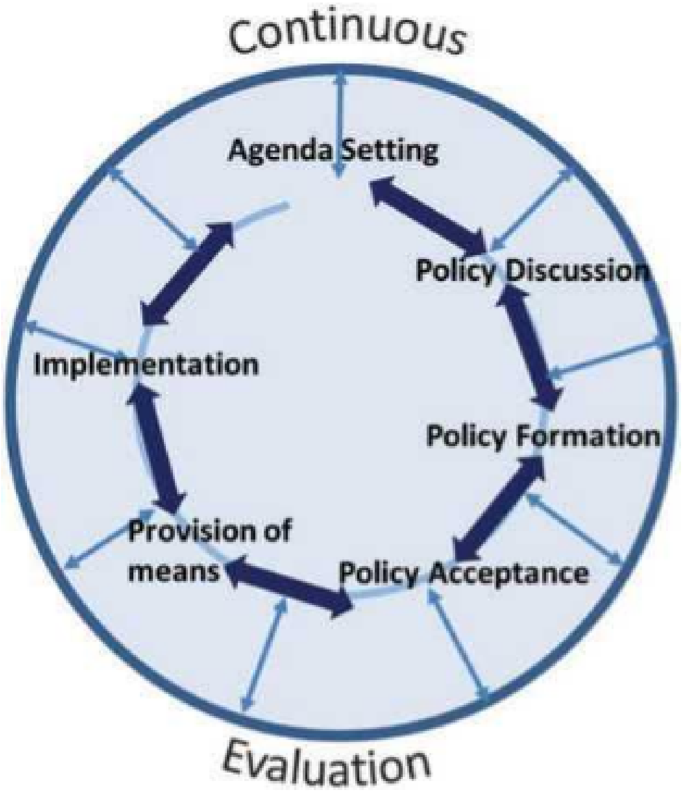
This stage refers to providing the necessary staff, legal, financial, and other resources. The use of data in this phase can be seen as an additional resource necessary for the effective implementation of the policy. There is empirical evidence that the use of data can increase efficiency and efficiency by reducing costs (Manyika et al., 2011). The availability of data and analysis techniques contributes to the development of pre-policy scenarios and facilitates the focus on expected results and the creation of policy self-assessment tools, creating a feedback cycle and can help identify and stop unsuccessful policies (Akanni, 2019).

The implementation of each policy is affected by the use of data. The very execution of new and recurring policies constantly generates new data, which then provides feedback on immediate results. This enables immediate (and long-term) evaluation of the effectiveness and improvement of future policy implementation through the immediate identification of problems and the necessary improvements. The most important improvement in the policy cycle is the new dimension that gives feedback. Direct data generation during application, rather than after, allows for great flexibility and adjusts the process itself. The directness of feedback allows for increased autonomy of public institutions at lower levels of the administrative pyramid, increasing flexibility and the ability to react promptly to the consequential results (Valle-Cruz et al., 2020). Another improvement is provided by developing the quality of data used by public administration. Because data is often collected over long periods of time, it is often out of date when it is made available to public services. This may not be appropriate for the development and implementation of new policies. The increase in data sources and the combination of different databases allows access to up-to-date data, as opposed to data that is updated only once or twice a decade. This requires careful processing and understanding of the data received so that variations in results are derived from real changes rather than from wrong estimates. The role of experts in every policy area is, in this case, very important. Cross-checking data from the public administration's databases with data from outside sources would be another way to improve the validity (Villa Alvarez et al., 2022).

This approach, by using data at each stage of the policy cycle, brings inevitable changes to the stages model. The redesigned cycle uses the data at all stages, gaining new possibilities as mentioned above. In contrast to the evaluation of policies after the final stage of their implementation, this formative approach provides evidence of bottom-up assessment and leads to a process of active stakeholder involvement (John & Parsons, 2005).

By theorizing and visualizing the approach, we find that the traditional cycle changes by including an external cycle of feedback and continuous evaluation at each stage. Instead of the evaluation being carried out at the end of the implementation of a policy, there is constant information supply for policy-makers at the previous stages of the process. An additional possibility is the use of machine learning tools in real time. This makes it possible to draw immediate conclusions. However, the use of all these tools requires attention as feedback depends on the stage of the process and may lead to fragmentary and erroneous conclusions and early abandonment of policies. A feature of the proposed policy cycle is the removal of the assessment phase from the end of the process and its inclusion in each other stage of the cycle as an integral part (Pencheva et al., 2020). The continuous evaluation approach combines traditional and formative evaluation. Thus, the continuous evaluation at each stage of the cycle formatting policy and is carried out throughout the process, while the final evaluation is essentially the assessment of the implementation phase. Increasing data generation makes such approaches possible by public administration and government.

Figure 2.12. Data analysis and continuous policy cycle



2.5 Discussion

Having presented the theoretical background of data use and the added value of data-driven policy, it becomes even more obvious that the small contribution of academic research to education policy is a picture that needs to be reversed. In summary, the literature review found the following.

The field of higher education is extensively discussed throughout the articles that we have reviewed. More research can be carried out in an academic environment as a result of an increase in the quantity of data that is readily available and the tighter contact that researchers have with academic institutions. In contrast, secondary and elementary education, which span a larger educational area and effect a bigger population of students, do not tend to attract the same level of study interest. At this level of education, there is great room for research due to the diversity of students' ages, topics, and learning levels, among other factors. The restricted access of the data, the processes for authorizing studies by authorities, the increased focus on students' personal data, etc. are all significant obstacles to the growth of further research.

A variety of available algorithms that enable researchers to evaluate the effectiveness of their findings in assessing student achievement have been found. Successful algorithms have been recognized. The majority of algorithms have been shown to have good levels of accuracy. In just a few instances did we observe accuracy levels that were either very low or extremely high. The use of complex algorithms did not result in any noticeable improvement. The majority of the studies have been achieved sufficient levels of accuracy make use of Decision Trees. Naive Bayes, C4.5, and Random Forest algorithms have increased use, but the KNN algorithm maintained its dominant position in the instance-based methods. Only three papers made use of the K-means clustering method. In fewer studies ensemble methods have been employed.

In the body of studies we reviewed, student grades were cited as a predictive variable the majority of the cases. It was common practice to combine grades with other data, such as demographic or academic information. There was no combination that resulted in a larger degree of accuracy, although the average level of accuracy achieved a sufficient level. Academic community was the most frequently mentioned audience in the papers that we reviewed. The bulk of examples were either tests aimed to evaluate the accuracy of methods or academics' ideas for new algorithms. On the other hand, the many different potential applications of such discoveries have been virtually completely overlooked.

In a few instances, a web application was created or study results were utilized to enhance instruction. More commonly, some hopes regarding the potential of applying the conclusions were expressed, although these statements were not unsupported by the facts. But from a more practical point of view, EDM can help education policy by providing it with the necessary evidence-based support. But, this objective has not yet been met. There have been only a few cases where it has been stated how the research findings can be applied in the real world. The only research that can be regarded realistic in terms of establishing methods for earlier detection of student fail and improving student outcomes via the feedback are those that examined the effects of feedback (Knowles, 2015; Burgos et al., 2018).

This review relied heavily on descriptive statistics. We have tried to give a full picture of a part of EDM that has to do with predicting academic success. After doing an in-depth study of the abstracts of the publications, we analyzed in full a large number of articles.In addition, there is a large amount of room for scientific study to be undertaken in a range of EDM subfields using various sorts of literature reviews.

In general, this review uncovered a sizable body of different types of research approaches, with a primary concentration on higher education. Education policy making and institutional decision making can benefit from the development of data mining methodologies. We propose that as an alternative, further study be conducted with the goal of incorporating findings into everyday teaching and educational policy decision-making. Internal scientific feedback from successful instances of the use of different algorithms and methods without real implementation might result in the downfall of a scientific area. On the other hand, the implementation of the results in real-world situations will expand the scope of study and will benefit both the academic research community and the larger educational community.

In our opinion, the most crucial for the future path for this discipline is to widen its reach outside university education and concentrate on lower education levels that cover a larger number of students, are more varied, and have a bigger influence on students' lives and society. This is the field's most promising future. It may provide new study possibilities, but more significantly, it can improve education and society.

The use of data analysis to primary and secondary education data can help meet the specific needs of the student population and provide a deeper insight into education systems. Methods that have proven useful in higher education can provide valuable information for teachers, administrators and educational authorities.

This thesis comes to offer the field the use of primary and secondary education data at the country level, which has not been done before. This can help in the process of data-based decision making and evaluation of the dimensions of the education system. It also highlights the important potential of using primary and secondary education data that is underused.

Chapter 3

Educational Data

Data use is a reality in business, science, technology and engineering. The growth of data use is considered to be equivalent to the industrial revolution (Richards & King, 2014). Others insist that virtually nothing has changed except the increase in its volume. Despite the opposite view, it is a common belief that the widespread use of data analysis has a catalytic effect on science, technology and politics, supporting or dispelling theories and assisting in decision-making (Williamson, 2016).

The use of data for policy-making is not new, but the way and possibilities of using data are changing the way of decision making at a theoretical and practical level. (Mandinach, 2012). Evidence - based decisions reduce uncertainty. However, the challenges posed by the use of data analytics techniques need to be addressed, as various sources of information compete in the policy making process and increase the requirements from decision makers to properly understand and use the data and conclusions derived from the analysis.

3.1 Data in Educational settings

The increased availability of educational data over the past decades has resulted from the widespread introduction of information technologies into education and has improved the way measures are taken and educational policies are studied, monitored, perceived, and evaluated (Laney, 2001). At the same time, the analysis of educational data can provide additional evidence in decision-making, thus leading to improved educational efficiency (Ward & Barker, 2013). In parallel, there is an increase in the computational resources necessary for the processing, analysis, and interpretation of the results. The sources from which the data originated in educational settings derive from the administrative structure and the learning process. Both sources provide important data and have their own contributions to educational research.

The introduction of new technologies into education has led to an increase in the volume of data through two channels. Firstly, it increased the recording and storage of data. Educational organizations now accumulate huge amounts of information about students. Information systems record and store data on students' profiles, e.g., demographic characteristics, academic background, courses followed, and scores. In some countries, data is now collected over a ten-year period, which makes it possible to look at the data over time. Secondly, it developed the learning management systems (LMS), which also records elements of students' behaviors, which is difficult to do in classrooms. Teachers are often able to use LMS to distribute educational materials, monitor tasks and communicate with students. This creates a large amount of data that includes: clicks on course modules, task scores, log files, and more, for each individual student. Beyond MIS and LMS, the introduction of innovations in education, such as innovative digital learning environments, provides additional possibilities. These innovations offer new teaching opportunities and, at the same time, collect data from the students they are teaching. Because there are so many different sources, there are a lot of different kinds of data (Kokkinos, 2019).

Data is a source of latent knowledge. Questions that were difficult to ask before the development of these data sources can now be answered. Big data, because of its volume, speed, and variety, can provide high-value inferences. The findings are applicable to both educational research and educational decision-making. The possibility of examining digital traces of students' actions allows for a more nuanced understanding of learning processes. The abundance of material provides insight into the impact of certain educational initiatives and helps to evaluate them. For example, studying student differentiation in learning using secondary administrative and learning process data can reveal educational inequalities. Personalizing the learning experience according to the learner's abilities and interests can significantly improve the learning experience and achievement. Data can now support or guide educational policies and decisions, which can then be evaluated through faster feedback loops. On the other hand, there are concerns about the quality and relevance of the data. Furthermore, it appears that some educational data, such as religion, falls into the category of sensitive information. As a result, a significant number of government regulations have been enacted.

The ways that data analysis is used in education can be broken down into three levels based on the educational settings in which they are used: the micro level, which includes things like clickstream data, the intermediate level, which includes things like text data, and the macro level, which includes things like administrative data. Micro-level data are captured interactions between actors. Most microlevel data points are automatically collected through the interaction of students and their learning environments. Such environments can be educational games, intelligent teaching systems, massively online open courses, and simulations. Mid-level data mainly includes texts and students' answers to questions during work writing in a digital learning environment or other activities, e.g., participation in social media forums or social media interactions. The recording of this raw data helps to draw conclusions about the evolution of students on a cognitive, social, or emotional level. The majority of macro-level data is collected at the institutional-administrative level. Macro-level data includes demographic characteristics of the students, courses they attend, data from the scores, etc. Long-term macro-data is generally collected for several years in several countries and is not updated frequently. This categorization should not be considered as absolute, as there may be a significant overlap between the data sources. For example, the written essays provide both intermediate-level text data, but the resulting scores are macro-level data.

Figure 3.1. Educational Data and Educational Pyramid

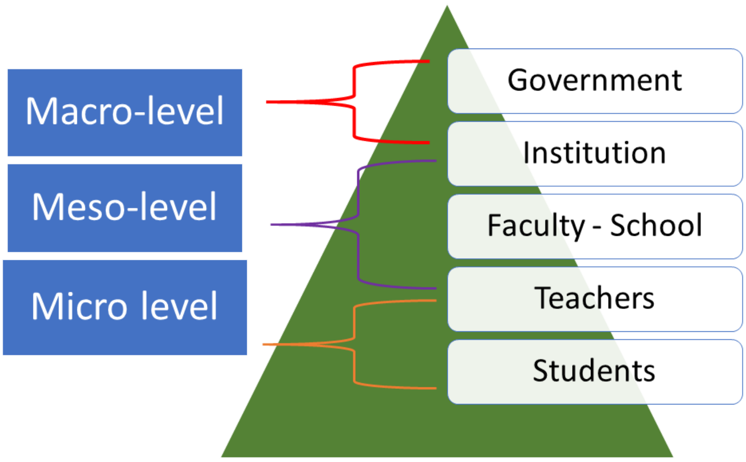
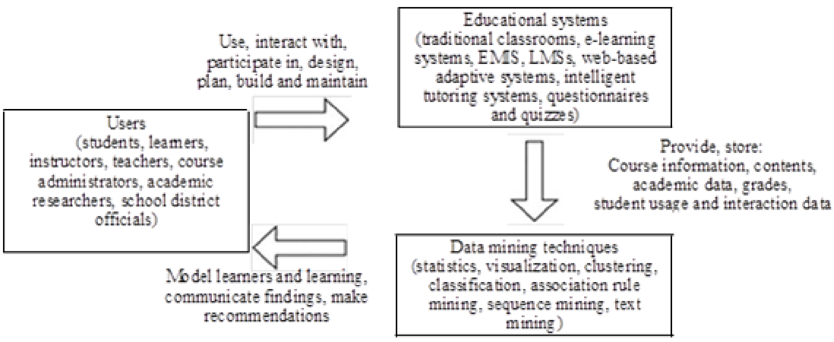


Figure 3.2. Educational Data sources, analysis and users



Micro-Level Educational Data

Exchanges between students and asynchronous learning platforms like MOOCs, intelligent teaching systems, simulations, and educational online games generate the majority of micro-level data. This type of data includes information on the actions of the learner and the context in which they are carried out. Usually, a few hundred students are examined, but the volume of data is quite large and can reach millions of enrollments. The nature of microlevel data makes it appropriate when immediate intervention is needed, e.g., providing feedback or moving students to a new level when they are ready. Micro-level data are sometimes validated on the basis of real-time observations of student actions (DeFalco et al., 2018; Botelho et al., 2019) or retrospective coding of subsets (Gobert et al., 2012) at the research level.

Some examples of studies that used micro-level data to analyze and understand the components of knowledge. For example, Toth et al., (2017) studied problem solving within the ”MICRODYN” learning environment and grouped how students' strategies were developed and changed over time. Bauer et al. (2017) examined problem-solving approaches in the scientific game ”Foldit” using visualization of clickstream data generated by the game. The authors identified several common problem-solving strategies and combined them with player performance. Readability and self-regulation often examine the student's ability to self-regulate learning processes (Roll & Winne, 2015), behaviors that are particularly important in less structured systems such as LMS and MOOCs. Park et al., (2018) developed a time management measure that identifies the late submission of students to online courses and identified three levels of time management skills. The researchers also used micro-level data to develop models (Fancsali et al., 2018), which use students' help to understand how students transition between using different learning resources.

An understanding of students' emotional states inspired work to identify the impact on different learning environments, such as intelligent teaching systems, game puzzles, and simulations (Botelho et al., 2019; Hutt et al., 2019). Finally, a widespread application of microlevel data is the assessment of students' knowledge based on their responses to problems.There are three popular methods. The Bayesian knowledge tracing (BKT), the Performance Factor Analysis and the Deep Knowledge Tracing (DKT) (Khajah et al., 2016). These methodologies estimate the probability of obtaining a degree based on students' skills. The increasing availability of datasets with platforms creates large datasets, which in turn promotes research. Micro-level data has been used in teaching in a variety of ways. Micro-level data is frequently massive; a student can provide thousands of thousands of data points. Data analysis allows for the study of these phenomena that occur in a matter of seconds (DeFalco et al., 2018; Botelho et al., 2019) . The collected data and conclusions can then be utilized to examine behavior over extended time periods (Slater & Mudryj, 2016). Micro-data analyses and models are quite simple to apply to learning interventions. Because are easier to acquire, even as case studies, substantial attention is paid on them (Papadogiannis et al., 2020), Too much focus on micro data is likely to have medium or long-term consequences. For example, the use of analysis solely for the immediate prediction of performance raises concerns about the long-term effectiveness of the models developed (Pardos et al., 2019) .

Meso-Level Educational Data

Meso-level data is primarily associated with written texts. As academic writing is nearly entirely done digitally, the availability of such data is growing. Text data can come from a variety of sources, including internet forums, website databases, intelligent teaching systems, code programming, and many more, in addition to courses and research work. The time it takes to acquire data can range from minutes to hours. The frequency and consistency with which actors participate in such activities varies. As a result, students may submit texts on an LMS every week or fortnight until they complete a module and then cease. As a result, their use of social media may shift during the course of the program (Romero & Ventura, 2010).

NLP (Natural Language Processing) is used in basic techniques to text data analysis to automate analytical procedures. The study focuses on the vocabulary, syntactic, and morphological aspects of the students' texts. Meso-level data research studies provide information on (a) cognitive processes (e.g., cognitive function, knowledge, and skills), (b) social processes (e.g., cooperative structures), (c) behavioral processes (e.g., student participation and disengagement), and (d) emotional processes (e.g., emotion, motivation) (Crossley & Kyle, 2018; Adrian et al., 2020).

Evaluation of cognitive function

Cognitive process research has focused on assessing and assisting learners' cognitive function, knowledge, and skills, as well as offering assistance to trainers (e.g., automated pupil feedback, automated job rating). With the using of large number of student writing samples, the potential for automating student rating increases. These studies often use text samples from hundreds or thousands of students to illustrate that assessing students' writing may be automated, greatly reducing human labor in essay grading (Head et al., 2017; W. R. Allen et al., 2018) . Such research focuses on mathematics (Adrian et al., 2020) and students' linguistic skills (Crossley & Kyle, 2018) . In addition to task assessments, support and feedback mechanisms for students have been established to help them learn in a variety of ways. For example, used an algorithm to deliver hints to programming students (K. Wang et al., 2018).

Study of Social Procedures and behavior

Other studies examined social processes by analyzing dialogues and collaboration patterns from online forums, intelligent teaching systems, and video transcriptions. These studies can involve thousands of students and millions of interactions. Hecking et al. (2016) studied data from MOOC discussion forums and discovered that social and semantic structures influenced interaction patterns and community building processes. They demonstrated that various forms of blended learning allow students to collaborate. Scheihing et al., (2017), for example, studied disparities in student interaction patterns using a micro blogging platform.

Behavioral Engagement Detection

Other studies examined student involvement and resource-seeking behavior, frequently utilizing hundreds of thousands of interactions from thousands of students. Epp et al. (2017), for example, studied communication behavior in online interactions, with a focus on the usage of personal pronouns. They discovered that students who were more engaged in class interacted more and used more personal pronouns, whereas students who struggled with classes were less engaged and used fewer personal pronouns. Joksimovic et al., (2015) studied MOOC student engagement patterns and their associations with Twitter, Facebook, and blog dialogues. They found that the topics discussed on all social media platforms were similar, and that the most pertinent topics arose rather quickly.

Emotional Processes studies

Emotional processes study the self-awareness and motivation of hundreds or thousands of students responding with learning opportunities. Crossley et al., (2018) , for instance, utilized NLP techniques to identify student capacity declines in an online education setting. Allen et al., (2018) also utilized NLP to infer essay writing characteristics and link them to emotional states of engagement and boredom. Crues et al., (2018) studied if the responses to open-ended questions on students' expectations during MOOC enrollment procedures and their relationship to age and gender identified 26 reasons for enrolling in the course that were related to the students' age but not their gender.

Text data can provide clues for students' understanding of their views on various topics and even their emotional impact. This data can also provide information about relationships and networks within an online community. Studies using text analysis can help trainers design courses and activities to improve student participation (Atapattu & Falkner, 2018). However, the applicability of various tools has not been extensively tested in all educational environments (Fesler et al., 2019). Researchers cannot ignore context factors, such as the stimuli to which students respond, because this may lead to inaccurate conclusions. These errors are particularly dangerous when associated with significant results, such as student grades.

Macro - level Educational Data

Compared to other levels, macro-level data is acquired slowly over extended periods of time. For example, university data comprises demographic and student entrance data, course enrollments and grades, course schedules and course descriptions, information about the degree and relevant prerequisites, and statistics on campus life. In the primary and secondary education information systems, corresponding data are collected. This data is obtained normally every few weeks and only twice every cycle (Williamson, 2016). Some student information, for instance, is typically collected only once and updated at the student's request. However, such information may empower leaders to engage in data-driven decision-making and enhance the school's efficacy. This section focuses on three common application domains for macro data that have developed in the literature: (a) early warning systems; (b) guidance and information systems; and (c) management support analyses.

Early warning systems

The detection of warning signs for students who are at danger of dropping out or quitting a course is crucial for educators and education administrators. The availability and exploitation of administrative data made the development of data-based early warning systems possible and permitted the creation of forecasting models (Baker et al., 2016). Several research strive to build early warning systems in real-world situations, albeit with limited sample sizes and a focus on higher education (Papadogiannis et al., 2020). The objective is the same: to improve students' academic achievement (Chaturapruek et al., 2018). Initial applications are utilized by early detection management systems to forecast student failure. Based on the LMS system's data, a study utilized demographic information and exam results to develop a system that estimated the likelihood of a student failing a course (Jayaprakash et al., 2014). The purpose of the study was to design a system of early intervention to increase the likelihood of student achievement; the results were contentious. A statistically significant rise of 2 to 5 percentage points was noticed in the average score. The subsequent chapter will provide a more in-depth analysis of this field (Papadogiannis et al., 2020) .

Guidance Systems

Course selection and guidance systems are complementary to early warning systems. They aim to help higher education students choose courses that suit them. One such example from Berkeley University is the AskOski system, which uses historical elements and machine learning to suggest courses related to students' interests while linking them to the requirements of obtaining a degree (Pardos et al., 2019). Also, Stanford's CARTA system uses multiple sources to assist with course selection (Chaturapruek et al., 2018). However, undesirable effects from the use of CARTA were also observed, resulting in a quarter-reduction in the GPA of those who used the system. These results show how important it is to know how administrative data and scores affect the choices and success of students.

Management-Facing Data Analytics

This approach provides a systematic framework for quantitative analysis and visualization of curricula, providing information on education. Education managers can be helped by data analysis in many ways. The use of simple visualization techniques was proposed by Méndez et al., (2017) and provides indications of the prospects of the curricula. Through the analysis of grades in many courses, Méndez et al., (2017) drew conclusions from course data in a computer science program. At the same time, they were able to identify patterns that led to the abandonment of studies through sequence analysis. Matz et al., (2017) tried to describe the student experience using macro-level data. They have shown that even limited administrative data can lead to a wealth of information. With a similar methodology, Mahzoon et al., (2018) created visualizations to describe the academic courses of students and to see their results. Finally, Davis et al., (2018) used data from 177 MOOCs to draw conclusions on common learning patterns in different courses. Papadogiannis et al., (2021c)used a large dataset containing all Greek students' marks to extract student achievement levels in primary and secondary education in Greece.

The applicability of various data analysis techniques, from simple to advanced, to macro-level data is a given. Educational organizations and universities have systematically collected data on student achievement, demographics, and other topics in recent years. However, data was rarely used for institutional reforms or for improving decision-making. By analyzing that data, schools can substantially improve their effectiveness. On the other hand, public access to the results of such analyses can also improve social accountability. The data provided to date is not complete and is rather fragmented (Papadogiannis et al., 2020). Open access to anonymized data and analyses can have a democratic effect, giving all students equal access.

However, this data may be subject to various restrictions. Initially, data is collected from different sources and systems, making it an invitation to preprocess. The main reason for diversification is the uniqueness of institutions. That uniqueness requires particular attention, as the application of the same analysis to all schools or institutions may lead to unreliable results. The data may not be sufficient for the purposes of the analysis. If, for example, the aim is to draw conclusions on the employability of graduates, administrative data must be combined with employment data. The inclusion of multiple data sources, such as employment records or graduates' social activities, can lead to better results, but at the same time raises concerns about the use of personal data. Finally, the extent to which the results of analyses are used is uncertain. It has also been reported that these interventions may have unintended consequences for the behavior of students (Chaturapruek et al., 2018).

3.2 The data used in this thesis

3.2.1 Data collection

This thesis aims to examine and formulate objective opinions on dimensions of the educational system as a whole as well as specific educational interventions. For this purpose, data were used that are related to the meso and macro levels. These data are available to educational authorities, from the level of school principals to the level of the country's Ministry of Education officials.

3.2.2 Students' data

The education system in Greece consists of three distinct levels: primary education, which lasts for a total of eight years and includes kindergarten and elementary school (ES); secondary education, which lasts for a total of six years and includes Junior High School (HS) (Gymnasium), Senior High School (SHS) (Lyceum), each of which lasts for three years; and Higher Education. The educational process, at all levels, is constantly generating data. This data concerns the educational and organizational levels. Prior to 2015, data on primary and secondary education were not collected in a systematic way. There was no easy way to combine the data because different applications were used for different parts of educational work. During the 2015-2016 school year, the Ministry of Education in Greece launched a new integrated Educational Management Information System (EMIS) known as ”my school”. The EMIS holds all data on students attending all 14 levels of primary and secondary education. School principals across the country are responsible for entering the data. Here are some examples of the types of information included in this EMIS data:

- Demographic characteristics of the student population, such as age, gender, profession of parents, nationality, religious affiliation, etc.

- Data on the student's academic achievement, such as grades per subject, attendance in class, and behavior.

- Data on the teaching staff, such as their contact information, the classes they teach, the years they've been in the school, the number of hours they teach each week, the qualifications they possess, and so on.

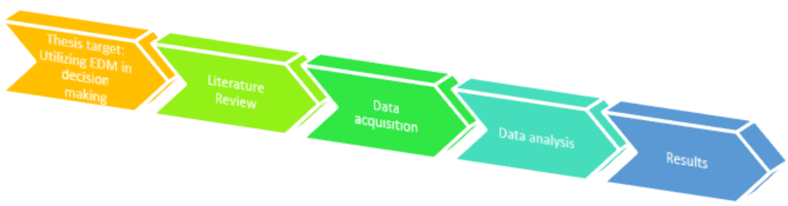
- Data regarding the educational institutions, such as their address, contact person, facilities and resources available, additional requests for teacher's recruitment and any other relevant information.

"My school" EMIS is the only tool that can currently support the exports of statistical reports and data about all students in the educational system in Greece. While the amount of data being collected grows all the time, new study opportunities emerge. To this day, access to these data is restricted to staff members working at various administrative levels within the education system. These staff are able to view or edit different data depending on their role, but they can do so only through the interfaces and predetermined reports offered by "my school".

In order to gather the data for analysis we requested student demographic and academic achievement data from the Greek Ministry of Education. The dataset covered the years 2016-17 to 2018-19. As students' scores were a key part of the data set, numerical scoring data were requested, which is used in the last two years of primary school (grades 5 and 6), as well as in junior high school. Furthermore, a dataset of demographic characteristics of students that exist in the Ministry of Education's EMIS was requested. These characteristics include the city and region of residency, the occupation of the guardians, nationality, religion, and others.

The analysis of the data started after acquiring and importing them into a new SQL database. Several steps are involved in the process. Preprocessing was certainly required (removal of noise or outliers, transformations, etc.). Then, using clustering, interesting patterns were retrieved, which had to be analyzed in order to derive conclusions. The following diagram depicts the flow of the current thesis.

Figure 3.3. Data analysis as procedure



The dataset

Data has been provided for the past three years in a row, as has already been mentioned. In point of fact, two separate subsets of the data have been presented for the years spanning from 2016–2017 to 2018–2019. The first dataset refers to fifth grade of elementary through first grade of junior HS, and the second to the three grades of junior HS. All of the general schools, as well as all of the music schools and all of the art schools, are represented in this data. It is easy to make comparisons between them due to the fact that the three distinct types of schools all cover the same subject matter during the relevant grade levels. No special education schools were included in these datasets.

- The students who began the 5^{th} grade of ES in the year 2016 are included in the first subset of the dataset. This subset follows these kids as they move on to their 1^{st} year of HS.

- The students who started the 1^{st} grade of HS in the year 2016 are included in the second subset of the data set, which follows those students all the way through the third year of HS.

When students are analyzed on a national scale, it is reasonable to expect that not everyone will have the same path in their education. There are some students that decide to leave school. Others arrive from other countries and start their education in Greek at a class that is appropriate for their age. There are some students who are required to take a class more than once, either because they did not maintain an adequate attendance rate (for instance, if they missed a significant portion of the school year due to medical reasons) or because they did not achieve satisfactory academic results (dropout).

The Structure of the Dataset

The dataset contains a subset of the demographic information that can be found in the ”my_school” database. Only a portion of the demographic attributes in a specific time point has been made available to us at this time. The data we had covered a three-year period, which we examined in this study. To put it another way, we do not know whether any of that information has changed during the course of the three years that we examine in this work.

Table 3.1 provides a summary of the demographic characteristics that can be considered in relation to this study. The Student_Id field requires a special note because it is not originally located anywhere in ”my_school,” nor is it one that can in any way be connected to a particular student. Due to the fact that the data has been anonymized, the ministry has added a “fake ID” before exporting it so that utilizing it; we will be able to keep track of the same student during the course of the three years. Other factors that can be considered are the student’s gender, the location of the school, and the jobs held by the student’s parents (or whoever is the legal guardian).

Additional data is summarized in Table 3.2 for each student: class, general grade, and absences. The information comprises the student’s Grade Point Average (GPA), which is computed according to the requirements that the Greek legislation sets for each grade, as

Table 3.1. Demographic data of the dataset.

Element	Type	Description
Student_Id	Character	Fake Student ID
Region	Text	50 counties + (6 regions of Attika and 2 regions for Thessaloniki)
Gender	Boolean	Male/Female
Guardian Occupation	Text	Free text (As completed from school, in Greek language)

well as the number of absences the student has had throughout the school year. There is no information that can be provided regarding the pattern of absences that occur throughout the course of the year.

Table 3.2. Grading data data of the dataset.

Element	Type	Description
Class	Character	E, ST (ES) and A, B, C (HS)
GPA	Numeric	Average score (according to Greek law), with an accuracy of two decimals
Number of absences	Numeric	One per day (Elementary) and Seven per day (Junior HS), without decimal digits

Also, specific data regarding the grades that the students have earned in each individual course topic is also provided, as can be seen in Table 3.3

Table 3.3. Lessons grades.

Element	Type	Description
Lesson	Text	ES and HS Courses
Lesson Score	Numeric	Numerical scoring, 1–10 (ES) and 0–20 (HS) without decimal digits

As expected, the list of accessible courses differs between the various types of schools and programs. The courses that are taught in the 5^{th} and 6^{th} grades of ES are summarized in Table 3.4 and the courses that are offered in the first three years of HS are summarized in Table 3.5.

Table 3.4. Lessons taught in ES

Lessons 5th and 6th grade
Greek Language
Geography
Social and Political Education
Religious Education
History
Mathematics
Physics
English Language
Second Foreign Language
Computer Science

Table 3.5. Lessons taught in HS

1st Class	2nd Class	3nd Class
Ancient Greek Language	Ancient Greek Language	Ancient Greek Language
Greek Literature	Greek Literature	Greek Literature
Greek Language	Greek Language	Greek Language
English Language	English Language	English Language
Religious Education	Religious Education	Religious Education
History	History	History
Mathematics	Mathematics	Mathematics
Home economics	Computer Science	Social and Political Education
Computer Science	Technology	Computer Science
Technology	Physics	Technology
Physics	Chemistry	Physics
Biology	Biology	Chemistry
Geography	Geography	Biology

3.2.3 Dataset cleaning

The fact that datasets of this magnitude are nearly never completely error-free should not come as a shock. The case of our data is not an exception to that rule. To begin, there are a few grades that have not been filled in. In some cases, this is because some students do not follow all courses (religious education is an example of a course that a number of students don't attend), or in other cases, this is because of mistakes that were made during the data entry process. In order to clean up the dataset, specific rules were needed to fill or delete missing values. In relation to the students' grades the following rule was followed.

- When only a subject grade was unavailable, we estimated what that grade would have been by taking the average of the grades from all of the other classes.
- When there were instances where multiple grades were absent, the entire record was removed from the dataset.

- In addition to missing grades, there are instances of data that do not make sense (for example, grades that are off the scale), which are caused by data entry errors. Because there was inaccurate data validation for the first school year (2016-17), this issue is limited to that year. Non-reasonable entries received the same treatment as empty entries.

In relation to missing values in demographic data, there are also instances where the data is lacking. A significant percentage of empty entries were found in the field of the parent's or guardian's occupation. Also, in a number of cases, 'Not answering' or 'Unknown' was entered. This problem arises as no control is applied to this field of EMIS. In contrast, the gender and region fields were fully completed as the data entry was done in the form of a list (gender) or automatically, based on school (region). In relation to the students' demographic data, the following rule was followed.

- If the empty entry was only for the field "guardian profession', the entry was accepted with corresponding handling only in the analysis on professions.

- If the empty entries were found in more than on demographic field, the entries eliminated.

Since one of our goals is to analyze the development of children's achievement from one grade to the next, only the records in the data set that correspond to the development of their age group in all three years were kept. This resulted in us having records for 85680 distinct students in the first subset of the dataset (Table 3.6), and records for 85344 distinct students in the second subset of the dataset (Table 3.7). This is the largest dataset that has ever been analyzed for elementary and secondary education in Greece with data analysis, and it is likely one of the largest datasets ever analyzed for students of these ages.

Table 3.6. Records from the 1st dataset before and after data cleaning.

Class	Initial Records	Final Records
5th Elementary School	101,644	85,680
6th Elementary School	104,559	
1st High School	111,785	

Table 3.7. Records from the 2nd dataset before and after data cleaning.

Class	Initial Records	Final Records
1st High School	96,359	85,344
2nd High School	99,431	
3rd High School	100,943	

After data cleaning, two new datasets were generated. The final two grades of ES and the first year of HS were included in the first dataset. In this way, the transition from elementary to secondary education is also examined. The second dataset contained the grades of students from all three grades of junior HS. After the X-means algorithm was applied, one new variable representing the achievement cluster were added to the original dataset. Consequently, it was feasible to make use of statistical methods in order to provide answers to the research questions. In particular, the following aspects of student achievement were studied: (a) the relative frequency of each achievement level; (b) the longitudinal stability of achievement cluster frequencies; (c) the differentiation of the average score (GPA) per cluster and its statistical significance (using non-parametric tests (Kruskal Wallis) due to the lack of homogeneity in variables); and (d) the effect of some demographic features on student achievement, the features being the profession of guardian, gender, and region of residence. The"$\chi 2$" statistical test was utilized for these studies.

3.2.4 Adding more data sources

In order to add characteristics that were missing from the dataset, an attempt was made to obtain economic data related to the students' area of residence, as this characteristic provided to us (nuts 3). These data were collected from the National Statistical Authority of Greece and related to the gross product of the area as well as the gross per capita product of the area. It should be noted that it was not possible to select the region using more stringent criteria, therefore, we were forced to accept the existence of heterogeneity in these characteristics. The data concerned the corresponding years to those used in the analysis.

In literature recorded a positive correlation between income and academic achievement, there is also evidence to suggest that this relationship may have become weaker in recent years. A study by Reardon et al., (2016) found that the income-achievement gap in the United States has remained relatively stable over the past 50 years, but that the correlation between income and achievement has weakened in some areas, particularly in urban areas. Similarly, the study of Watts (2020) found that while there is still a positive correlation between income and academic achievement, this relationship has weakened in recent years, particularly for low-income students. The study suggests that other factors, such as access to high-quality educational resources, may be more important predictors of academic success for these students.

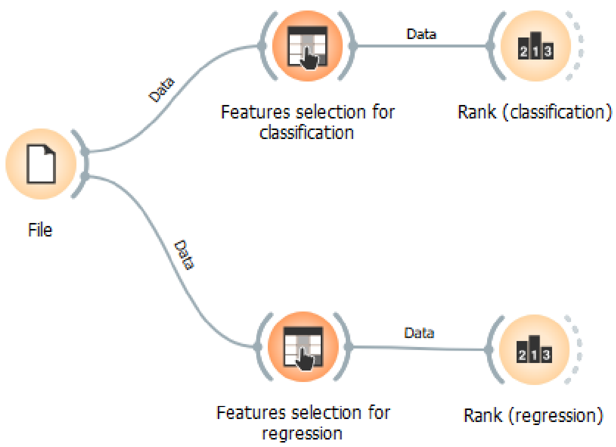
For examining the importance of these characteristics, linear correlation with GPA of the three grades was first examined (for presentation purposes only the case of high school dataset is presented). It is observed that there is no correlation and therefore no direct explanatory power is expected for these variables with respect to the students' grade.

Table 3.8. Correlations - Pearson's r

Variable	GPA_A	GPA_B	GPA_C	income_2016	income_2017	income_2018	GDP_2016	GDP_2017
1. GPA_A								
2. GPA_B	0.94							
3. GPA_C	0.899	0.948						
4. income_2016	-0.052	-0.042	-0.038					
5. income_2017	-0.051	-0.041	-0.038	0.999				
6. income_2018	-0.054	-0.044	-0.04	0.998	0.999			
7. GDP_2016	-0.048	-0.041	-0.046	0.536	0.537	0.539		
8. GDP_2017	-0.048	-0.041	-0.046	0.535	0.536	0.539	1	
9. GDP_2018	-0.048	-0.041	-0.046	0.536	0.536	0.539	1	1

Using data analysis techniques, we can divide the attributes into methods that rank them based on their explanatory power in classification or regression. A direct approach is provided by the Orange "Rank" widget. The Rank widget scores variables according to their correlation with discrete or numeric target variables, based on some scorers (like information gain, chi-square and linear regression) and methods that support scoring, such as regression, classification, etc.

Figure 3.4. Features selection in Orange platform



The scoring methods used for classification techniques are:

Information Gain:

A feature selection is a popular technique for selecting relevant features in classification tasks. It is based on the concept of information entropy and calculates the amount of information gained about the class variable by knowing the value of a particular feature. It is a fast and computationally efficient method that is suitable for large datasets with many features. However, it has some limitations such as favoring features with many distinct values or categories, which can lead to overfitting. In recent years, extensions and variations of information gain have been proposed to overcome these limitations, such as gain ratio and chi-squared. Additionally, conditional information gain and mutual information have been developed to capture interactions between features (Yu & Liu, 2003).

Gain Ratio:

A modification of information gain that adjusts for the bias towards features with many distinct values by taking into account the intrinsic information of the feature itself. It has been used in many applications, such as bio-informatics, text classification, and image analysis, and is a more robust method than information gain. However, it can be computationally expensive for datasets with many features and may not be suitable for datasets with highly correlated features. Despite its limitations, gain ratio is still a popular and effective feature selection method for many real-world applications (Li et al., 2018).

Gini:

The Gini method for feature selection is a popular technique for selecting relevant features in classification tasks. It is based on the concept of the Gini impurity index, which measures the probability of misclassifying an instance based on the distribution of the class labels in a particular subset of the data. It is a robust method that is not affected by the scale of the features or the number of classes, and has been shown to perform well in many real-world applications. However, it has some limitations, such as favoring features with many distinct values or categories, and may not be suitable for datasets with highly correlated features (Li et al., 2018).

AN OVA:

ANOVA (Analysis of Variance) is a statistical method used for feature selection in classification tasks. It is based on the concept of the F-statistic, which measures the variance between different classes relative to the variance within each class. ANOVA has been used in many applications, including bio-informatics, image analysis, and social media analytics. It can handle both continuous and categorical data, and can handle datasets with highly correlated features. However, it can be computationally expensive for datasets with many features, and assumes that the data is normally distributed. Despite these limitations, ANOVA remains a popular and effective feature selection method for many real-world applications (A. Yang et al., 2019).

Chi-square:

The chi-square method for feature selection is a statistical method used for selecting relevant features in classification tasks. It is based on the chi-square statistic, which measures the independence between a feature and the class label. It is computationally efficient and non-parametric, and has been used in many applications, such as bio-informatics, text classification, and social media analytics. However, it has some limitations, such as assuming that the features are independent and favoring features with many distinct values or categories, which can lead to overfitting. Nevertheless, it remains a popular and effective feature selection method for many real-world applications (C. Wang et al., 2017).

ReliefF:

Relief-based feature selection is a popular method for selecting relevant features in classification and regression tasks. It is based on the idea of identifying feature subsets that are most useful for discriminating between instances of different classes, and calculates the weight of each feature based on its ability to distinguish instances that are close in feature space but have different class labels. It is robust to noisy and irrelevant features, and has been found to perform better than other feature selection methods in terms of classification accuracy and stability. However, it is a greedy algorithm that evaluates each feature independently, and may not be suitable for datasets with complex relationships between features and the target variable (Y. Zhang et al., 2019).

FCBF (Fast Correlation Based Filter):

Fast Correlation Based Filter is a feature selection method that is based on the correlation between features and the class label. It has been used in many applications, such as bioinformatics, image analysis, and social media analytics. It is efficient and non-parametric, and was found to be a useful feature selection method for predicting the prognosis of breast cancer patients. However, it is a greedy algorithm that may not always find the optimal subset of features, and may not work well with datasets with many noisy or irrelevant features (Yao et al., 2017).

Scoring methods for regression are:

Univariate Regression:

A feature selection method that is based on the correlation between individual features and the class label. It has been used in many applications, such as gene expression analysis, image analysis, and financial forecasting. It is easy to implement and can be used with a wide range of regression models. However, it does not account for the interactions between features, and may not work well with datasets that have many noisy or irrelevant features. Despite these limitations, it remains a popular and effective feature selection method for many real-world applications (C. Wang et al., 2017).

FCBF (Fast Correlation Based Filter): as described above.

Figure 3.5. Feature's importance for classification

		#	Info. gain	Gain ratio	Gini	ANOVA	χ^2	ReliefF	FCBF
1	C b Cluster	4	1.015	0.517	0.360	NA	44847.716	0.462	1.079
2	C a Cluster	4	0.803	0.418	0.272	NA	41107.783	0.409	0.000
3	C EPAGG	100	0.084	0.019	0.028	NA	307.094	0.054	0.000
4	C GENDER	2	0.055	0.055	0.017	NA	2300.749	0.061	0.039
5	C profess	9	0.045	0.017	0.015	NA	57.559	0.036	0.000
6	C DIEYTH	58	0.019	0.004	0.006	NA	897.723	-0.003	0.005
7	C GDP_2018	57	0.017	0.003	0.005	NA	532.869	-0.005	0.000
8	N GDP_2017		0.002	0.001	0.001	NA	94.676	-0.000	0.000
9	N income_2017		0.002	0.001	0.001	NA	133.483	0.000	0.000
10	N GDP_2016		0.002	0.001	0.001	NA	92.767	-0.000	0.000
11	N income_2018		0.002	0.001	0.001	NA	121.284	0.000	0.000
12	N income_2016		0.002	0.001	0.001	NA	121.284	0.000	0.000

Figure 3.6. Feature's importance for regression

			#	Info. gain	Gain ratio	Gini	ANOVA
1	C	GPA_B	96	1.718	0.275	0.050	NA
2	C	GPA_A	95	1.306	0.213	0.034	NA
3	C	EPAGG	100	0.183	0.041	0.003	NA
4	C	DIEYTH	58	0.086	0.016	0.001	NA
5	C	GDP_2018	57	0.084	0.016	0.001	NA
6	C	GENDER	2	0.061	0.061	0.001	NA
7	C	profess	9	0.060	0.023	0.001	NA
8	N	income_2017		0.007	0.004	0.000	NA
9	N	GDP_2017		0.007	0.004	0.000	NA
10	N	income_2018		0.007	0.004	0.000	NA
11	N	income_2016		0.007	0.004	0.000	NA
12	N	GDP_2016		0.007	0.003	0.000	NA

All of these techniques revealed that our initial assessment of the economic features importance was correct. The economic data had no explanatory power in relation to grades or achievement levels. As we can see from the Table 3.8, the correlation between grades or levels is too strong, whereas the correlation with economic characteristics is almost insignificant. It is important to note, however, that the strength of the correlation between income and academic achievement can vary depending on a number of factors, including the population being studied, the measures used to assess academic achievement, and the specific indicators of income that are examined.

It is important to note the weakness of using average income or average per capita income in such large geographical areas. It is not possible to identify individual groups living in specific conditions. Nor is it possible to distinguish between incomes in urban, semi-urban, or rural areas, and so on. The occupation of the guardian, on the other hand, is more important because it differentiates between students, as opposed to a value for the whole area of residence, which is present in income.

In addition, the use of average income as an indicator of the economic well-being of a country or region has limitations, and this is true for Greece as well. The basic reason why the average income may not provide a reliable indicator in Greece is the significant income inequality that exists within the country. According to data from the Hellenic Statistical Authority (ELSTAT), in 2019, the top 20% of the population had an average disposable income that was 6.3 times higher than the bottom 20%. This means that the average income figure may not accurately reflect the economic reality for many people in Greece.

Additionally, the use of average income as an indicator does not provide information on the distribution of income within the population. For example, a country could have a high average income, but if that income is concentrated in a small percentage of the population, it may not reflect the standard of living for the majority of people.

So, when Greece's economic health is studied, researchers look deeper than the average income; they study data such as the poverty rate, the difference between rich and poor, and the unemployment rate, among others. All these measures refer to a big geographical area with one number (nuts3), annihilating their importance to the forecast of students' scoring or achievement level.

Chapter 4

Students’ achievement levels

The importance of academic achievement of students is crucial for several reasons. Is a critical factor in determining a student’s future success, both in terms of their career and their overall quality of life. Students who perform well academically are more likely to have access to a wider range of educational and career opportunities, and are better equipped to handle the challenges of the modern workforce (Papadogiannis et al., 2021c). Academic achievement is also an important indicator of the effectiveness of educational institutions and policies. Schools, districts, and governments often use academic achievement data to assess the quality of their education systems and to identify areas where improvement is needed.

For all these reasons, in this thesis, students’ academic achievement was a key variable. Using lessons’ score data from students in the country, students were initially categorized into specific achievement levels. These levels constituted a new variable, which was examined both over time and in relation to specific demographic characteristics of students. This made it possible to examine the stability of school achievement levels as well as variations by non-academic characteristics.

4.1 Data collection and analysis

The workflow diagram describing the steps of the data analysis process in this thesis is shown in Figure 4.1. The first step was to obtain data from the Ministry of Education in CSV format files containing student scores by subject and grade and other academic as well as demographic characteristics. These CSV files contained data on all students in the country for grades 5, 6 of elementary school and grades 1, 2, and 3 of junior high school.

The second step was to create a MySQL database and import the data from the CSV files into tables by grade and year. Pivoting was also needed in order to create tables with columns of subjects and rows of students. Primary keys were defined to identify students by psevdo_id so that the tables could be linked later if needed.

In the third step, basic preprocessing and cleaning of the data were performed. This involved applying rules for missing or non-logical values, identifying and handling outliers, and converting data types if necessary. Also in the preprocessing phase, ILO categorization was applied to the guardian occupation in order to limit the number of occupations registered in the free field ”guardian occupation” in the EMIS.

The fourth step was to run a clustering algorithm on the pre-processed data to group students into distinct achievement groups. Instead of applying a common algorithm such as K-means, the minimally invasive and fast X-means algorithm was used. The use of the algorithm affected the performance levels of each student.

In the fifth step, a new ”cluster” variable was added to each table to indicate the achievement level. This way, each student row was assigned an achievement level by the clustering algorithm. This step allowed us to study how clusters differ in terms of other variables.

The sixth step was to perform chi-square independence tests between the new cluster variable and other relevant categorical variables (gender, occupational category, region of residence) to determine if there is statistical independence.

The final step involved answering the research questions. The answer to the first research question about achievement levels was based on the results of clustering. The following research questions were answered by the results of the independence tests and by using descriptive statistics.

4.1.1 The algorithm used

Starting from the clustering process, the most commonly used algorithm is K-means, but it has three key drawbacks: it has limited computational scaling, it employs a predetermined number of clusters, and it is vulnerable to local minima (Yuan & Yang, 2019). An alternate algorithm, which is an extension of K-means, is X-means, an algorithmically accelerated technique that is offered as a possible answer to these problems. The algorithm calculates the positions and number of clusters by optimizing the Bayes Information Criterion (BIC) or the Akaike Information Criterion (AIC). This clustering algorithm is very fast, and it is also statistically stable (Pelleg & Moore, 2000).

In K-means algorithm K is fixed and provided by the user. In X-means this assumption is changed, and it is sufficient for the user to specify a range in which the true K is expected to lie. The algorithm begins at the lower limit of the provided range and continues to add centroids until it reaches the top limit. During this process, the set of centroids with the best score is kept track of and then retrieved at the end. The algorithm’s output is a number in this range, which is best rated by a model selection criterion like the minimum BIC.

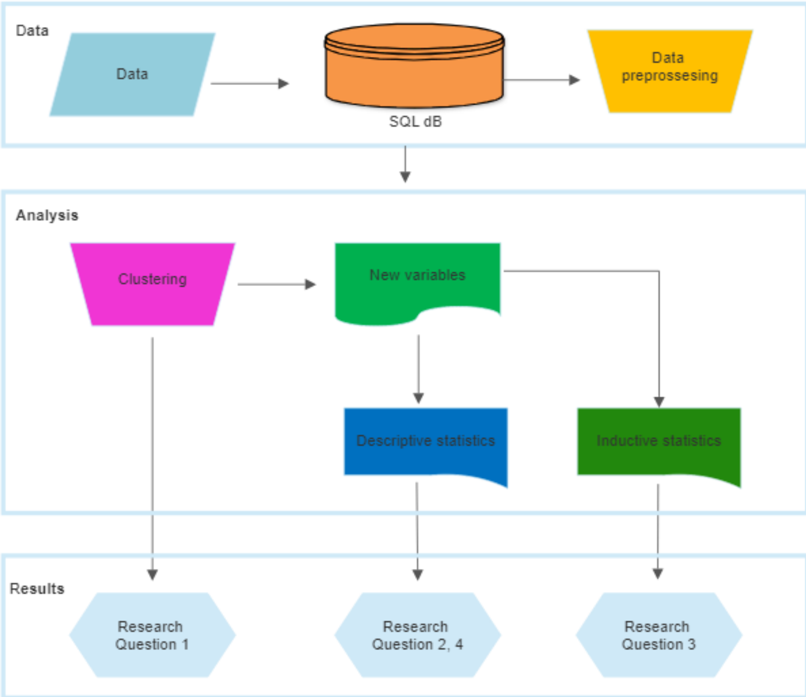


Figure 4.1. Data analysis and results

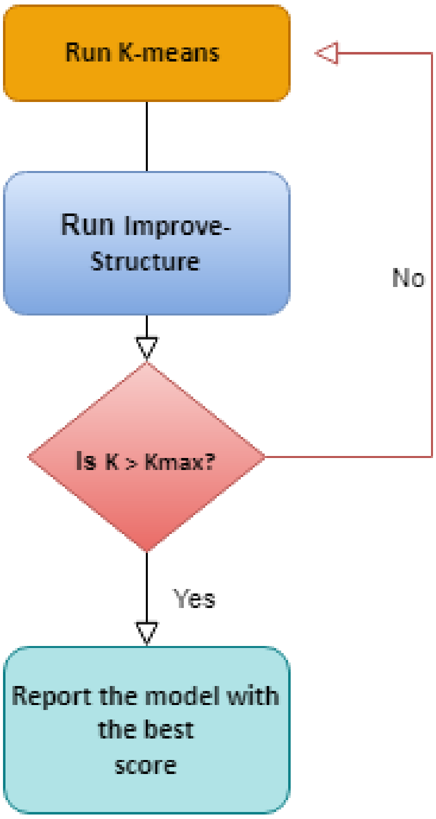
Another approach is also used to speed up the computational process (Improve-Structure). By using a KD-Tree and keeping statistical measurements on its nodes, the speed can be increased. A list is generated, of those centroids that need to be considered in each region (Pelleg & Moore, 2000). This list improves speed and scaling, allowing easy computation.

The X-means algorithm starts by dividing the data into two clusters using the K-Means algorithm. Then, it repeatedly splits the cluster with the largest variance until the variance of each cluster became falls a certain threshold. This results in a dynamic determination of the optimal number of clusters, as the algorithm will stop splitting when the variance can no longer be reduced by dividing the data into smaller clusters. This is a key advantage of X-means over K-Means, which requires the number of clusters to be specified in advance.

Another advantage of X-means is that it is more robust to noise and outliers. Since X-means splits clusters based on variance, it is less likely to be affected by outliers or noisy data. This makes it a good choice for datasets with a high degree of variability or noise. X-means is also a fast and efficient algorithm, as it uses a divide-and-conquer approach. By dividing the data into smaller clusters and using the K-Means algorithm on each of these clusters, X-means can converge faster than traditional K-Means. In addition, X-means uses the Bayesian Information Criterion (BIC) to determine the optimal number of clusters, which results in a more accurate and stable clustering solution.

The Improve-Structure function is a key component of the X-means algorithm and is used to dynamically determine the optimal number of clusters. The function splits a cluster

Figure 4.2. X-means algorithm



if its variance is greater than a certain threshold. This process is repeated until no further improvements can be made, resulting in a final clustering solution. The Improve-Structure function starts by dividing each cluster into two sub-clusters using the K-Means algorithm. The sub-clusters are then evaluated based on the Bayesian Information Criterion (BIC) to determine if they are a better fit for the data than the original cluster. If the BIC of the sub-clusters is lower than that of the original cluster, the split is accepted, and the process is repeated for each sub-cluster. This continues until no further improvements can be made, or the variance of the sub-clusters is below a certain threshold. The BIC is a measure of the quality of a clustering solution, taking into account both the goodness of fit and the complexity of the model. By using the BIC to evaluate the sub-clusters, the Improve-Structure function can ensure that the clustering solution is not only a good fit for the data, but also simple and easy to interpret. One important thing to note about the Improve-Structure function is that it does not guarantee that the global optimum solution is found. However, it has been shown to be effective in practice for a wide range of data sets and provides good results in a reasonable amount of time.

In conclusion, the Improve-Structure function is a critical component of the X-means algorithm, allowing it to dynamically determine the optimal number of clusters. By splitting clusters and evaluating the sub-clusters based on the BIC, the Improve-Structure function provides a flexible and efficient method for clustering data. The function has been widely used in various fields, including data analysis, image processing, and pattern recognition, and has been shown to provide good results for a wide range of data sets.

Finally the X-means algorithm is a flexible and efficient method for clustering data. It provides several advantages over traditional K-Means, including dynamic determination of the optimal number of clusters, robustness to noise and outliers, and faster convergence. The X-means algorithm has been widely used in various fields.

Cluster's selection criteria

The X-means algorithm runs until it reaches the maximum number of clusters given by the user and then extracts the number of clusters with the best score. Pelleg and Moore's criterion was the lower BIC.

$$BIC = -2logL(\theta) + log * n * dim(\theta) \tag{4.1}$$

Where L is the maximum likelihood, log the neutral logarithm, n the sample size and dim(θ) the length of the parameter vector. The model with the smallest BIC value is selected, since it best approximates the true model.

The Akaike Information Criterion was proposed by (Akaike, 1974) and is one of the most important and popular methods of model selection. Its applications are found in linear model problems regression and time series, in Survival Analysis and generally in all in all cases of comparison of parametric models. Its formula is as follows:

$$AIC = -2logL(\theta) + 2 * dim(\theta) \tag{4.2}$$

Where L(θ) is the maximum likelihood, log the neutral logarithm and dim (θ) the length of the parameter vector. In practice, after first calculating the value of AIC for each candidate model, we choose the one with the lowest ”score”, since it is the one closest to the actual model. Observing the form of the BIC, it is clear that it has many similarities with that of the AIC, as here too the maximum likelihood is used and a term penalty term (log n dim(θ)), so that by combining them from the candidate models, those that are both accurate and simple are selected.

4.2 Steady achievement levels

4.2.1 Ranking the achievement levels

Our primary aim was to examine whether the academic achievement of students can be grouped into specific levels of academic achievement or not. Intuitively, we are aware that teachers can detect which students in their courses are strong, which students are average, and which students are either not participating in class activities or are ”very weak”; however, we have decided to follow the statistics rather than the intuition of the teachers.

We used X-means (Pelleg & Moore, 2000), an extension to K-means that also estimates the value of K, in order to avoid any bias that comes from looking for a specific number of clusters. This allowed us to avoid looking for a specific number of clusters. After applying the clustering technique, it was found that there are four (4) different levels of academic achievement among the students. These specific levels cover both ES and HS and are similar at each grade level. The centroids and the standard deviations of each class's grades, covering all five school years, are shown in Table 4.1. The dataset contains information on each and every student in Greece, and Table 4.2 presents the BIC values which resulted, broken down by class and school year.

Table 4.1. Cetroinds (CD) per achievement level

Cluster	**CD 5th ES**	**CD 6th ES**	**CD 1st HS**	**CD 2nd HS**	**CD 3rd HS**
A	9.988/0.002	9.979/0.023	19.070/0.038	19.100/0.088	19.076/0.014
B	9.727/0.029	9.680/0.264	17.144/0.071	16.934/0.170	16.826/0.062
C	9.008/0.018	8.912/0.081	15.205/0.118	14.811/0.120	14.696/0.183
D	8.180/0.066	8.263/0.574	13.073/0.256	12.845/0.346	12.962/0.471

The first point to note is that rather than finding the three unique categories of academic achievement, which is intuitively divided (strong, medium, week), we came across four separate groups. The data-based approach which seeks to eliminate bias and listen to what the numbers have to say is given a boost as a result of this finding. The fact that the X-means algorithm was run fifteen times in two datasets (one per class and school year) and each time yielded the same number of clusters is, however, a far more significant observation. The separate levels of academic achievement throughout primary and secondary school, or at the very least from the fifth grade of elementary school to the end of high school, were four. As a result, we can draw the conclusion that this is not a random outcome. Throughout the rest of this thesis, we will refer to these levels as “Very strong”, “Strong”, “Weak”, and “Very weak”, respectively. The standard deviation within the clusters was fairly low in virtually all of the cases, which leads us to believe that the

Table 4.2. BIC values per grade and year

Class	School Year	No of Clusters	BIC–Value
5^{th} Elementary School	16–17	4	770,325.77
	17–18	4	924,904.65
	18–19	4	938,884.65
6^{th} Elementary School	16–17	4	886,493.46
	17–18	4	917,255.56
	18–19	4	742,378.62
1^{st} High School	16–17	4	467,536.44
	17–18	4	485,932.55
	18–19	4	494,685.45
2^{nd} High School	16–17	4	401,889.33
	17–18	4	404,857.79
	18–19	4	566,802.06
3^{rd} High School	16–17	4	475,555.07
	17–18	4	378,631.03
	18–19	4	416,664.98

clusters are quite distinct from one another. One group, the ”Very Weak,” students’ group, is an exception to this rule as the standard deviation was bigger. This is to be expected given that this group contains all possible grades, all the way down to virtually zero.

Figure 4.3 depicts a clear example of the difference between levels of academic achievement in the 1^{st} HS. Beyond the statistical significance of the difference between all levels, there are large absolute differences in the difference between the centroids of each level. In high school, the difference reaches five points on the 20-point scale. Also evident is the overlap between the mean achievement levels (”B” and ”C”) and the partial overlap in the ”C” and ”D” groups.

It was also found that the average grades per subject are clearly separated between student achievement levels. There was no case of a subject in which the grades did not follow the level ranking. For example, Figure 4.4 shows average grades by subject and level. There is a clear differentiation between levels in the Greek language, Mathematics, Science etc. Only in the subjects of Gymnastics and Music are the average scores more closely placed, as would be expected since in these subjects the methods of teaching, knowledge, skills and assessment differ. The same picture was found in all classes in both datasets.

Additionally, we are able to observe that the gap between the four levels of academic achievement in elementary school is quite small, whereas in high school the gap is significantly larger (as GPA percent, Figure 4.6). This is primarily due to the fact that, whereas in junior high school nearly all of the grades from 1 to 20 can be used, in elementary school

Figure 4.3. Box plot per achievement level GPA (1st HS)

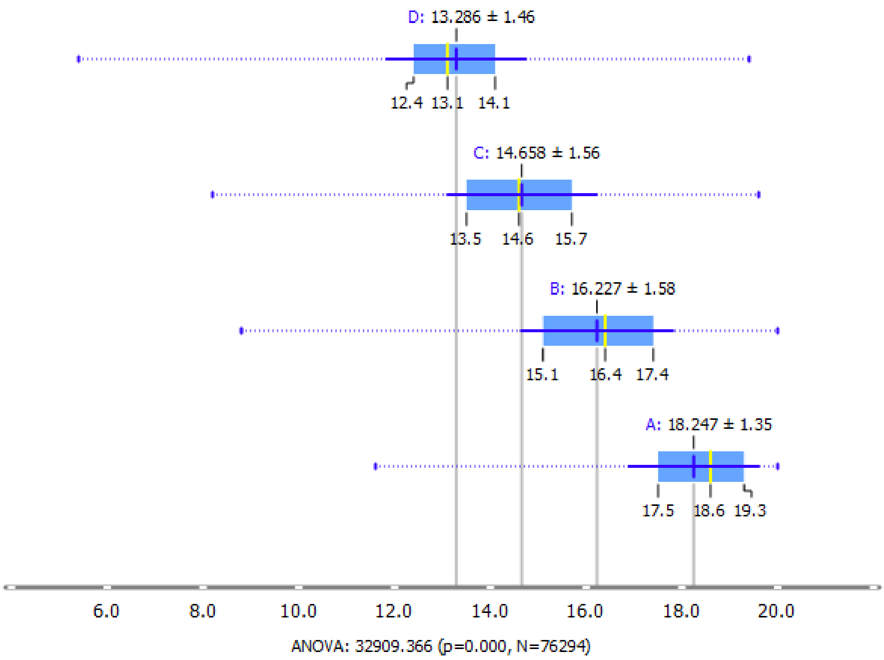
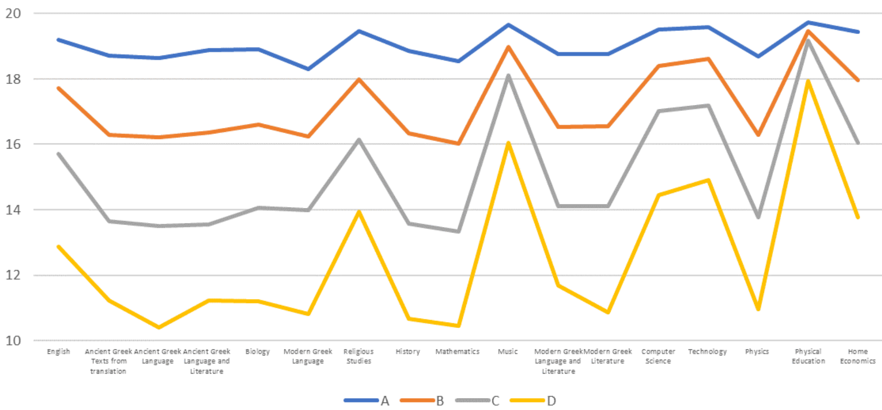


Figure 4.4. Average grades by subject and level

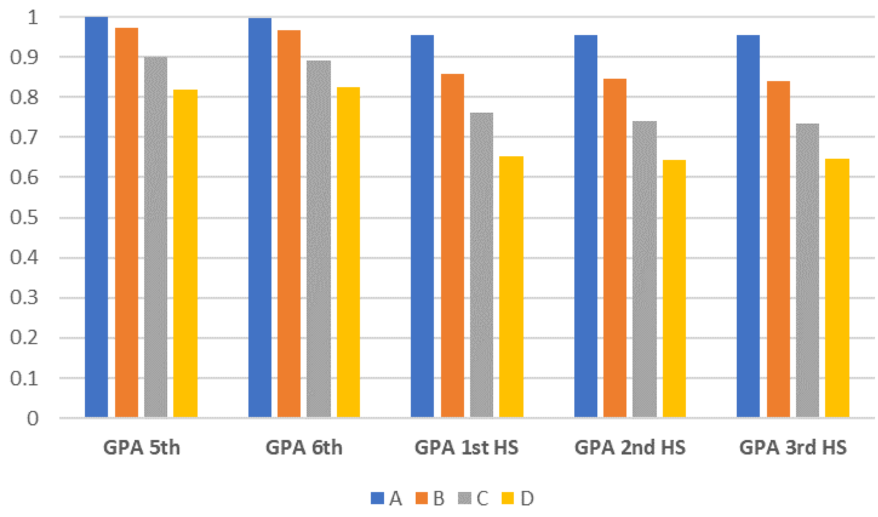


the majority of grades are restricted to the 7–10 range, and lower grades are rarely, if ever, used. So, what we're seeing here isn't necessarily a sign of different grading, but rather of different academic achievement.

Achievement differences between grades

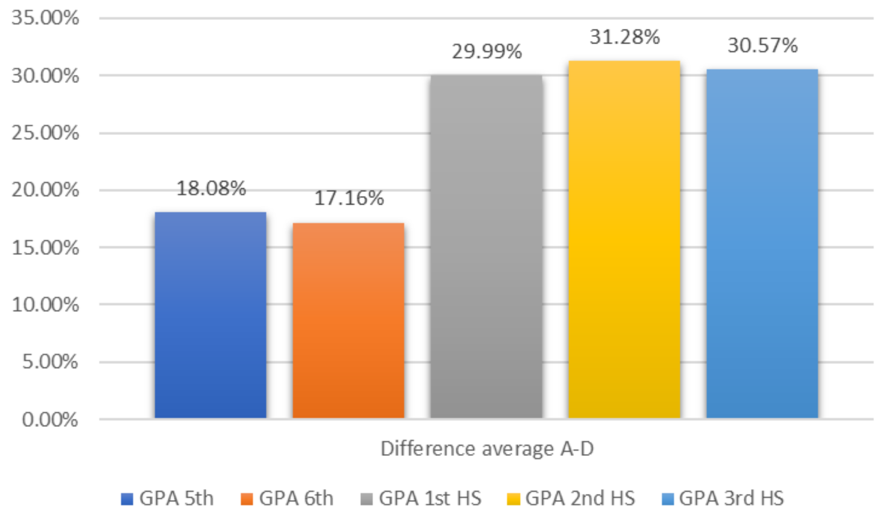
By looking at the data as percentage differences from maximum grade scales, the picture becomes more clear. The difference between maximum and minimum performance, as percentage of maximum grade, equals 18.08% in the 5th grade and 17.16% in the 6th

Figure 4.5. GPA per level, as % of the maximum of grading scale



grade. In contrast, in high school, it shoots up to 29.98% (1st grade), 31.25% (2nd grade), and 30.57% (3rd grade). The difference in scores, as a percentage, between very strong and very weak students is generally much larger in junior high school. This shows the difference in the levels of difficulty of the lessons and the difficulty of transition of students in secondary school.

Figure 4.6. Difference between GPA of ”A” and ”D” level, as % of the maximum of grading scale



Differences between lessons - first dataset

Examining the differences between the average of each performance level by subject leads to signs of whether subjects act as causes of variation in achievement levels. The tables below detail the differences between the means of the achievement levels. In the fifth grade, the main subjects in which there was a difference in performance between the very strong and very weak students are, in descending order: History (2.9153), Geography (2.8086), Maths (2.7337) and Science (2.7136). The main subject of Greek language has shown slightly smaller difference. The differences between 'strong' and 'very weak' students are similar. It seems that these subjects show a greater ability to separate performance, while in contrast Music, ICT, Art and Physical Education do not show significant differences between levels.

Figure 4.7. Differences between levels per lesson (5th ES, 1st dataset.)

Academic achievment level 5th Class	English	Geography	Greek Language	Arts	Religious Education	History	Social and Civic Education	Mathematics	Music	ICT	Science	Physical Education
Average Diff A-B	0.6408	0.5889	0.8107	0.0220	0.0809	0.6697	0.1818	0.7940	0.0445	0.0627	0.5427	0.0197
Average Diff A-C	1.4311	1.5723	1.6307	0.0572	0.4061	1.7083	0.7190	1.6596	0.1325	0.1761	1.4738	0.0572
Average Diff A-D	2.3185	2.8086	2.6539	0.2376	1.6661	2.9153	2.0689	2.7337	0.3762	0.5373	2.7136	0.2156
Average Diff B-C	0.7904	0.9834	0.8200	0.0353	0.3252	1.0386	0.5372	0.8655	0.0880	0.1133	0.9310	0.0374
Average Diff B-D	1.6777	2.2197	1.8432	0.2157	1.5851	2.2457	1.8871	1.9396	0.3316	0.4746	2.1708	0.1959
Average Diff C-D	0.8873	1.2363	1.0232	0.1804	1.2599	1.2071	1.3499	1.0741	0.2437	0.3613	1.2398	0.1584

For sixth grade, the largest differences are found in the subjects of History, Geography, Greek Language, Mathematics and Physics. The average difference between the category of "very strong" and "very weak" students exceeds 2.5 points on the 10-point scale. The difference between very strong and "strong" students remains less than one. Very significant are also the differences between the average levels of the 'strong' and 'very weak' students, which range from 1.6675 in English Language, to 2.2808 in Geography. The differences between the 'weak' and 'very weak' categories are also considerable. There does not appear to be any discriminatory ability for the subjects of Art, Music, ICT and Physical Education. All these subjects are complementary to teaching and in most cases, are taught by teachers other than the main class teacher.

Figure 4.8. Differences between levels per lesson (6th ES, 1st dataset.)

Academic achievment level 6th Class	English	Geography	Greek Language	Arts	Religious Education	History	Social and Civic Education	Mathematics	Music	ICT	Science	Physical Education
Average Diff A-B	0.6667	0.5041	0.7779	0.0384	0.0814	0.7276	0.1673	0.7459	0.0628	0.0697	0.3975	0.0202
Average Diff A-C	1.4525	1.4659	1.5807	0.0909	0.3921	1.7497	0.6569	1.6336	0.1612	0.1901	1.3156	0.0594
Average Diff A-D	2.3441	2.7849	2.6528	0.3127	1.6089	2.9225	1.9904	2.7539	0.4441	0.5698	2.6454	0.2547
Average Diff B-C	0.7859	0.9618	0.8027	0.0525	0.3107	1.0221	0.4896	0.8877	0.0984	0.1204	0.9181	0.0392
Average Diff B-D	1.6775	2.2808	1.8749	0.2743	1.5276	2.1949	1.8231	2.0079	0.3813	0.5001	2.2479	0.2346
Average Diff C-D	0.8916	1.3190	1.0722	0.2218	1.2169	1.1728	1.3335	1.1203	0.2829	0.3797	1.3298	0.1954

The differences between the level centers of 'very strong' and 'very weak' students for the first year of secondary school are bigger for the subjects of History, Biology, Ancient Greek, Greek, English and Home Economics. The same is true for the differences between "strong" and "very weak" students. The average differences in these cases are greater than

Figure 4.9. Differences between levels per lesson (1st HS, 1st dataset.)

Academic achievment level 1st Class HS	English	Ancient Greek Language	Biology	Modern Greek Language	Religious Studies	History	Mathematics	Music	Modern Greek Language and Literature	Computer Science	Technology	Physics	Physical Education	Home Economics
Average Diff A-B	0.1288	0.2097	0.1937	0.0563	0.2009	0.2440	0.1937	0.0563	0.2002	0.0712	0.0793	0.0207	0.0320	0.1716
Average Diff A-C	2.8998	3.8504	4.5814	1.3936	3.2365	4.6985	4.5814	1.3936	3.7682	1.9908	1.6635	0.5814	0.4575	3.2406
Average Diff A-D	6.1704	6.8124	6.8949	3.5473	5.9240	7.1937	6.8949	3.5473	6.5401	4.4458	4.2891	2.9047	1.2955	5.9524
Average Diff B-C	2.7710	3.6408	4.3876	1.3373	3.0356	4.4545	4.3876	1.3373	3.5680	1.9196	1.5842	0.6020	0.4896	3.0691
Average Diff B-D	6.0416	6.6027	6.7011	3.4911	5.7231	6.9497	6.7011	3.4911	6.3399	4.3746	4.2099	2.9254	1.3275	5.7808
Average Diff C-D	3.2706	2.9620	2.3135	2.1538	2.6875	2.4952	2.3135	2.1538	2.7719	2.4550	2.6256	2.3233	0.8380	2.7117

6 points on a scale of 20. It should be noted, however, that the difference between "weak" and "very weak" students is also very significant (more than 2 points), which leads to the need for repetition of the class almost exclusively among students who were initially characterized as "very weak".

Differences between levels per lesson - second dataset

In the second dataset, we discovered that, similar to the first, the differences in averages by level of academic achievement are clearly bigger in several subjects. For each grade, the differences between levels were calculated, and it was found that in the first grade of junior high school, the most significant differences between the levels of "very strong" and "very weak" students are found in the subjects of History, Mathematics, Ancient Greek, Modern Greek, Biology, Geography, Foreign Languages and Religious Studies.

Figure 4.10. Differences between levels per lesson (1st HS, 2nd dataset.)

Differences	English	Ancient Greeks	Biology	Geography	Modern Greek Language	Religious Studies	History	Arts	Mathematics	Music	Modern Greek Language	Home Economics	Computer Science	Culture	Technology	Physics	Physical Education	Second foreign Lang
A-B	1.3931	2.0254	2.1728	2.1264	2.1489	1.5064	2.6920	0.5199	2.8294	0.6714	2.1698	1.5365	1.0544	0.5177	0.7100	1.4417	0.7571	1.7468
A-C	3.2157	4.3824	4.6871	4.6057	4.2814	3.4862	5.5444	1.1603	5.5498	1.5828	4.4210	3.5860	2.3070	1.2671	1.7966	2.9093	1.6747	4.0851
A-D	6.2249	7.1953	7.2344	7.1695	6.7506	5.9761	8.0368	2.6718	7.9011	3.4509	7.1126	6.1343	4.4194	2.9409	3.9579	4.9343	2.9159	6.6832
B-C	1.8226	2.3570	2.5143	2.4793	2.1325	1.9798	2.8524	0.6404	2.7205	0.9114	2.2512	2.0495	1.2526	0.7494	1.0866	1.4677	0.9176	2.3383
B-D	4.8319	5.1699	5.0617	5.0431	4.6017	4.4697	5.3448	2.1519	5.0717	2.7795	4.9428	4.5978	3.3651	2.4232	3.2480	3.4926	2.1588	4.9364
C-D	3.0093	2.8129	2.5473	2.5638	2.4692	2.4899	2.4924	1.5115	2.3513	1.8681	2.6916	2.5483	2.1125	1.6737	2.1614	2.0250	1.2412	2.5981

All these differences are greater than 7 points on a scale of 20. The differences between "strong" and "very weak" students in the same subjects are similar and are around 5 points on the 20-point scale. The average differences between the categories of 'weak' and 'very weak' students are also remarkable. There are small differences in the subjects of Arts, Music, technology, and Physical Education.

Figure 4.11. Differences between levels per lesson (2nd HS, 2nd dataset.)

Differences	English	Ancient Greeks	Biology	Geography	Modern Greek Language	Religious Studies	History	Arts	Mathematics	Music	Modern Greek Language	ICT	Culture	Technology	Physics	Physical Education'	Chemestry	Second foreign Lang
A-B	1.5046	2.1443	2.1971	2.0420	2.1770	1.3713	2.7935	0.5504	2.9053	0.6963	2.1941	1.1091	0.5384	0.7133	2.1542	0.9684	1.7101	1.8969
A-C	3.3482	4.5249	4.8084	4.5212	4.2773	3.2921	5.7102	1.2552	5.7223	1.6056	4.4238	2.3744	1.3139	1.7444	4.2564	2.0236	3.5425	4.2385
A-D	6.2612	7.3541	7.4180	7.1698	6.7332	5.7911	8.1186	2.9310	8.0055	3.4370	7.1272	4.4404	3.0529	3.7361	6.4586	3.2238	5.3860	6.6888
B-C	1.8435	2.3805	2.6113	2.4791	2.1003	1.9208	2.9168	0.7048	2.8169	0.9093	2.2297	1.2653	0.7756	1.0311	2.1022	1.0551	1.8325	2.3416
B-D	4.7566	5.2097	5.2210	5.1278	4.5562	4.4198	5.3252	2.3806	5.1001	2.7407	4.9331	3.3313	2.5145	3.0228	4.3044	2.2553	3.6760	4.7920
C-D	2.9131	2.8292	2.6097	2.6486	2.4560	2.4990	2.4084	1.6758	2.2832	1.8313	2.7034	2.0660	1.7389	1.9917	2.2022	1.2002	1.8435	2.4504

The picture remains the same in the second grade of junior high school, where the teaching of Chemistry begins. The differences concern the same subjects, with History and Mathematics again showing differences of more than 8 points between "very strong" and "very weak" students. Chemistry also proves to be a good indicator of clusters' split. The differences between the other levels of academic achievement are similar.

Figure 4.12. Differences between levels per lesson (3^{rd} HS, 2^{nd} dataset.)

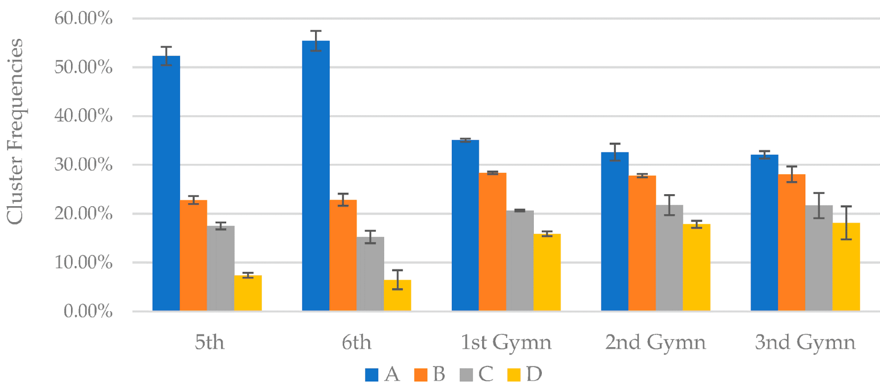
Differences	English	Ancient Greeks	Biology	Linguistic Teaching	Religious	History	Arts	Mathematics	Music	Modern Greek Language	Social and Political Science	ICT	Technology	Culture	Physics	Chemistry	Physical Education
A-B	1.4684	2.1823	2.5296	2.3306	1.3715	3.0065	0.5261	3.2474	0.6852	2.3568	1.7994	1.0241	0.5581	0.7481	1.7663	2.4954	0.2101
A-C	3.4880	4.8973	5.4593	4.7296	3.4180	6.1727	1.1333	6.2749	1.4705	4.9810	4.2039	2.2515	1.2167	1.8568	3.5451	5.4895	0.4528
A-D	6.4845	7.7796	7.7363	7.1442	5.9820	8.3714	4.6956	8.3469	3.9095	7.7010	6.7259	4.5244	3.6704	4.0386	6.0911	7.8246	1.0930
B-C	2.0196	2.7150	2.9296	2.3991	2.0464	3.1661	0.6072	3.0276	0.7853	2.6242	2.4045	1.2274	0.6585	1.1087	1.7788	2.9941	0.2427
B-D	5.0161	5.5973	5.2067	4.8136	4.6105	5.3649	4.1695	5.0996	3.2243	5.3442	4.9264	3.5003	3.1122	3.2906	4.3248	5.3292	0.8829
C-D	2.9965	2.8823	2.2770	2.4145	2.5640	2.1988	3.5623	2.0720	2.4390	2.7201	2.5219	2.2729	2.4537	2.1818	2.5461	2.3351	0.6402

In summary, it can be seen that the differences between the average grades by level are most significant in the core subjects of secondary school, which are also tested in the final written examinations. The written exams may be a factor in these differences because it is an objective test of the most important things taught. On the other hand, subjects that belong to the group that are not examined at the end of the school year do not show much discriminatory power. These are subjects in Arts, Music and Physical Education.

4.2.2 Achievement levels' magnitudes

In the previous sections, we have concentrated on the centroids and the differences between clusters; however, we have not looked at the overall size of the clusters. The sizes of each cluster per class are shown in Figure 4.13.

Figure 4.13. Frequencies per cluster. (A is Very strong, B is Strong, C is Weak, and D is Very weak)



The difference in distributions between elementary school and high school is very clear. In elementary school, the majority of students are considered to belong to the group of students with very strong academic achievement. However, by the time students reach high school, only about one third of students are considered to have very strong academic achievements, and weaker academic achievements become relatively more common. On the other hand, when students reach high school, the differences in frequencies at each level become more pronounced.

This is an evidence that there is a higher dispersion in the distribution of grades that students obtained when they were enrolled in junior high school. The average performance of ”very strong” students in junior high school is showing a little decline over time, and the frequency of ”very weak” students is increasing; table 4.3 demonstrates that. In addition to the consistency of the level of performance in each category, there is also a consistency in the frequency of occurrence of the various levels.

Table 4.3. Frequencies of achievement cluster by grade.

Cluster	5th ES	6th ES	1st HS	2nd HS	3rd HS	Mean/St. Dev ES	Mean/St. Dev HS
A	52.30%/1.90%	55.40%/2.03%	35.08%/0.32%	32.57%/1.75%	32.07%/0.78%	53.8%/1.97%	33.24%/0.95%
B	22.80%/0.80%	22.87%/1.22%	28.38%/0.27%	27.81%/0.37%	28.07%/1.59%	22.84%/1.01%	28.09%/0.74%
C	17.50%/0.70%	15.25%/1.29%	20.64%/0.17%	21.78%/2.04%	21.69%/2.59%	16.37%/0.99%	21.37%/1.60%
D	7.40%/0.50%	6.48%/1.98%	15.91%/0.50%	17.84%/0.74%	18.17%/3.38%	6.94%/1.24%	17.31%/1.54%

Since the frequency distributions for each level of achievement in junior high school are different from those in elementary school, it is reasonable to wonder why this sudden change happened. While the percentage of students that fall into the poorer performance category has more than doubled, the number of students receiving an exceptional mark (very strong) has decreased to one-third of the total.

This highlights the difference in the level of difficulty of the lessons in secondary school, as well as the difference in the students’ transition to the new school. The change from elementary school to junior high school is not an easy step in a child’s development because there are many changes in the social and educational environment. Students not only have to adjust to larger facilities, new teachers, new demands, new curricula, and a larger number of classmates. They also have to adapt to their new way of thinking and acting in secondary school. Poor school transitions have been found to have long-term adverse effects on mental health (Waters et al., 2012). According to Zeedyk et al. (2003), the transition to secondary school is one of the most difficult moments in a student’s educational career.

Students often have conflicting feelings about the transition (Sirsch, 2003). Children expect to have more independence, to take on new tasks and to meet new friends. At the same time, students worry that they will be bullied by older students, that they will have to study harder, that they will receive lower grades, and that they will be confused in a larger, new school.

In relation to assessment, the non-competitive nature of assessment in primary school should also be emphasized, which refers not only to performance but also to other characteristics such as effort, initiative, creativity, and cooperation with peers and other individuals, etc. (van Rens et al., 2018).

The importance of a smooth transition can also be seen in the well-known consequences of a poor transition, according to the literature. Students also feel excluded, unwanted, and undervalued by others. This can lead to withdrawal from school (L. W. Anderson et al., 2000), poor academic achievement and dropping out of school (Waters et al., 2012), and conflict between the child and the school (L. W. Anderson et al., 2000).

In this study, we observed that as we moved from one grade to another, the percentage of students in the class whose academic achievement was extremely low continued to increase. This demonstrates, however, that as the years pass, a larger number of students are falling further and further behind.

4.3 Findings

The aim of the first research question was to examine whether students' academic achievement can be grouped into specific levels. It was found that there are four distinct levels of academic achievement. The specific levels were identified across different grades and years, which confirms the validity and reliability of the results. The centroids for each cluster were clearly separated, with the small standard deviation within clusters, indicating that the clusters were distinct from each other with few outliers. The mean grades in each course followed the ranking of the levels.

It also found an increase in the gap between primary and secondary school achievement levels. Although the number of levels remained the same, the percentages of students in the achievement levels changed. Fewer students now fall into the "very strong" category and more into the "very weak" category. This highlights the increased difficulty of the subjects and challenges in adapting to the new school environment.

The thesis aimed to objectively evaluate dimensions of the educational system and specific educational policies in light of providing equal opportunities for students. With student achievement as the main tool of study, it was necessary to objectively determine the levels of student achievement in order to then examine whether the school can function as an equal opportunity mechanism both overall and at the level of individual educational policies. Examining the possibility of objectively determining levels of achievement was the subject of the first research question.

Using unsupervised learning, four levels of student achievement for primary and secondary schools emerged. The classifications "very strong", "strong", weak," and "very weak" result from the performance hierarchy and are stable over time. These ratings are based on the twenty-point scale for junior high school and the ten-point scale for elementary school. The designation of a student as "strong" betrays a direct assessment that approximates a particular "picture" and is linked to some corresponding latent features.

This grouping could lead to the development of new educational interventions more adapted to the level of student achievement. The aim of these policies should be to improve the upward mobility of students' achievement. The results of this research showed significantly higher mobility between levels "b" and "c". These are those who are classified as "average" students. Improving this two-way mobility between levels should be a goal of educational policy. A clear picture also emerges with regard to students who fall into the categories of ”very strong” and ”very weak”. More than two-thirds of those who are initially classified in one of the two categories do not change. Although this is very positive for the ”very strong” category, it raises major questions about the effectiveness of the education system for ”very weak” students. Although the particular attributes of each student’s ”profile” in these groups are unknown, ”weak” students are expected to improve faster than ”very weak” students.

To improve the achievement level of students classified as ”very weak”, the education system must implement targeted intervention programs. Additional support should be provided to help these students gain fundamental skills and close achievement gaps. Smaller class sizes and one-on-one tutoring could ensure ”very weak” students receive individualized attention from teachers (Hawley et al., 1984). Specialized remedial teaching courses designed for their level of understanding are needed (Grissmer et al., 2000). Teachers require extensive training to effectively support diverse learners with differentiated instruction methods. Increased monitoring and evaluation of student progress is also important to assess the effectiveness of interventions early on. Finally, engagement with parents/guardians is critical to developing support at home. With the right mix of academic, social-emotional, and family support, more ”very weak” students can significantly improve their achievement levels over time (Brigman & Campbell, 2003).

Chapter 5

Longitudinal analysis of students' achievement levels

The second research question examined the longitudinal dimension of students' achievement levels. Through examining achievement levels from year to year, it is possible to draw conclusions about the stability of factors affecting student achievement. It is also possible to examine the effectiveness of the school system in terms of its ability to improve student achievement. Furthermore, the longitudinal examination can show weaknesses in the education system in relation to student performance and provide clear evidence of the effectiveness of specific educational interventions in improving students' achievement.

Because we had two datasets available, the first for the 5^{th}, 6^{th} grade in elementary school and the 1^{st} grade oh junior high school and the second for the three junior high school grades, a three-year longitudinal analysis was performed on both datasets. A summary of the data used can be found in the Tables 5.1 and 5.4. The consistency of the results on both datasets was then studied.

5.1 From ES to HS, longitudinal analysis

The pattern of student achievement from elementary school to junior high school was examined using data from 5^{th} and 6^{th} grade in elementary school and the 1^{st} grade in junior high school. It also studied how students' Grand Point Average scores change over the time of their high school career in three different grades. At first, Table 5.1 illustrates the consistency of students' academic achievement throughout their time in elementary school. A significant stability in achievement levels is found. 90.1% of students who were rated as ”very strong” in fifth grade are still rated as excellent in sixth grade. In addition, 62.3% of them were rated as “very strong” in fifth grade of elementary school continued to be rated as “very strong” when they were in the first grade of secondary school.

Table 5.1. Frequencies of achievement level over time, based on initial clustering. (1st dataset)

		Initial achievement level (5th ES)			
Cluster	Class	A	B	C	D
A	6th Class ES	**90.10%**	**9.20%**	0.70%	0.00%
	1st Class HS	**62.30%**	30.00%	7.10%	**0.60%**
B	6th Class ES	34.80%	50.50%	13.70%	1.00%
	1st Class HS	14.20%	**42.60%**	35.50%	7.70%
C	6th Class ES	4.50%	29.50%	53.90%	12.10%
	1st Class HS	1.50%	**17.10%**	47.90%	33.50%
D	6th Class ES	0.50%	5.60%	36.50%	57.40%
	1st Class HS	0.20%	4.00%	**28.20%**	**67.60%**

There is a greater degree of variation in achievement levels between classes for students who fall into the B and C achievement levels. Only 42.6% of students who were rated at level B in the fifth grade of elementary school are able to maintain this level of performance in junior high school. Of those who were rated in category B in the fifth grade of elementary school, only 9.2% were characterized as excellent in the sixth grade of elementary school.

On the other hand, it would appear that the typical high school student is functioning at a lesser level. The fact that 28.20% of students who were placed in category C in the fifth-grade fall into the lowest D category in the first class of junior high school and 17.10% of students who were placed in category B fall into category C in first grade of junior high school is an sign of the increasing difficulty that students face when they attend high school. Those students who were placed in the lowest performance category in the fifth grade of elementary school also remain in that category (67.60%) in the first grade of high school while very few students (0.6%) are able to excel in their studies. A visual representation of these differences between achievement levels are displayed in the following stacked column charts. The blue color represents the best achievement (very strong students), while the yellow represents the lowest (very weak students). We can easily observe the stability from one grade to the next, especially for very strong and very weak students.

Based on the initial clustering of student performance in the 5th grade, Figure 5.2 shows a distinctly different distribution of GPA in junior high school. Student performance in 6th grade in elementary school and in the 1st grade of high school appear to be linked with the initial clustering of student achievement.

Table 5.2 provides a more in-depth icon of the fundamental descriptive statistics of GPA. These statistics are derived from the initial categorization of students into groups that

Figure 5.1. Achievement level frequencies from 5th to 6th ES

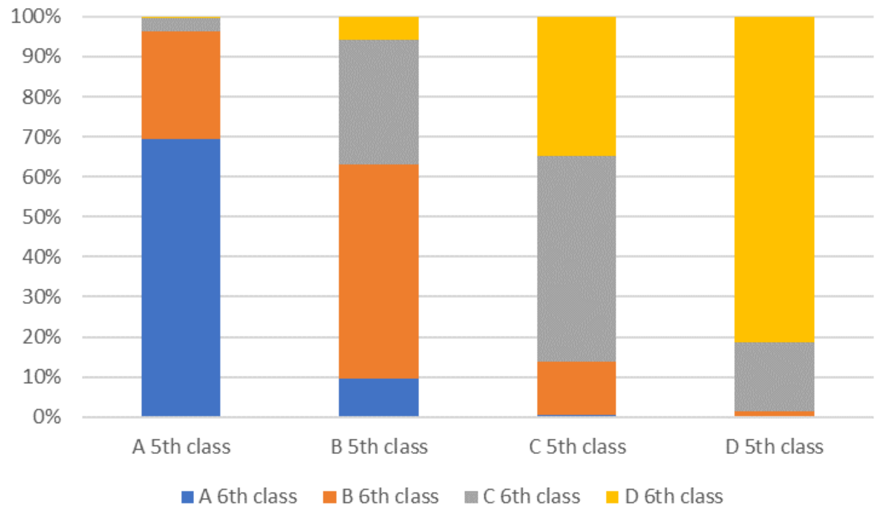
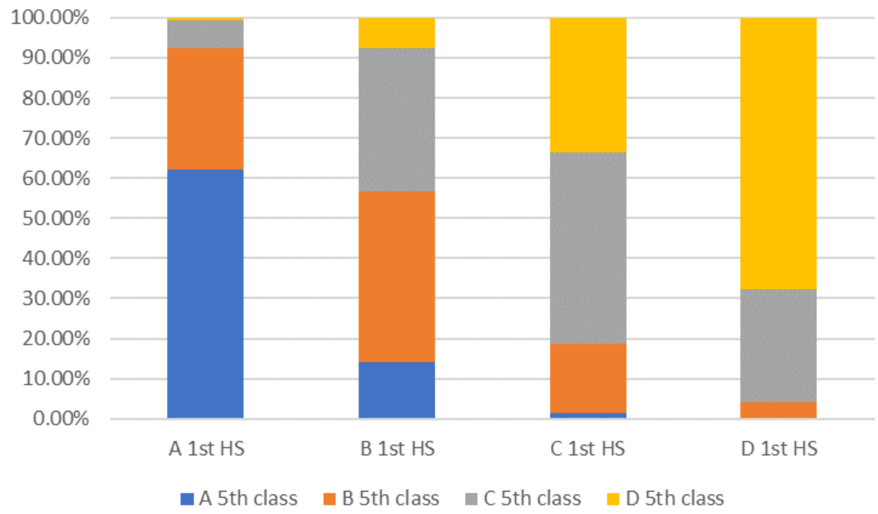


Figure 5.2. Achievement level frequencies from 5th ES to 1st HS



took place in the 5th grade of elementary school. There is also a more normal distribution of student achievement levels in the 1st grade of high school compared to 6th grade of elementary school.

The Kruskal-Wallis Test was also used to test the statistical significance of GPA variance depending on the initial classification of the students. There were found statistically significant differences in GPA between the various initial categories of the students in the next two grades (Table 5.3).

The achievement path that students follow based on their initial grouping in 5_{th} grade is made more clear through the following Sieve diagrams (Figures 5.4, 5.5). The blue areas in these charts show that the Grade Point Averages have a higher frequency than expected

Figure 5.3. GPA boxplots based on the initial clustering (1st dataset).

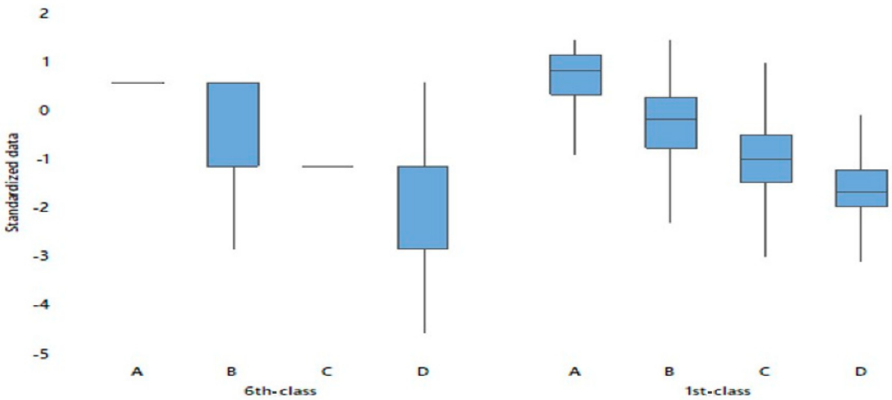


Table 5.2. Descriptive statistics of GPA based on the initial clustering (1st dataset).

Variable	5th Class Level	Mean	SE Mean	St. Dev	Median
6th Class ES GPA	A	9.9784	0.00073	0.146	10
	B	9.7059	0.00344	0.467	10
	C	9.1154	0.00441	0.515	9
	D	8.4096	0.00992	0.671	8
1st Class HS GPA	A	18.247	0.0068	1.348	18.6
	B	16.227	0.0117	1.581	16.4
	C	14.658	0.0133	1.558	14.6
	D	13.286	0.0215	1.455	13.1

in a χ^2 contingency table. In contrast, the red areas indicate Grade Point Averages that have a lower frequency than expected. In each case, the statistical test of independence shows strong differentiation. For the 6th grade, the ”red area” contains grades below 9.5 for ”very strong” students, and, as expected, the maximum of the scale is out of reach for ”very weak” students. In contrast, the blue areas are dominated by combinations of very strong or very weak students for the top or bottom of the score scale.

Even though it’s been two years since students were first put into groups based on their achievement levels (Figure 5.5), the picture is still clear and strong. This connection is illustrated by the sieve diagram’s diagonal split. The number of students who are ”very weak” is four times higher than what would be expected. Almost all of the ”very weak” students in the overall population had a score of less than 15.15.

Table 5.3. Kruskal-Wallis test of GPA based on the initial clustering (1st dataset).

Class	Statistic	*p*-value
5th Elementary School	57,365.28	0.0000
6th Elementary School	44,992.44	0.0000
1st High School	41,402.00	0.0000

Figure 5.4. Sieve diagram - GPA 6th grade per initial achievement level

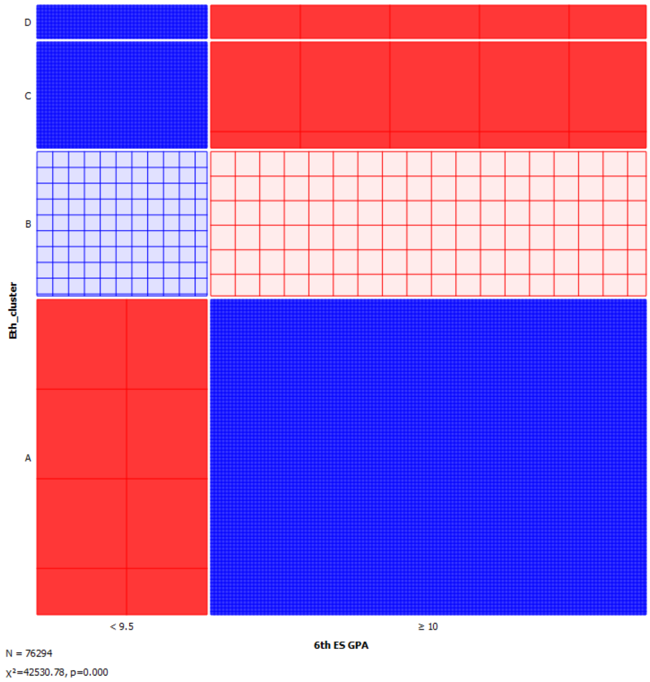
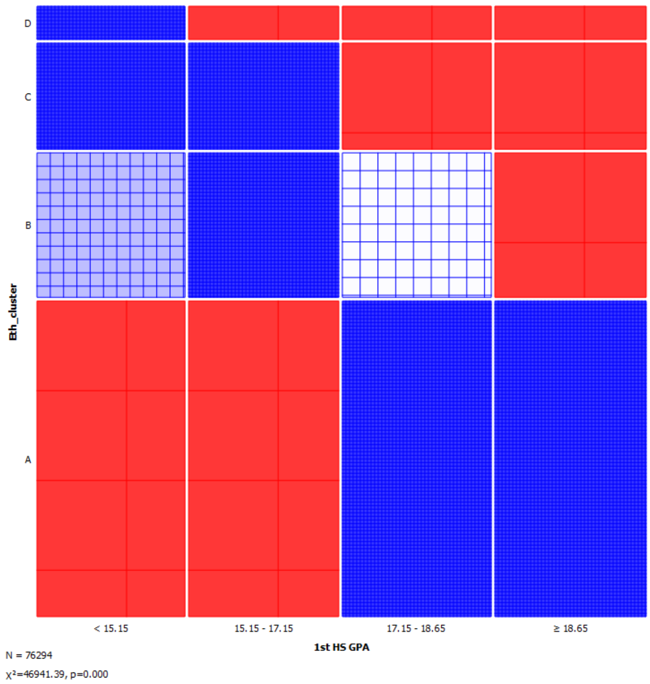


Figure 5.5. Sieve diagram - GPA 1st grade HS per initial achievement level



5.2 Junior HS, longitudinal analysis

The second dataset included the grades of the students in three sequential grades of junior high school. When we apply the same approach to the grades' data, we find that

78.3% of the students who were “Very strong” when they were in the first class are still high achievers when they are in the third class of junior high school (Table 5.4). In addition, a total of 21.60% (20.20% +and 1.40%) of the “Very strong” first graders see a decline in their performance when they move on to the third class, and 16.30% (15.90% and 0.40%) see a decline when they move on to the second class, although there were just a few students who had the best performance in the first class who ended up in the category with the worst performance.

Table 5.4. Frequencies of achievement level over time, based on initial clustering. (2nd dataset).

Cluster	**Class**	**1st _High School Cluster**			
		A	**B**	**C**	**D**
A	2nd class	83.70%	11.00%	0.10%	0.00%
	3rd class	**78.30%**	15.50%	0.60%	0.00%
B	2nd class	**15.90%**	65.70%	**13.20%**	**0.40%**
	3rd class	**20.20%**	60.50%	**22.90%**	**1.40%**
C	2nd class	0.40%	**22.30%**	60.90%	12.00%
	3rd class	1.40%	**22.50%**	61.00%	34.20%
D	2nd class	0.00%	0.90%	25.70%	**87.60%**
	3rd class	0.10%	1.50%	15.60%	**64.40%**

There hasn’t been a shift in the pattern of students who are placed in the category of ”very weak” in the first year of junior high school. In the third year of junior high school, almost two-thirds (64.40%) of these students are still performing at the same “Very weak” level. Also, from the students that underperformed in their first year of junior high school, 87.60% stayed at the lowest achievement level in the second class. Some of them show some signs of improvement, but not enough for them to be placed in a group higher than B. In the stacked charts (Figure 5.6, 5.7), we can observe the stability of ”very strong” and ”very weak” levels, as well as stability in the ”strong” (B) and ”weak” (C) levels. This stability showing an increasing lack of mobility between performance levels in junior high school.

Also, the fact that there are mixed patterns among students ranked in the average achievement (B, C) is combined with the fact that the percentage of students belonging to the strong level (B) who decrease their achievement is higher (22.30% in the second and 22.50% in the third grade) than those who increase it (13.20% in the second and 22.90% in the third grade)..

A consistent pattern emerges in the distribution of Grade Point Averages (GPA) in the second and third grades, based on initial clustering (Table 5.5). The averages are very close to being the same in both classes, and students who performed exceptionally well had the smallest standard deviation in GPA. This is because their scores were closer to the

Figure 5.6. Achievement level frequencies from 1st to 2nd HS).

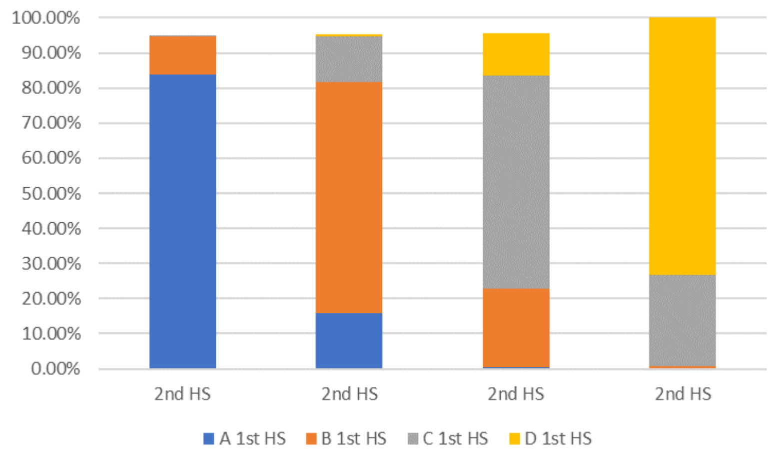
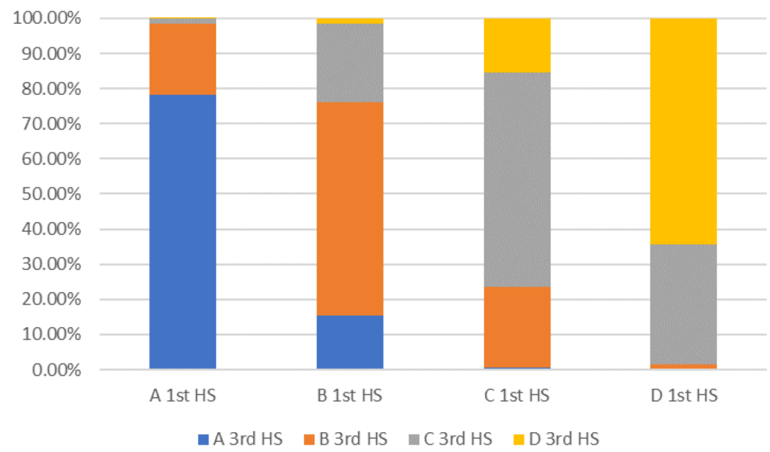


Figure 5.7. Achievement level frequencies from 1st to 3rd HS



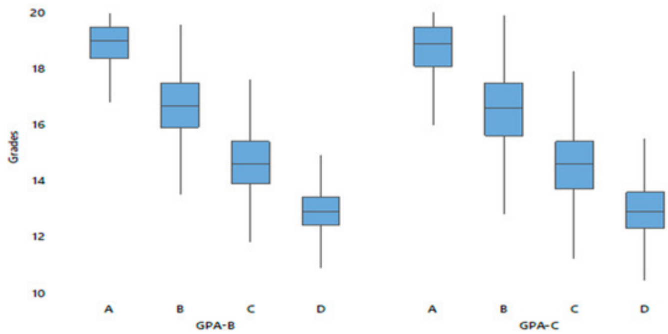
top of the scale, which ranged from 0 to 20.

Table 5.5. Descriptive statistics of GPA based on the initial categorization (2nd dataset).

Variable	A Class Cluster	Mean	SE Mean	St. Dev	Median
B class HS GPA	A	18.846	0.00522	0.855	19.0
	B	16.674	0.00742	1.095	16.7
	C	14.65	0.00873	1.075	14.6
	D	12.939	0.00900	0.876	12.9
C class HS GPA	A	18.716	0.00635	1.040	18.9
	B	16.525	0.00910	1.341	16.6
	C	14.61	0.01000	1.234	14.6
	D	12.996	0.00996	0.969	12.9

The non-parametric Kruskal-Wallis Test was used in order to analyze the statistical significance of GPA differentiation, which was based on the initial characterization of students who were enrolled in the first year of junior high school. The results were similar

Figure 5.8. GPA boxplots based on the initial clustering (2nd dataset).

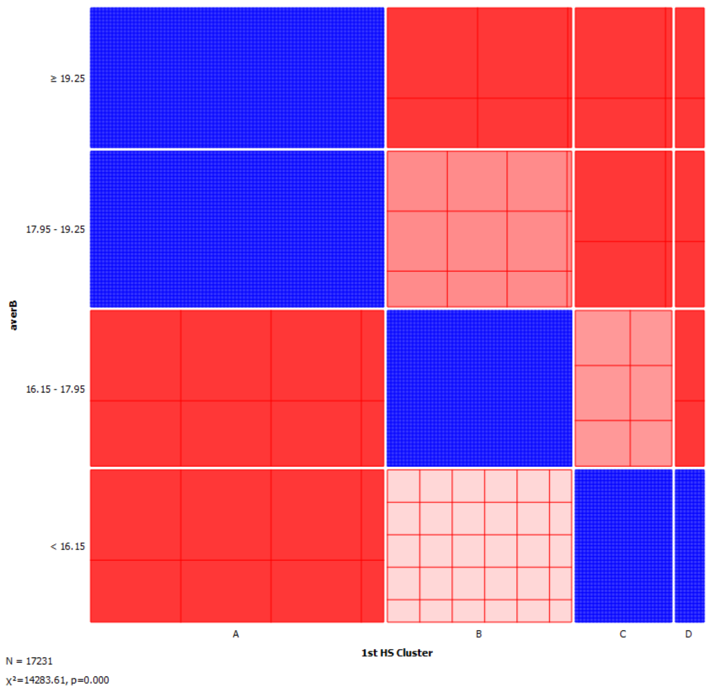


to those of the first dataset. Different initial classifications lead to statistically significant differences in GPA (Table 5.6).

Table 5.6. Kruskal-Wallis test of GPA based on the initial clustering (2nd dataset).

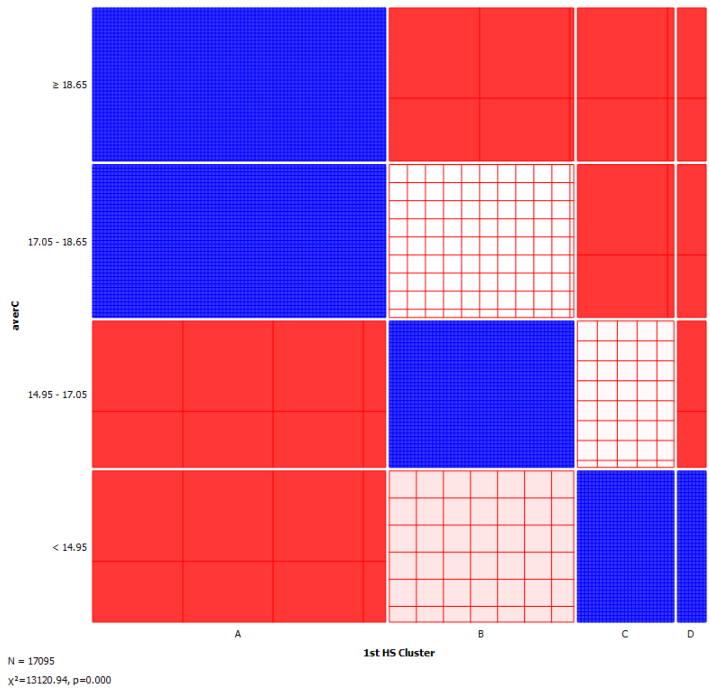
Class	Statistic	*p*-value
1st High School	77,827.42	0.0000
2nd High School	69,654.86	0.0000
3rd High School	63,991.20	0.0000

Figure 5.9. Sieve diagram - GPA 2nd grade HS per initial achievement level



The achievement path that students follow in junior high school based on their initial grouping in 1_{st} (Figures 5.9, 5.10). The blue areas in these charts show that the Grade Point

Figure 5.10. Sieve diagram - GPA 3rd grade HS per initial achievement level



Averages have a higher frequency than expected in a χ^2 contingency table. In contrast, the red areas indicate Grade Point Averages that have a lower frequency than expected. In each case, the statistical test of independence shows strong differentiation. The picture is more clear than in elementary school. ”Very strong” students expected to gain a GPA above 17.95 in the 2_{nd} grade and above 17.05 in the 3_{rd} grade. On contrary, ”very weak” students expected to gain a GPA below 16.15 and 14.95 respectively.

5.3 Findings

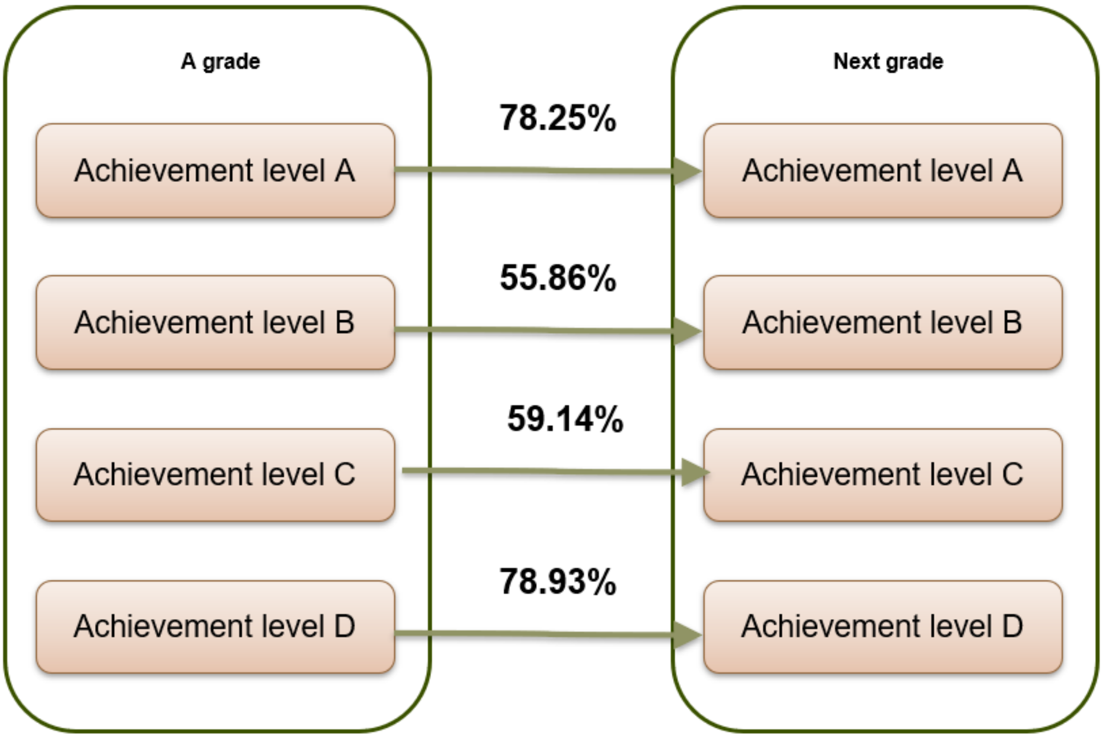
The second research question tested the longitudinal stability of students’ achievement levels over three years using two data sets. The first dataset, which included data from primary and secondary schools, showed stable levels of achievement, particularly for the ”very strong” and the ”very weak”. However, intermediate levels showed greater variation, with trends of falling performance suggesting increased difficulty in high school. The use of inferential statistical tests revealed statistically significant differences in GPA based on initial achievement levels.

The second dataset showed similar results. Most students who started out very strong and very weak remained at the same achievement levels. Results were similar with respect to GPA using non-parametric statistics. Both data sets demonstrated the stability of performance levels at extreme achievement levels, both in terms of scores and cluster size.

This finding is very positive for the ”very strong” students, but shows a very problematic situation for the ”very weak”.

The study in this chapter showed that there is stability between achievement levels based on the initial categorization of students. More intense is the stability between the levels of very strong and very weak students. As shown in the next graph, on average, the percentage of those who remain at the same level of high achievement from grade to grade reaches 78.25%. Similarly, for very weak students, this percentage equals 78.93%. Lower but highly significant levels of stability in the average levels of performance are found.

Figure 5.11. Average stability of student achievement level from one grade to the next



According to the study, most students tend to remain at the same relative level of achievement for several years. This stability indicates stability in the factors that influence achievement. Although stability provides predictability and consistency in learning, it also means that disadvantages are more difficult to overcome without constant intervention.

Schools can expect to face difficulties in substantially improving the outcomes of students performing well below grade level only through routine practices. Targeted, long-term interventions are needed (Brigman & Campbell, 2003). From a policy perspective, stability indicates the importance of early intervention strategies. It also suggests that internal and external social factors affecting students need to be considered. Ongoing assessment and personalization are crucial for increasing support for students as their needs and environment change over the years. One-size-fits-all approaches are less effective.

Chapter 6

School as institution of equal opportunities

In the years following the war, economic expansion and knowledge possession contributed to a significant improvement in the social and economic position of individuals with a high degree of knowledge. School achievement, along with the development of other abilities, is a significant factor in knowledge acquisition. Students from lower socio-economic status (SES) might benefit from this trend.

Following this school of thought, we would like to think that our educational system is a great equalizer that gives all children a fair shot at success and the chance to realize their dreams. The school as institution of equal opportunities. If this is in fact the case, then one may anticipate that certain demographic data pertaining to the students themselves do not have a link with their academic achievement. In other words, one might hope that other demographic data are uncorrelated with academic achievement.

6.1 Data preparation

Examining whether the school functions as an equal-opportunities school, was the aim of third research question. This aim supported by studying the independence of students' achievement from above non-academic factors. In order to test this idealistic expectation, the maximum number of demographic characteristics related to socio-economic status that are available in the EMIS database was requested. Because the General Data Protection Regulation (GDPR) puts limits on what can be shared, only a few socioeconomic and demographic features about the students have been made available. For this reason, we could not study factors that maybe give a more detailed picture of the overall profile of the students.

The available variables were the level of school achievement of the students, as obtained from the previous research question, and the occupation of the guardians as a demographic (non-academic) variable. The variable of guardians' occupation was a free field in the information system, resulting in incomplete records. However, a very high percentage of 70% were completed. For this reason, inferential statistics were used in order to draw conclusions. In particular, the χ^2 statistical test was used, which we briefly present below.

Another issue that raised was the large number of professions that were registered in the information system. It was imperative to group them in some way in order to draw conclusions. For this purpose, an internationally recognized categorization, that of the International Labor Organization, was preferred (ILO, 1990). In this way, the hundreds of occupations that had been registered were categorized into specific groups according to the criteria set by the International Labor Organization. The new variable was added to the data set, allowing statistical testing to be carried out in relation to the differentiation of the performance level variable in relation to the variable of the guardian occupation category.

The new variable, created after running the algorithm determining the achievement level that each student belongs to, was also added to the dataset. It was an ordinal variable, so that the clusters-levels of achievement could be sorted from lowest to highest performance category. With this new variable, it was possible to look at how achievement levels differed based on the demographics of the dataset. The guardian's occupation was reclassified using the ILO categorization (ILO, 1990).

The choice of appropriate statistical techniques for data analysis is closely related to the type to which each of the variables under consideration belongs. The non-parametric x^2 test can be used to examine whether or not there is independence between categorical variables. The conditions for the application of this criterion are as follows:

1. The variables are qualitative (on a categorical or ranking scale).
2. The sample must have been selected in a random manner.
3. The observations must be independent of each other.
4. All expected frequencies are greater than 1.
5. At most, 20% of the expected frequencies are less than 5.

The conditions for applying these criteria are fulfilled in the datasets. The variables are categorical. Random sample selection is not applicable since the data is complete and the variations in relation to the population are only related to the data cleaning, from which records that did not provide information were removed. As can be seen from the relevance tables, all expected frequencies in all cells of the tables are greater than unity, and there are no cases where expected frequencies are less than 5. Therefore, all the conditions are met in the case of these datasets.

It should be stressed that for the purposes of analyzing such a large proportion of the total population, the importance of inferring results through hypothesis testing is limited.

Essentially, we refer to the whole population and we are interested in studying the frequencies - probabilities of occurrence of the individual categories and their differentiation. For this purpose, we studied the percentage difference between expected and observed frequencies at each level of student performance, based on the contingency tables.

In order to test the above hypothesis, we have compared the frequency of students who fall into each of the four levels of academic achievement, based on their parents' occupation to the theoretical frequencies from contingency matrix of χ^2 tests.

6.2 Occupation of the guardians' effects

According to Cole et al., (2005), the education level, the occupations, and financial circumstances of the parents define the family's social and economic status. These family-related factors have an impact on the child's academic path in terms of economic benefits, and a child who comes from an underprivileged family at the start of compulsory education is at a disadvantage compared to his/her peers who come from privileged parents and have already acquired a culture that is consistent with the dominant culture that the school promotes from the family environment.

The social, economic, and educational level of the parents has a significant impact on the academic path and development of the child, either directly by providing or not providing material benefits such as tutorials and foreign languages, or indirectly by providing or not providing an educational culture and culture, which is the most catalytic influence (Coleman, 1968; Bourdieu, 1973). High-class parents provide their child with the best educational experiences outside of school, which, when combined with the high goals they set, creates a productive field for a successful and rewarding educational life (Cole et al., 2005).

In this study, was found that there were significant differences between the actual percentage and the theoretical percentage of the frequency of each occupation in the levels of "very strong" and "very weak" students ($\chi 2 = 603495.000$, *p-value* = 0.0001). The percentage difference between observed and expected performance for the low-performance and high-performance categories are presented in Table 6.1.

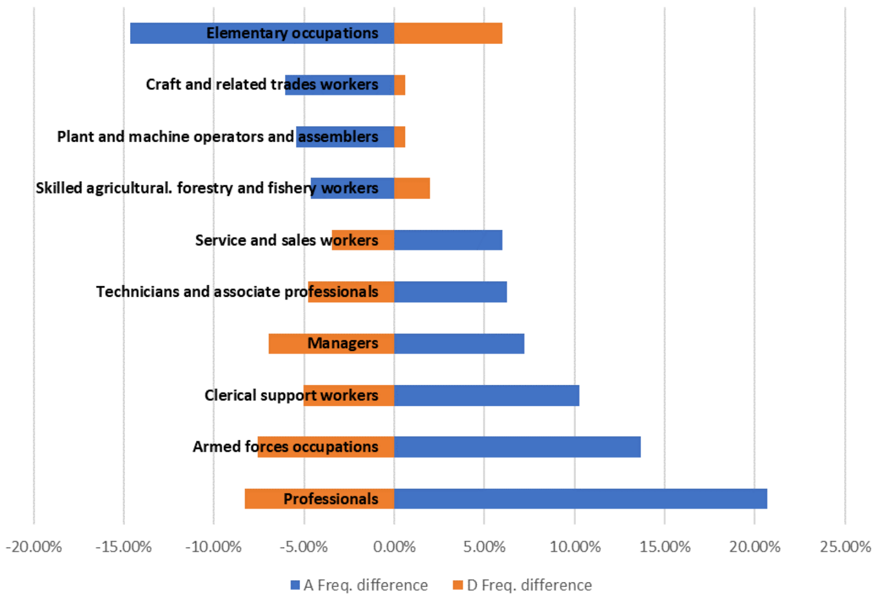
There appears to be a disparity. Students whose parents or guardians are self-employed professionals, teachers at all levels, officers in the military services, private and public officials, have been shown to receive higher grades than predicted. On the other hand, the students whose parents or guardians identify themselves as unskilled workers, manual workers, farmers, or stock breeders are significantly likely to have poor levels of academic achievement.

The falsity of our idealistic hypothesis is glaringly obvious right away. There are some professions in which children are significantly more likely to be "very strong" students

Table 6.1. Differences between the actual and theoretical frequencies of levels A (Very Strong) and D (very Weak) (based on ISCO categorization).

ISCO	Achievement level frequencies	
	A Average/ St. Deviation	D Average/ St. Deviation
Professionals	20.72%/14.70%	−8.28%/5.12%
Armed forces occupations	13.69%/1.99%	−7.56%/2.43%
Clerical support workers	10.30%/11.62%	−5.01%/4.18%
Managers	7.23%/8.63%	−6.97%/3.42%
Technicians and associate professionals	6.26%/12.37%	−4.75%/5.46%
Service and sales workers	6.00%/13.74%	−3.45%/6.06%
Skilled agricultural. forestry and fishery workers	−4.61%/11.56%	1.97%/5.91%
Plant and machine operators and assemblers	−5.42%/5.07%	6.30%/3.14%
Craft and related trades workers	−6.03%/8.62%	6.30%/4.85%
Elementary occupations	−14.64%/5.77%	5.98%/4.27%

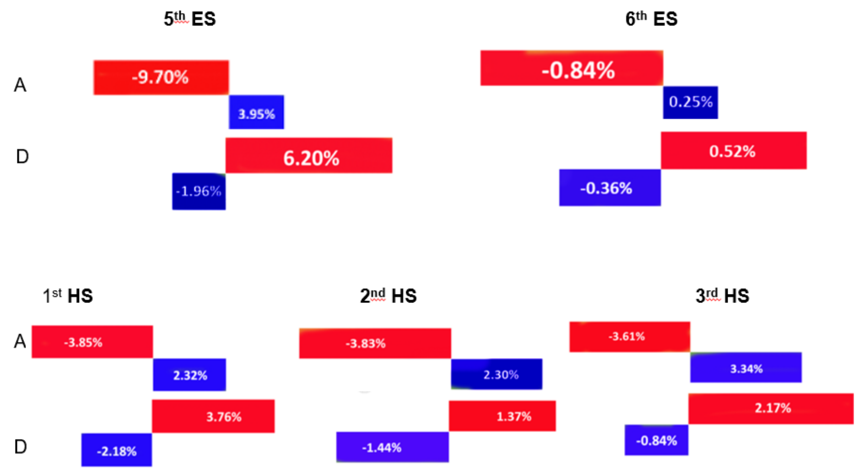
Figure 6.1. Differences between the actual and theoretical frequencies of levels A (very Strong) and D (very Weak) (based on ISCO categorization).



and considerably less likely to be ”very weak” students, whereas for other professions, the situation is exactly the opposite. This is especially true when it comes to the case of very excellent academic achievement, where the children of self-employed professionals have a far larger possibility of achieving well in school than other children do. It is just this kind of performance that, in a few years’ time, will make it possible for them to enroll in one of the most prestigious schools, and by building on it, they will be able to continue their ascent into the upper social strata.

Looking over time at the differences in performance in the categories of ”very strong”

Figure 6.2. Differences in achievement levels of students from ”Intellectual” (blue) and ”Manual” (red) occupations.



and ”very weak” students between the different categories of guardians' occupations, this pattern becomes evident. The same phenomenon is observed in all grades of primary and secondary school. Students with guardians who come from professions of a more intellectual type appear to outperform, compared to students from backgrounds in which their parents are in mainly manual professions.

Children whose parents are employed in manual occupations have a lower possibility of following such a path and are therefore more likely to remain in the same social classes. To put it another way, the findings of this chapter suggest that the educational system in Greece is not, in fact, the great equalizer that its proponents assert it to be.

6.3 Analyzing the overperformance category

We also studied the relative frequency of over-performance in the ”very strong” category, which includes the specific professions that were reported, in order to study the likely differences in performance in the category of ”professionals” over-performing. The high frequency of certain occupations should be noted. By studying the over representation rates (Table 6.2), we found that students with guardians who were doctors, educators, chemical or environmental engineers, pharmacists, lawyers, teaching staff of higher education institutions, and electric engineers had a more than 25% over representation. Doctors declared 587, although a few more declared their medical specialties. Also, 1545 guardians were declared educators. Students in both of these occupational categories over-performed by more than 30%,. These two findings combined reinforce the importance of over-performance. On the opposite end of the spectrum, lower-than-expected students with guardians who were freelance professionals performed lower than expected, making

Table 6.2. Over - under achievement in "Professionals" category.

Guardian's Occupation	**Rank**	**N**	**Expected**	**Percentage**
Doctor	A	587	394	32.85%
Educator	A	1545	1077	30.27%
Chemical Engineer	A	74	53	28.54%
Environmental Engineer	A	10	7	27.89%
Pharmacist	A	48	35	26.89%
Lawyer	A	187	137	26.73%
Teaching Staff of HE	A	57	42	25.78%
Electrical Engineer	A	138	103	25.45%
Forester	A	23	17	24.76%
Agronomist	A	184	146	20.84%
Agronomist - Surveyor	A	36	29	19.88%
Veterinarian	A	24	19	19.88%
Dentist	A	70	56	19.65%
Mechanical Engineer	A	227	183	19.31%
Civil Engineer	A	315	258	18.20%
Architect	A	64	53	17.37
....	A	...	...	....
Psychologist	A	10	11	-5.76%
Artist	A	15	18	-18.58%
Journalist	A	40	48	-20.18%
Freelancer	A	2792	3820	-36.83%
Music Artist	A	42	59	-40.79%
Contractor	A	58	82	-40.90%

up the largest category with 2790 cases. In conclusion, students with guardians who are doctors or educators perform better in this dataset.

6.4 Findings

This chapter examined whether the school system functions as an equal opportunity institution by investigating the independence of students' achievement levels from non-academic demographic factors such as their guardians' occupation. Statistical analyses using chi-square tests revealed significant differences between students' actual achievement levels and the theoretical levels expected based on their guardians' occupations. Specifically, students from more intellectual professional backgrounds such as doctors, educators and engineers significantly outperformed expectations for the high-achieving "very strong" category, whereas students from manual occupational backgrounds such as manual workers, farmers and stock breeders unperformed expectations for this category.

Additionally, students from manual backgrounds significantly over-represented the low-achieving ”very weak” category compared to expectations.

A closer analysis within the over-performing ”professionals” category found that students with guardians working as doctors, educators, chemical engineers and other specific intellectual professions exhibited over-representation rates of 25-30% or more in the high-achieving category. In contrast, students whose guardians were freelance professionals, artists or contractors represented lower than expected proportions in this high achievement group. The differences in achievement levels based on occupation persisted across all grades of primary and secondary schooling.

Taken together, these findings indicate that students' academic achievement is strongly linked to their guardians' occupation. This suggests that non-school factors related to one's socioeconomic background, such as differing cultural and material resources available at home, provide an unequal starting advantage not compensated for by the school system (OECD., 2004). This lack of independence from non-academic factors also statistically demonstrates the failure of school to function as an equal opportunity school (Bourdieu, 1966; Coleman, 1968). Therefore, the study based on data, demonstrates that the education system in this context does not fully function as an equal opportunity institution as it fails to achieve independence of students' achievement from external demographic factors.

Chapter 7

Other demographic effects

Other demographic features that were present in the datasets concerned a) the gender of the students and b) their region of residence. This chapter examines the effects produced on student achievement levels by these characteristics. Research question four also concerns the effects of gender and region of residence on student achievement.

Differences based on these features provide useful information on the achievement of boys and girls, which has been a constant topic of academic research. Differentiation by region of residence can provide clues to areas of the country that need further intervention in order to improve student performance.

7.1 Gender effect

The study of the impact of gender on student performance has lately been rekindled in the context of attempts by educational institutions in various nations throughout the world to promote gender equality in education. Since the early 1990s, when it became apparent that boys were increasingly underperforming in various examinations, the impact of gender on academic achievement has been the subject of inquiry and research, mostly in Anglo-Saxon countries (Crombie et al., 2005; Meelissen & Luyten, 2008; R. B. King & McInerney, 2014b).

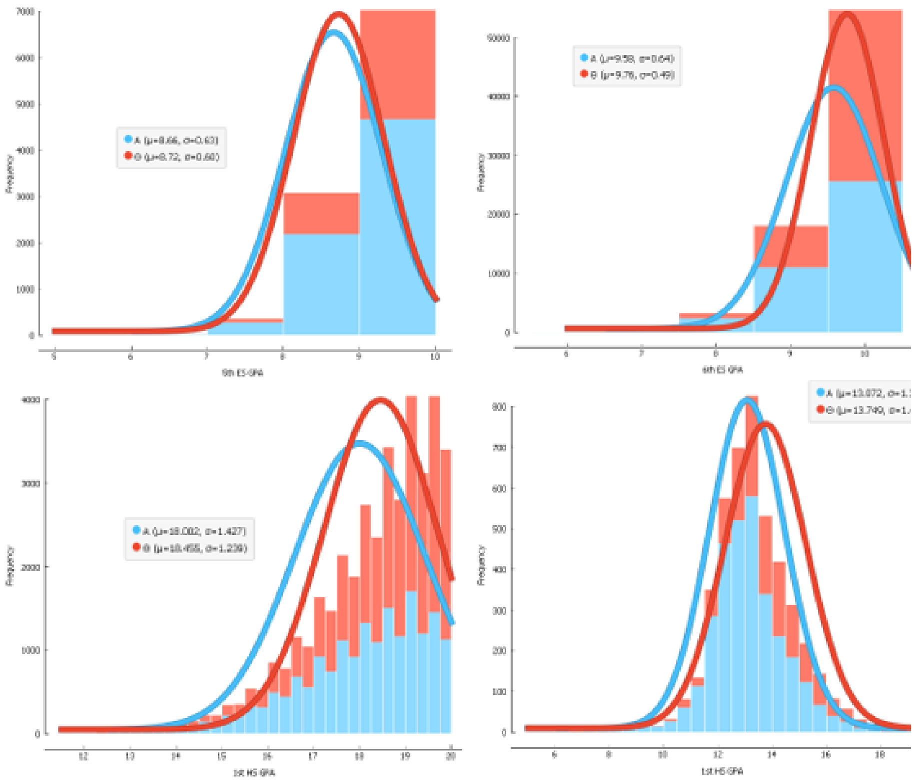
By using data for the whole country, this study provides additional evidence, as there is no need for an induction of the results. In contrast, most research on the issue of gender-based achievement is done on data from international competitions or small samples. First, the GPA distributions between boys and girls overall were examined.

In our study, the analysis of the distributions of GPA scores between boys and girls yielded insightful findings (Figure 7.1). The examination of overall performance has highlighted a slight disparity between the two genders. It is worth noting that the variance across the entire student population is minimal, especially at the primary school level.

However, as students progress to secondary schools, the disparity becomes more pronounced, nearing a difference of approximately one point, specifically 17.317 for boys and 16.352 for girls. These findings shed light on the overall performance patterns between boys and girls at different educational stages. There seems to be a better performance of girls than boys. But the difference is not particularly large.

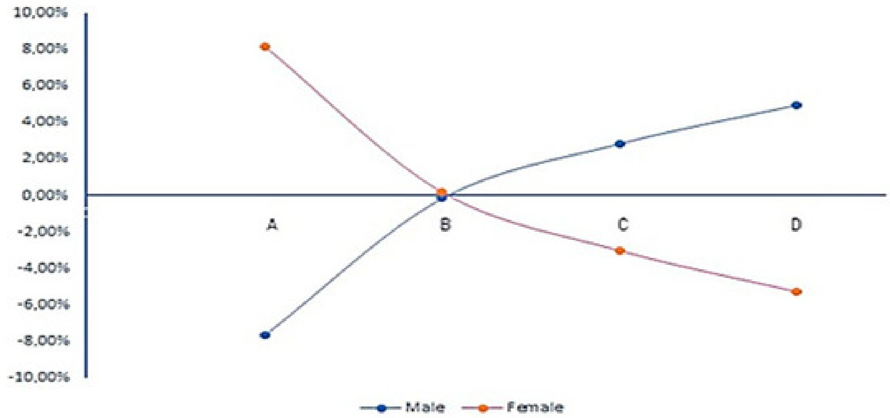
The differentiation between ”very strong” and ”very weak” students in relation to their gender does not change in secondary school. In ”very strong” category of the 1st junior high school the difference in average performance is smaller, with girls having an average score of 18,445 and a standard deviation of 1.239, while boys have an average of 18.000 and a standard deviation of 1.427. Also in the category of ”very weak” students, female students have an average score of 13,749 and a standard deviation of 1,494 in the first grade, while boys have an average score of 13,072 and a standard deviation of 1,385. These differences do not appear to be particularly large.

Figure 7.1. Distributions of GPA (based on gender).



Looking deeper and in relation to student achievement levels, we discover a different situation. When the frequency of boys and girls at the two extreme levels of achievement is examined, the situation looks quite different. From the statistical tests significant differences were found ($\chi 2$ = 17,514.29, *p-value* < 0.0001). The percentage differences between the observed and expected frequencies of the four different clusters with regard to gender are significant. Figure 7.2 demonstrates that females have a frequency of high performance that is higher than expected (8.15%), while their frequency of low performance is lower than what would be expected. In contrast, males have a higher frequency than expected of having low academic achievement and a lower frequency than expected of having good academic achievement. The probability of boys and girls occurring varies by achievement level in all three classes presented in Figure 7.4.

Figure 7.2. Differences between the actual and theoretical frequencies of levels A (Very Strong) and D (Very Weak) (based on gender).



A comprehensive examination of the pattern of achievement between boys and girls provides us with a clear and enlightening picture (Figure 7.3). It is evident from our findings that girls consistently outperform boys in the ”very strong” category starting from 5th grade up until the 3rd grade fo HS. In contrast, boys demonstrate superior performance in the ”very weak” student category during the same time period. Furthermore, it is noteworthy that the underperformance rates of boys escalate progressively over time, reaching a threefold increase by the time they reach third grade. This longitudinal analysis reveals a discernible and concerning negative trend in boys’ performance as they progress through high school.

To gain a comprehensive understanding of the overachievement of female students in primary and lower secondary schools, let us turn our attention to Figure 7.4. This visual representation highlights a notable discrepancy in the likelihood of boys being categorized as ”very strong” in the 5th grade, with the situation appearing to worsen as they advance to subsequent grades. Conversely, when it comes to the category of ”very weak” students, boys are more than twice as likely to fall within this classification. Upon comparing the individual charts, a strikingly clear linear and inverse trend between achievement levels in all grades becomes apparent. These findings further emphasize the consistent and significant disparities in academic achievement between boys and girls throughout their educational journey.

There are many studies demonstrating that female students have been outperforming

Figure 7.3. Differences between girls (pink) and boys (blue)

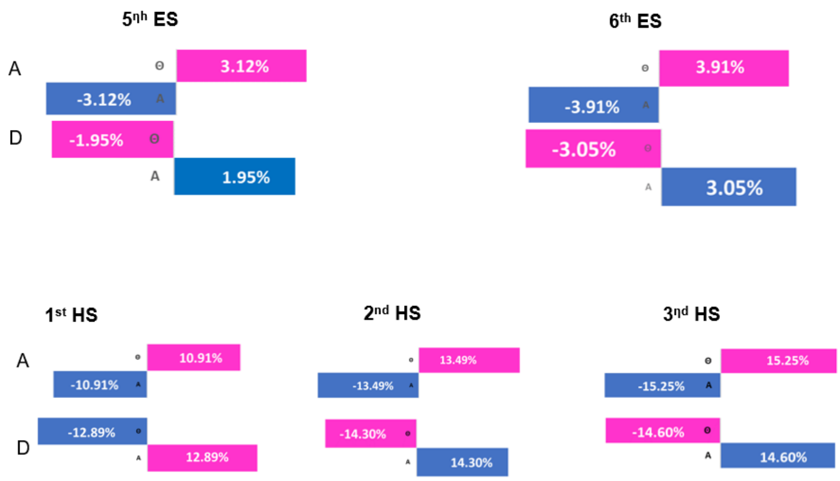
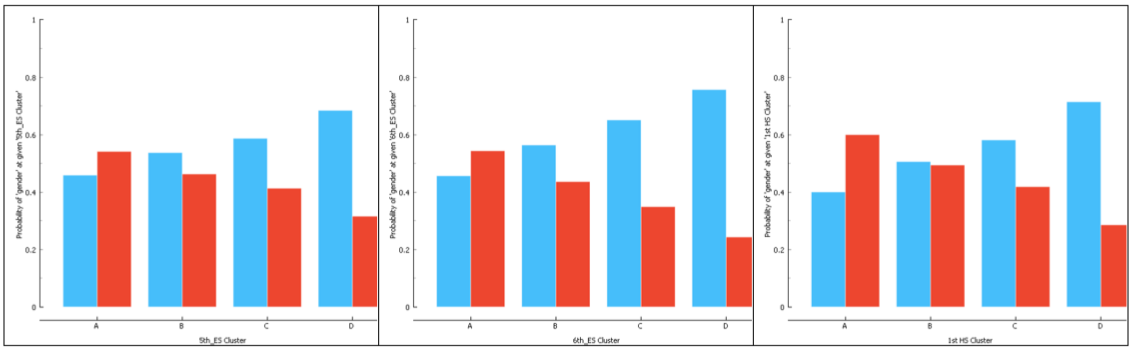


Figure 7.4. Probabilities of achievement level by grade and gender, 1_{st} dataset.



male students in recent years. Even in fields where boys have historically excelled, like mathematics and science (Crombie et al., 2005; Herbert & Stipek, 2005; Meelissen & Luyten, 2008; Valiande et al., 2011; R. B. King & McInerney, 2014b). It should be noted that while our findings are consistent with the trend showing that there is a general over-achievement of girls, according to the literature, the issue seems to be much more complex, with significant variations observed depending on the study context (Bouiri et al., 2022). But the general conclusion of all these studies is that girls will do better at school than boys. This applies to different subjects and different levels of study in primary and secondary education.

The current study showed, using aggregate country-level data for three consecutive years, that girls' school performance is better than boys'. The fact that the analysis is based on an objective assessment of achievement levels rather than on statistical measures of the total population or on the researcher's determination of achievement levels provides additional support for the results. The better performance of girls was also confirmed at the level of the overall education system. For reference, other studies on smaller samples

(Crombie et al., 2005; Herbert & Stipek, 2005; Butler, 2014; OECD, 2015; Crues et al., 2018; Breda & Napp, 2019).

This study found that girls are better students than boys. This is a general finding found in many studies. The findings are consistent with many previous researches (Crombie et al., 2005; Vandecandelaere et al., 2012; R. B. King & McInerney, 2014b). In the study of Bouiri et al., (2022), girls appeared to do better than boys in secondary education. Boys in particular develop gender stereotypes and consider themselves academically superior in motivation, ability, performance, and self-regulation. Zell, Krizan and Teeter (2015) found that the effects of gender were small. Gender differences in reading and writing appeared in the study of (Reilly et al., 2019). The key finding in the study of Terrier (2016) is that teachers' gender biases have a high and significant effect on girls' progress relative to boys in both mathematics and French. During high school, teachers' gender bias against boys explains 6% of the lag of boys over girls in mathematics.

The study of Siddiq & Scherer, (2019) revealed a better performance by the girls in ICT and significant variation across studies of these gender differences. This difference was also partly due to the level of education of the students, which was slightly higher in primary schools. The study of Breda & Napp (2019) using data from PISA competition showed also that girls do better than boys. They found no gender differences in how students see themselves in mathematics, how interested they are, or how they feel about it. Student differences in how well students do in maths and reading are a much better way to explain why girls are less likely than boys to want to study maths.

The different conclusions that emerge relate mainly to the factors affecting performance rather than the general trend. The general trend in the studies confirms the findings of the current thesis. Girls do better at school than boys and even better in subjects such as mathematics and ICT. Not only do boys express lower motivation on average (Butler, 2014), but they are also less engaged (Wong et al., 2012) and do worse in secondary education than girls (Voyer & Voyer, 2014). Although the generalization due to the secondary data available to us may be dangerous as it treats boys and girls as two homogeneous groups, it seems that the general picture is this. The different conclusions in the studies that emerge relate mainly to the factors affecting performance rather than the general trend. The general trend in the surveys confirms the findings of the thesis.

What this thesis contributes to the research field is the use of complete country-level data. The findings of our research reflect the situation regarding the over-achievement of girls at the country level over a period of three school years at the elementary and middle school levels. In contrast, the other studies were limited to samples or competition data, with all that this implies for the generalizability of the findings. Previous research primarily looked at factors influencing scores, but this study strengthened understanding of the baseline reality: that girls in the country tend to achieve at higher levels than boys in primary and secondary grades. It helps contextualize factors within the overarching pattern revealed through population-level analysis.

Overall, through its large-scale design and total data, this thesis meaningfully advanced knowledge about gender disparities, from the level of samples and assessments to declarations that can be confidently applied to the country's students and education policies. This statewide perspective provides valuable insights to inform targeted interventions aiming to narrow gender gaps.

7.2 Region of residence effect

The region of residence can be a source of variation in student performance, to the extent that it corresponds to different conditions in relation to climatic, economic and social circumstances. By analyzing student achievement levels by region, we can gain insights into how the local environment affects student achievement. For example, changes in weather, differences in temperature, weather conditions, and seasonal variations have a significant impact on students' educational experiences and outcomes. Similarly, the economic conditions prevailing in the district, as well as income levels, affect academic achievement.

In addition, social dynamics within specific regions, such as cultural norms, structures and networks within the community, and socio-demographic characteristics, create a complex web.This study can provide further insights into how the specific factors in each region lead to specific educational outcomes. Exploring these elements helps identify patterns and trends, providing a basis for interventions and other measures. For all these reasons, it was considered useful to examine in a similar way the differentiation in the level of achievement.

The dataset included the field "Educational Directorate" to which each student belonged. There is a correspondence between the education directorate and the region of residence since most education administrations are identical to the provinces into which the country was divided. The only difference is in the metropolitan areas of Athens and Thessaloniki, which are divided into individual directorates.

It should be pointed out that the specific level of administrative structure is very broad and it is not possible to identify all individual social and economic conditions. In the EMIS database there are fields that would allow a much more precise identification of the social and economic conditions in which students live. Unfortunately, however, they were not provided, despite our persistent efforts.

The chi squared test was ran, taking into account the elementary and junior high school districts where the students live ($\chi 2$ = 6612,839, *p-value* = 0.0001). The largest and smallest percentage differences in the high and low achievement level were found. The areas that display the greatest disparity between levels of academic achievement are compared in Table 7.1

Table 7.1. Differences between the actual and theoretical frequencies of levels A (Very Strong) and D (Very Weak) (based on region).

Region	**A**	**D**
Kastoria (Pr.)	18.40%	−6.95%
Arta (Pr.)	16.17%	−3.75%
Rodopi (Pr.)	16.02%	−5.83%
Trikala (Pr.)	14.69%	−2.72%
Karditsa (Pr.)	13.75%	3.43%
Karditsa (Sec.)	11.82%	−4.02%
Chios (Pr.)	11.47%	−5.02%
Larisa (Sec.)	11.00%	−6.10%
Chios (Sec.)	10.93%	−4.65%
Kastoria (Sec.)	10.45%	−8.53%
Ioannina (Pr.)	10.39%	−4.24%
Thessaloniki (Pr.)	10.13%	−5.10%
......	....	.....
Lefkada (Pr.)	−12.50%	5.29%
Corfu (Pr.)	−10.87%	5.51%
Religious Directorate	−14.17%	6.52%
Lasithi (Pr.)	−10.52%	7.00%
Rethymnon (Sec.)	−10.31%	7.62%
West Attica (Pr.)	−13.04%	8.63%

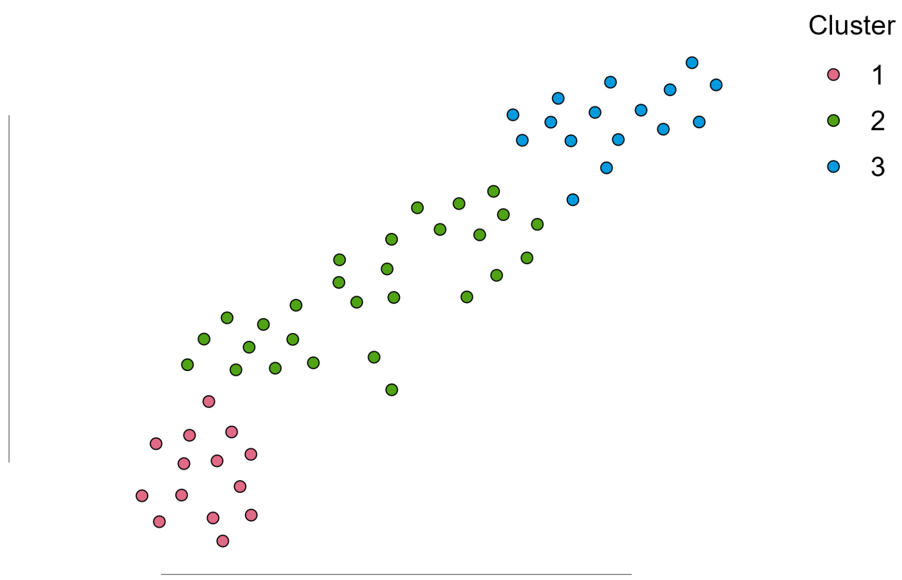
It was also examined whether it is possible to group the overachievement or underachievement rates of students into the categories of ”very strong” and ”very weak” students. Using the percentages in Table 7.1 as input, the K-Means algorithm was run in conjunction with the Elbow method to pre-determine the number of clusters. Three clusters were identified, into which the over- or under-performance rates of the students were categorized Table 7.2.

Table 7.2. K-Means Clustering on regions’ differences

Clusters	**N**	**R^2**	**AIC**	**BIC**	**Silhouette**
3	58	0.782	36.860	49.220	0.470

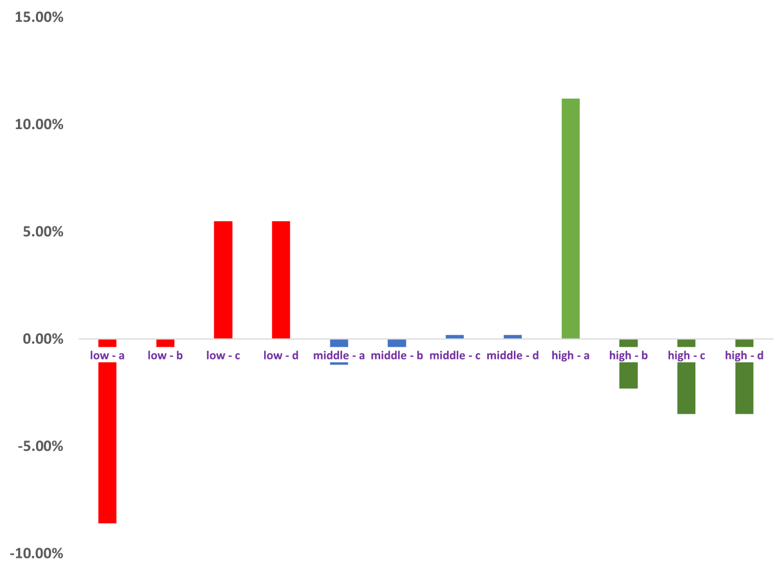
In this way, three clusters of districts were obtained based on the level of performance. The districts were categorized as low, medium, and high achievement areas. Figure 7.5 shows the three groups of areas. The average percentage of over- and under-achievement in each achievement level and the low, medium, and high achievement areas were also calculated.

Figure 7.5. Clusters of regions by achievement level.



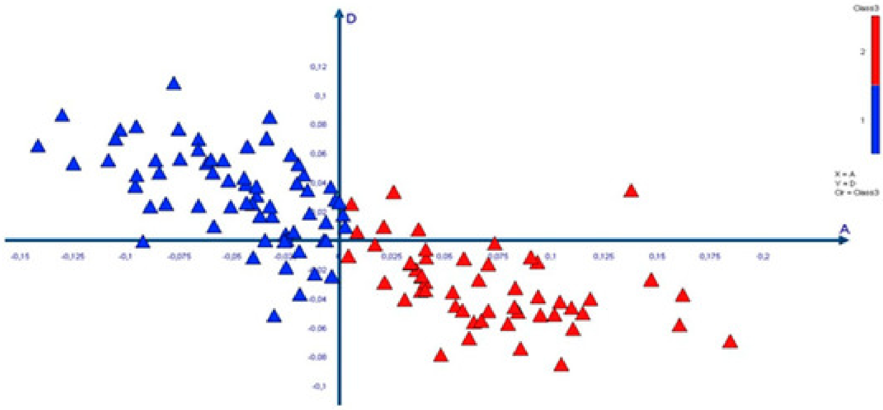
As can be seen in Figure 7.6, the high-performance regions have a higher percentage of overperformance than the low-performance areas. The picture is the opposite in the low-performing areas, where the very weak category of students dominates. Average-performing districts are not characterized by overachievement or underachievement.

Figure 7.6. Over and underachievement, based on regions' clusters



Another interesting fact is that the percentage of students who are characterized as ”very strong” is lower in regions that have a high frequency of students scoring in the ”D” category. As a result, the majority of the observations in Figure 7.7 are located in the second and fourth quartile. There appear to be regions in which factors favoring education outweigh others.

Figure 7.7. Differences between the actual and theoretical frequencies of levels A (Very Strong) and D (Very Weak) (based on region).



Finally, the main findings on student achievement in the different regions of residence refer to the significant differences between them, with some having a much higher or lower percentage of students who failed or excelled. Districts can be clustered into three groups according to their overall level of performance: high, medium and low. In high performing districts, the percentage of students who excelled was higher, while in low performing districts, the percentage of failures was high.

In the very last line of Table 7.1, there is a reference to Western Attica, which is a region that is considered to be degraded and is home to a significant population of minority groups, such as the Roma. In addition, there is a lower percentage of kids with low performance in places that have high levels of student achievement and vice versa. Roma students are a special category. A report o the European Union Agency For Fundamental Rights (2014), which included 11 countries in the European Union showed that around 14% of Roma children do not receive systematic schooling in the EU. Greece was in first place, with 43% of Roma stop attending school. Of the Roma who are in school age , between 30% and 40% have not attended school, also the percentage of children who eventually manage to graduate from primary education and continue is very low. It is noteworthy that West Attica does not show a similar underperformance in secondary education, a fact that seems linked to the almost zero percentages of Roma children attending high school.

7.3 Findings

The effects of the area of residence on student performance did not allow the extraction of specific patterns. This was due to the large geographical area covered by the regions, which constituted the variable provided. Consequently, each instance included many different characteristics that could not be identified. For example, there is a large difference in population density within the districts, which include urban, semi urban, and rural areas. The density of the student population also differs strongly between urban and rural areas, but unfortunately, such a distinction was not possible from the data we received. Mountainousness and insularity are also differentiating factors, but no such differentiation was obtained from our data.

Despite the obstacles of using such a wide region, some useful conclusions were drawn. The areas of residence can be divided into three groups, with the high-performing group of areas being over-performing very high achievers. The group of low-performing areas shows an over-presentation of very weak students. A linearity was also found in the rates of over-presentation and under presentation in the high and low performing categories.

The only exception was the consistent presence of a particular geographical area in the very low-performing category. This is the region of West Attica, which is home to a significant number of minorities, such as Roma students. This is a minority that has been the target of many educational interventions. Even today, the Roma have a hard time fitting in at work and in society because of their unique traits and because of stereotypes and prejudices. In particular, the traits of poverty and social exclusion are particularly pronounced in the Roma population, where they are confronted on a daily basis with political, economic, social, and cultural inequalities. Levels of paid employment for Roma vary across Europe, but are usually significantly lower than those of the majority population. A high level of discrimination against job-seekers is also evident in many cases (Bartlett & Gordon, 2011). Research suggests that Roma employment is characterized by low skills, low pay, and precarious employment combined with limited opportunities for progression (Ahmed & Rogers, 2016).

In terms of the education and literacy of Roma children, the educational disadvantage among them is evident in all countries. A 2008 study in Greece showed that 54.7% of Roma did not go to school at all, 33.4% completed only some classes of primary school, 7% finished primary school, 3.4% attended some classes of secondary education, 0.5% graduated from secondary education; and about 1% attended some classes of secondary education. Frazer and Marlier (2011) found that 54% of parents said their children have never been to school. This shows how socially isolated they are in school.

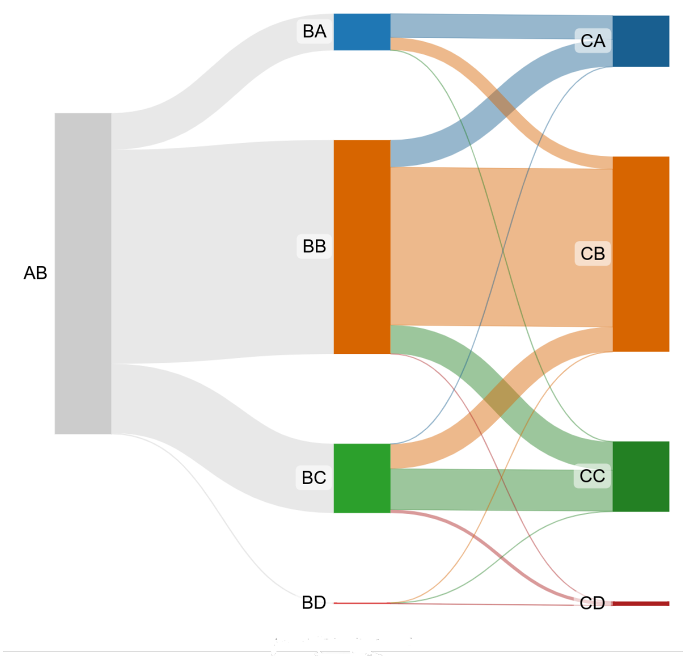
In our study, a well-known problem of educational policy has essentially reappeared. The education of Roma students has been a policy issue for decades. Many different efforts have been made at local and national level, with different results. The present study indicates, and through its performance, that the problem remains.

Chapter 8

Remedial teaching evaluation

Inequality in student achievement is an international phenomenon (Autor, 2014; H. C. Hill, 2017; Easterbrook & Hadden, 2021; Paulus et al., 2021).In our study we found a stability in the level of ”very weak” students, from year to year (Figure 8.1). This is a challenge for educational policy. One proposed response is the educational intervention of Remedial Teaching. This was initiated in Greece from primary education, with the aim of providing assistance to disadvantaged students to improve their performance. This particular intervention requires significant financial and human resources. It is considered successful and has been extended to secondary education. In this chapter we use the data to conduct an objective, data-based evaluation of the impact and effectiveness of this intervention.

Figure 8.1. Transitions between achievement levels from 5th ES to 1st HS



8.1 Data and analysis

Using the first dataset we were able to test the progression of student achievement in 5^{th}, 6^{th} grade of ES and 1^{st} grade of HS. However, performance is not fully comparable due to differences in curricula. Table 8.2 shows the courses for each grade. We note that there are e.g. ”Social and civic education” not taught in the first secondary school and ”Ancient Greek language” not taught in primary school. In addition, there are subjects that are not generally taught such as ”Italian language” and ”Religious education” for which some students request and receive an exemption. Consequently, achievement is not directly comparable. In order to overcome this, the students' Grade Point Average was used as the main parameter for assessing academic achievement and monitoring students' progress.

The scoring of students differs; in primary school, the scores are set in a range from 1 to 10, and in secondary school, from 0 to 20. High scores are more difficult to achieve as students move to more difficult grades. To overcome these problems, the approach used in the first research question is followed. Students' GPAs are grouped into different performance groups using unsupervised learning. It is found that there are four performance groups for the years under study. The clusters statistics identified are summarized in Table 8.1.

Table 8.1. Groups of academic achievement, contains the GPA and St. Deviation.

Cluster	**5^{th} grade ES**	**6^{th} grade ES**	**1^{st} grade HS**
A	9.988 / 0.002	9.979 / 0.023	19.070 / 0.038
B	9.727 / 0.029	9.680 / 0.264	17.144 / 0.071
C	9.008 / 0.018	8.912 / 0.081	15.205 / 0.118
D	8.180 / 0.066	8.263 / 0.574	13.073 / 0.256

Each group was then ranked by GPA in descending order. Since this ranking takes into account the way GPA are distributed within the same class, it overcomes the problems of different grading in different classes. In this way, even if students' GPA have dropped from 5^{th} to 6^{th} grade, students who have the same academic achievement will remain in the same achievement group.

Table 8.3 shows the percentages of over-representation or under-representation of each professional category of guardians at the levels of 'very strong' and 'very weak' student achievement. It becomes clear that there are 20.72% more students with guardians who are professionals than would be the case if the parent's occupation did not affect academic achievement and students were uniformly distributed across achievement levels. Similarly, among the ”very weak” category, there are 8.28% fewer students with professional guardians than expected. In contrast, students with guardians who practice manual oc-

Table 8.2. Courses on the school's schedule. (A "✓" means that the course is available, a "X" means that it is not, and a ("✓") means that the course is available only in some schools.)

Course	**5th grade ES**	**6th grade ES**	**1st grade HS**
Ancient Greek Language	X	X	✓
Arts	✓	✓	✓
Biology	X	X	✓
Chemistry	X	X	✓
Computer Science	✓	✓	✓
English Language	✓	✓	✓
French Language	(✓)	(✓)	(✓)
Geography	✓	✓	✓
German Language	(✓)	(✓)	(✓)
Greek Language	✓	✓	✓
Greek Literature	X	X	✓
History	✓	✓	✓
Home Economics	X	X	✓
Italian Language	(✓)	(✓)	(✓)
Mathematics	✓	✓	✓
Music	✓	✓	✓
Physical Education	✓	✓	✓
Physics	✓	✓	✓
Religious Education	(✓)	(✓)	(✓)
Skills Workshops	X	X	✓
Social and Political Education	✓	✓	X
Technology	X	X	✓

cupations are underrepresented in the "very strong" category and over represented in the "very weak" category.

Table 8.3. Guardian occupation and academic achievement.

Guardian occupation	**Group A**	**Group D**
Professionals	+20.72%	-8.28%
Armed forces occupations	+13.69%	-7.56%
Clerical support workers	+10.30%	-5.01%
Managers	+7.23%	-6.97%
Technicians and associate professionals	+6.26%	-4.75%
Service and sales workers	+6.00%	-3.45%
Skilled agricultural, forestry and fishery workers	-4.61%	+1.97%
Plant and machine operators and assemblers	-5.42%	+0.63%
Craft and related trades workers	-6.03%	+0.63%
Elementary occupations	-14.64%	+5.98%

Students from professional guardian groups that are over represented in "very strong" achievement and underrepresented in ”very weak” achievement are classified as ”privileged”. In contrast, students from professional guardian groups who are underrepresented in very strong achievement and over represented in ”very weak” achievement are classified as ”disadvantaged” (Table 8.4).

Table 8.4. Guardian occupation of privileged and of disadvantaged students.

Disadvantaged students	**Privileged students**
Elementary occupations	Professionals
Craft and related trades workers	Armed forces occupations
Plant and machine operators and assemblers	Clerical support workers
Skilled agricultural, forestry and fishery workers	Managers
	Technicians and associate professionals
	Service and sales workers

Because the data we had available did not include variables related to the implementation of the Remedial Teaching policy, we used a ”black box” approach that allows us to evaluate the effectiveness of the policy even in the absence of more specific data. For this purpose, the education system as a whole was examined in the light of the objectives of Remedial Teaching, i.e. to support disadvantaged students. We have therefore examined whether the education system has characteristics of equity in achievement. The analysis was not static but studied the progress of students over a period of three years. Thus, the short- and medium-term effects of remedial teaching on students' performance are studied.

8.2 Assessing the Remedial Teaching (a black box evaluation)

Education systems are characterized by complexity. They are governed by the implementation of many different educational policies. These educational interventions are implemented in parallel, overlapping, or contradictory ways. It is a process that takes many years as has been found many times (Birman, 1981; Maroy, 2009; Braun et al., 2010; Bradley & Migali, 2012; Bellei & Munoz, 2023). In the Greek educational system, along with the policy of remedial teaching, policies are being practiced to make the school more open to other cultures (Spinthourakis & Karakatsanis, 2011; Karountzou et al., 2021), to welcome the children of immigrants (Escaño et al., 2022), to support children with special needs (P. Miller, 2013), to reduce the dropout rates of Roma students (Nikolaou, 2009; Gkofa, 2017; Wallace et al., n.d.), to promote STEAM (Spyropoulou et al., 2020; Chaidi et al., 2021), to develop social skills (Cinque, 2016; Patrinopoulos & Iatrou, 2019) and many others. An objective, evidence-based evaluation of effectiveness and efficiency must take into account a specific policy, taking into account that the observed outcome depends on the simultaneous implementation of several policies.

Education systems are characterized by complexity. They are governed by the implementation of many different educational policies. These educational interventions are implemented in parallel, overlapping, or contradictory ways. It is a process that takes many years as has been found many times (Birman, 1981; Maroy, 2009; Braun et al., 2010; Bradley & Migali, 2012; D. Hall et al., 2015; Bellei & Munoz, 2023). In the Greek educational system, along with the policy of remedial teaching, policies are being practiced to make the school more open to other cultures (Spinthourakis & Karakatsanis, 2011; Karountzou et al., 2021), to welcome the children of immigrants (Escaño et al., 2022), to support children with special needs (K. J. Miller et al., 2013), to reduce the dropout rates of Roma students (Nikolaou, 2009; Gkofa, 2017; Escaño et al., 2022; Wallace et al., n.d.) to promote STEAM (Patrinopoulos & Iatrou, 2019; Spyropoulou et al., 2020), to develop social skills (Cinque, 2016; Demetroulis & Wallace, 2020) and many others. An objective, evidence-based evaluation of effectiveness and efficiency must take into account a specific policy, taking into account that the observed outcome depends on the simultaneous implementation of several policies.

In an ideal research situation, the functioning of the educational system would be examined through a comparison between a ”normal situation” and the situation resulting from the addition of only one of the policies under consideration. This is not possible since: The education system is complex and large and cannot be simulated in operation, so it can only be studied under real conditions. Ethical and legal issues are raised about applying different policies to different segments (experimental groups) of the student population, so it is not possible to compare with the benchmark setting.

Many external factors affect the functioning of the education system, such as changes in society, making observations not directly comparable. An alternative option for evaluating remedial teaching was to use the black box method. We will represent the educational system with a car and reinforce teaching with the car’s braking system. Ideally, we would like to consider the braking system separately from other parts and systems of the car. However, as this is not possible, we look at the whole car (the educational system) from the point of view of the brake system (the support of disadvantaged students and the provision of equal opportunities). If the car’s braking system works well enough (disadvantaged students are helped more), then no further changes are needed. If, on the other hand, the car does not stop in time (disadvantaged students are helped), then the intervention in question is not delivering the expected results in combination with the underlying policies.

8.2.1 Short-term effectiveness of remedial teaching.

The purpose of remedial teaching is to help disadvantaged students with very poor academic achievement, so we focus on students who fall into the very low academic achievement D category.Using the first data set, we examined students who attended fifth grade in elementary school during the 2016–2017 school year to track their progress. This group consists of 3,755 students, or about 5% of the 69,349 students with a known parental occupation in the dataset. In Figure 8.2, disadvantaged students are shown in brown, and advantaged students are shown in blue. To examine the effectiveness of remedial teaching,

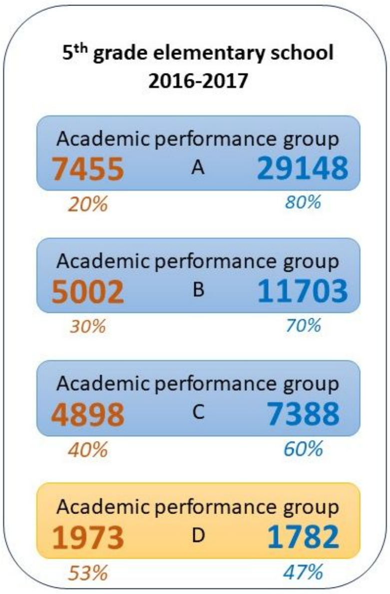


Figure 8.2. Remedial teaching, students analyzed.

I studied the progress of disadvantaged students in the following year. The improvement in the level of one student from group D of very weak academic achievement to a higher level of achievement leads to the conclusion that he or she has received sufficient overall support. Figure 8.3 summarizes this progress.

It is found that 2,631 students (70%) of the very weak students have improved their proficiency level and are now ranked at a level above that of the very weak students. This is a very high rate of mobility and contrasts with previous research that has shown that students rarely move from one academic achievement group to another (Papadogiannis et al., 2021c). This finding is consistent with the view that remedial teaching is a successful progressive intervention for helping low-achieving students. Moreover, the extent to which very weak students are helped exceeds expectations; more than half of very weak students improved their performance.We then examine whether the improvement observed is in line with the objectives of the policy of remedial teaching.

That is, whether disadvantaged students receive this benefit Looking at the occupa-

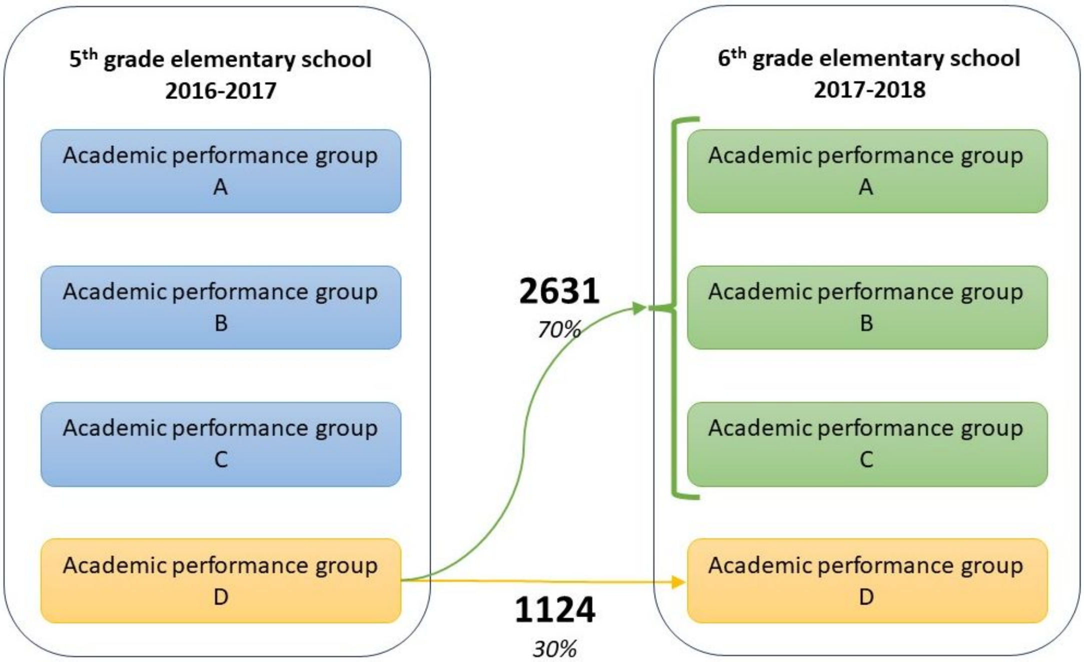


Figure 8.3. Student progression from 5th to 6th grade in ES.

tions of the guardians of the students who improved their performance, we find that both advantaged and disadvantaged students have a comparable improvement (72% vs. 68%) (Figure 8.4). Finally, in the short term (from one grade to the next), remedial teaching is effective in helping students improve their academic achievement. It fails, however, to be effective in achieving its initial goal of supporting disadvantaged students. If anything, privileged students appear to receive slightly more support than disadvantaged students.

8.2.2 Long-term effectiveness of remedial teaching.

The previous section found that remedial teaching (in combination with other interventions) helps a significant proportion of very weak students escape from the lowest academic achievement category. The performance of students in the year after remedial teaching is examined to determine whether they have managed to remain in the highest academic achievement groups or have regressed to a lower level of achievement.

The result is shown in Figure 8.5. The majority of formerly ”very weak” students who had improved their level of academic achievement as a result of remedial teaching reverted to the category of ”very weak” students. One in three students, however, maintained higher levels of academic achievement, which is not an insignificant achievement. Overall, an impressive 879, or 23%, of the 3,755 students who were classified as ”very weak” in grade 5 in elementary school have moved up and maintained higher academic achievement. This includes students who may have remained in Group D in sixth grade in elementary school

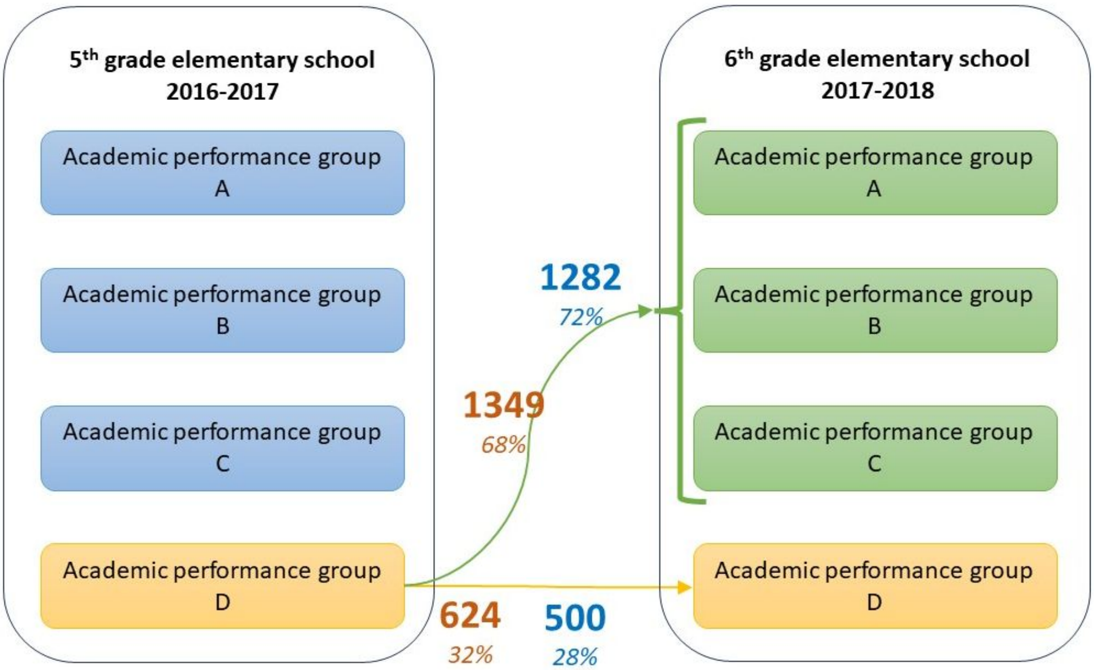


Figure 8.4. Privileged and disadvantaged student progression from 5th to 6th grade in ES.

but performed better in first grade in middle school. Given that students generally remain in the same academic achievement groups as they progress through the school system, the fact that nearly one in four ”very weak” students improves their academic achievement in a way that is sustainable over time is a very impressive positive result that can be attributed to remedial instruction in the absence of other obvious contributing factors.

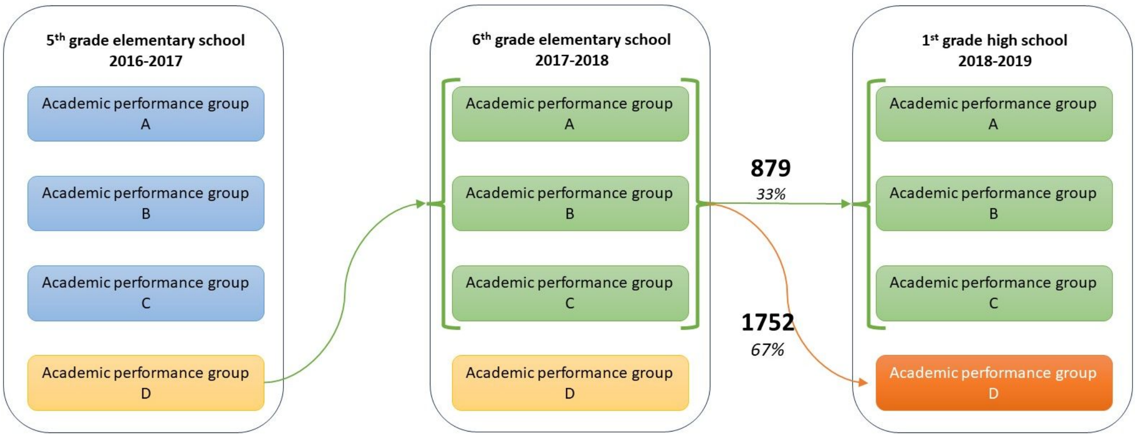


Figure 8.5. Student progress from 6th grade in ES to 1st grade in HS.

The relationship between the categorization of students into privileged or disadvantaged and their ability to maintain higher academic achievement was analyzed to determine whether remedial teaching serves its stated purpose. The goal as created is to help disadvantaged students or rather to provide more opportunities for disadvantaged students.

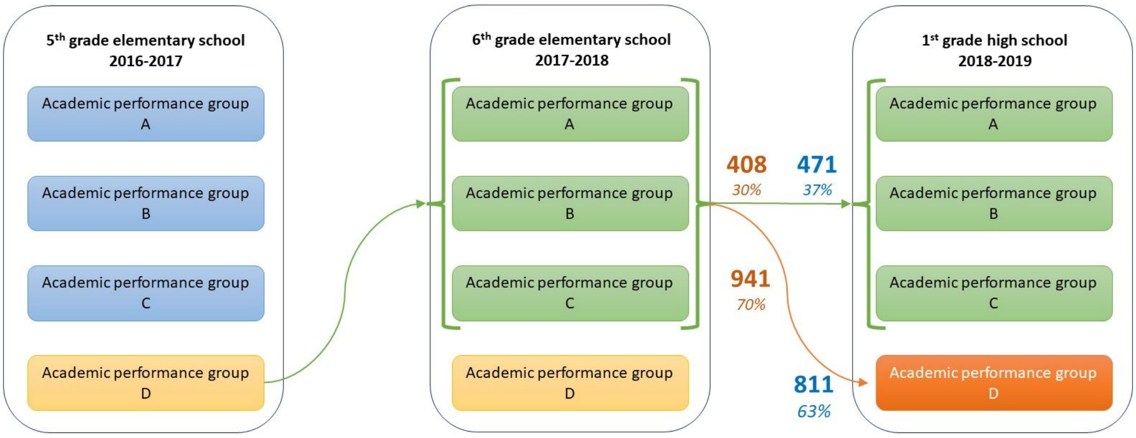


Figure 8.6. Privileged and disadvantaged student progression from 6th grade in ES to 1st grade in HS.

Figure 8.6 compares the second year performance of privileged and disadvantaged students who were originally helped by the intensive teaching policy. Unfortunately, and contrary to the expected focus of remedial teaching, it was found that privileged students are more likely to maintain higher academic achievement than disadvantaged students (37% vs. 30%).

Remedial teaching helps some students improve and maintain their achievement level. Nevertheless, in contrast to prevailing beliefs, it can be argued that the implementation of remedial teaching policies does not effectively contribute to the establishment of a school environment that promotes equal opportunities. If anything, it operates in a contrary manner, serving as an additional facet of the educational system that confers greater advantages to students from privileged backgrounds as opposed to those who are disadvantaged, widening the education gap.

8.3 Findings

Applying a black-box approach and utilizing an extensive data set, we conducted an evaluation analysis of educational policy related to remedial teaching. This educational policy is widely promoted and represents a significant financial commitment by the Greek government. Its primary objective is to promote equality of opportunity within the education system by providing assistance to students facing socio-economic disadvantages. The analysis of the data revealed several findings that challenge the prevailing view on the effectiveness of the policy.

First, in terms of the immediate effects of the policy, it is observed that a remarkable 70% of students classified as disadvantaged are able to make sufficient progress to move into a higher academic achievement category in the following school year. This exceeds expected results. Moreover, a significant proportion of these students, namely one third, show continued improvement in academic achievement even after the end of remedial teaching in secondary school. This serves as a verification of the effectiveness of the policy in terms of significant improvement in students' academic achievement.

Looking at the findings from an equal opportunities perspective, this policy appears to disproportionately favor privileged students, both in the immediate and medium term. In the immediate period, privileged students demonstrate improved academic achievement, while in the medium term they return to the category of disadvantaged students at lower rates than non-privileged students.

In conclusion, while remedial teaching has shown some success in improving achievement level in the short term, particularly for privileged students, the analysis highlights the limitations of this policy in promoting lasting equality of opportunity. A more holistic approach that addresses both academic and socioeconomic challenges may be needed to drive meaningful change.

The use of Management Information System data to objectively evaluate an education policy shows its value. Similarly, more education policies can be evaluated by comparing their goals to their results. The fact that remedial teaching policy results contradict the preexisting, subjective, intuition-based assessment emphasizes the need for objective, data-based education policy evaluation. These findings should help us move toward a paradigm shift where policies are evaluated objectively, based on data, rather than subjectively, based on hope and intuition.

Chapter 9

GPA predictive power

In this thesis, Grade Point Average (GPA) was a key variable. The reason had to do with the theoretical underpinnings of the measure and its properties in measuring and evaluating student achievement (Carrillo-de-la-Peña et al., 2009; Kuh et al., 2011). GPA quantifies a student's overall achievement on a measurement. It is an objective measure that is applied uniformly to all students and allows for direct comparison between students. It indicates the degree of academic success, which indicates a student's proficiency (Brookhart et al., 2016). Due to the centrality of GPA in the research, its predictive power was examined separately in order to test the reliability of the research conclusions.

9.1 Data and analysis

Using both GPA and lessons' grades as inputs, we have found that in primary and secondary school, students' achievement can be classified into four achievement groups. We also found stability in the same achievement group in subsequent school years, as is shown in Figure 9.1. A clear picture of the centroids in HS per grade and achievement level, is shown in Table 9.1. From the study of the deviations, it appears that the four achievement levels are well separated.

In order to evaluate how the GPA in current performance reflects future performance or otherwise assess the predictive power of GPA, we first examined the cluster sizes as originally calculated, comparing them to the sizes of the corresponding levels in subsequent years, as calculated by using GPA.

The sizes of the groups in the three tiers are shown in Table 9.2. It becomes apparent that the relative sizes of the groups do not change substantially in the three grades of high school. Looking at the corresponding percentages in Table 9.3, this becomes even more apparent. We see that about 30% of students have excellent academic achievement, while about 15% have very weak academic achievement, regardless of the grade of the school.

Figure 9.1. Stability in achievement level.

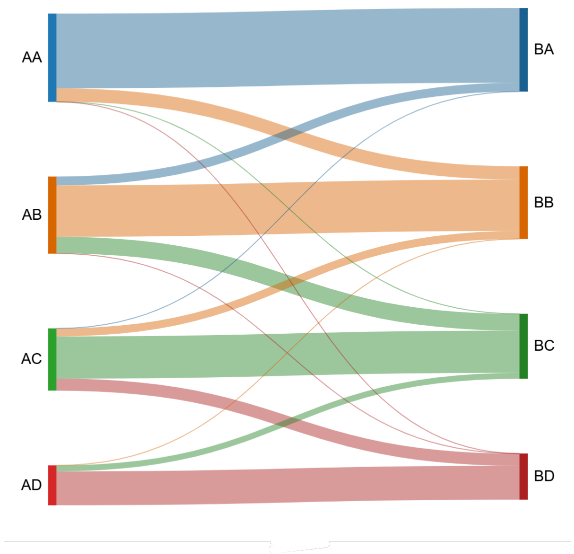


Table 9.1. Cetroinds (CD) per achievement level

Cluster	CD 1st HS	CD 2nd HS	CD 3rd HS
A	19.070/0.038	19.100/0.088	19.076/0.014
B	17.144/0.071	16.934/0.170	16.826/0.062
C	15.205/0.118	14.811/0.120	14.696/0.183
D	13.073/0.256	12.845/0.346	12.962/0.471

Table 9.2. Results of clustering per school year, based on GPA.

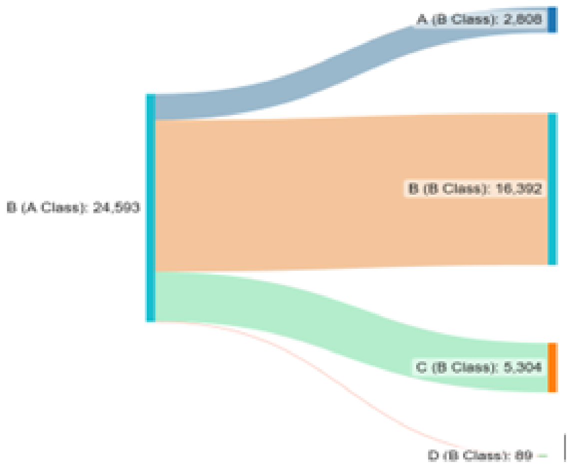
Cluster	1st grade	2nd grade	3rd grade
A	28,153	26,666	26,864
B	24,593	23,189	24,111
C	19,846	20,741	20,865
D	12,752	14,748	13,504

Table 9.3. Results of clustering per school year as %.

Cluster	1st grade	2nd grade	3rd grade
A	33.0%	31,0%	31.0%
B	29.0%	27.0%	28.0%
C	23.0%	24.0%	24.0%
D	15.0%	17.0%	16.0%

We then examine whether the members of the achievement groups remain the same from one grade to the next. Examining the transition from 1st to 2nd grade for students who had strong academic achievement in 1st grade in high school, we observe that, in 2nd grade, 67% maintained their strong achievement, 33% moved to a nearby academic achievement level (either better or worse), and about 0.4% of the latter had drastically different aca-

Figure 9.2. The progress of students with strong academic achievement in the school year 2017-18 as they moved from the 1st to the 2nd grade of HS.



demic achievement, moving to a non-neighboring academic achievement group. A clear visualization of this finding is shown in Figure 9.2.

In Table 9.4 it can be observed that a majority of 75% of students are able to sustain their academic achievement from first to second grade in high school. Significantly, major changes in academic achievement are very uncommon, transpiring in less than 0.5% of cases. Overall, based on these observations, we can conclude that Grade Point Average is a relatively good rough predictor of students' academic progress, both in one year and in the long term. The GPA can be used to predict the group of future academic achievement in the vast majority of cases, and importantly almost never produces seriously inaccurate predictions.

Table 9.4. Transitions between achievement groups from 1st to 2nd grade in HS.

Type of change	**Students**	**%**
No change	64.507	76,0 %
Change	20.837	24,0 %
Big change	236	0,3 %

The classification of academic achievement groups can be unclear, resulting in inaccurate predictions of the next achievement groups. Based on students' subsequent GPA, we examined whether GPA can predict future academic success.The lack of uniformity in the high school student achievement rating system across grades is the first issue to be addressed. In high school, a range of scores from 0 to 20 is used, but as students progress to higher levels of education, high achievement tends to become more difficult. In our data, the average score for the first year of high school was 17.21 and 16.77 in the second year. Therefore, students who perform well in both grades may see their GPA decrease by 0.44 points. The mean absolute deviation between the expected and actual second grade GPA is

0.58, with a standard deviation of 0.47 and a median of 0.47. The predicted and observed numbers are not significantly different.Similarly, the prediction of the third grade GPA showed that the mean difference was 0.7 points, with a standard deviation of 0.61 and a median of 0.52. Again, the predicted GPA values are remarkably close to the actual GPA values. Overall, it can be concluded that GPA is a reliable indicator of future academic achievement.

9.2 Alternatives GPAs

So far, the data show good predictive validity of GPA and a positive correlation between GPA in the first year of high school and subsequent academic achievement. A widespread perception among teachers and students is that some subjects, such as language and mathematics, have greater academic significance than others, such as physical education. Besides, the inclusion of physical education in the calculation of Grade Point Average (GPA) has been a matter of debate (Langendorfer et al., 2001). We further explored the potential advantages of adopting an alternative approach to GPA calculation. We therefore used the relative importance of each course, as derived from various data extraction metrics, as weighting factors. The aim is to more accurately assessing students' abilities and improving the prediction of their future academic achievement. We used the corresponding widget of the Orange data mining platform to calculate the relative weights of several courses. This tool provides a variety of scoring algorithms providing many combinations of weights. There are five different methods for calculating the importance of courses: χ^2, Information Gain, Gain Ratio, ANOVA and Relief F. In Table 9.5 we can see the subjects' weights, per method.

Using these weights, we calculated five different GPAs as weighted averages of their grades in their 1st, 2nd and 3rd grade subjects. We then calculated the percentage of students who either retained or shifted into differential achievement groups during the transition from first to second grade. In Table 9.6 we see that the predictive characteristics of the traditional and the alternatives GPAs are comparable, with the conventional grade point average showing a slightly greater advantage.

The Relief F criterion in the weighted GPA calculations appears to marginally improve the prediction of the next school year. In the mid-term, the conventional GPA outperforms the weighted GPA. In general, it appears that all alternatives have similar performance. Consequently, none of the other approaches we studied produced much better predictions than conventional GPA. Therefore, our analysis does not warrant a change in the GPA calculation technique.

Table 9.5. Courses' weights, as calculated by the Orange platform using different scoring options.

Lesson	chi^2	inf. gain	gain ratio	ANOVA	reliefF
Greek Literature	39,816.80	0.693016	0.350941	52,476.20	0.090410
Greek Language	44,856.70	0.670554	0.336379	48,928.90	0.082805
Ancient Greek Language	48,731.50	0.703587	0.361261	52,657.10	0.089857
Religious Education	35,999.96	0.486662	0.249384	29,996.70	0.072567
History	40,278.40	0.682574	0.344007	55,619.80	0.090181
Mathematics	40,398.50	0.601192	0.304467	44,838.70	0.083461
Home Economics	36,581.80	0.488409	0.250016	30,264.70	0.081459
Computer Science	23,896.10	0.313992	0.160368	15,433.20	0.032100
Technology	28,216.30	0.334923	0.222179	16,125.40	0.030976
Physics	38,961.90	0.545619	0.273215	38,949.40	0.038838
Biology	54,942.50	0.676912	0.339833	41617.60	0.105455
Geography	39,618.00	0.585534	0.294713	42,298.20	0.098573
Chemistry	57506.50	0.690078	0.346513	42860.60	0.107945
Social and Political Education	41787.50	0.582667	0.291426	38393.70	0.092651
Music	24,156.20	0.240835	0.157425	9,186.05	0.043716
Physical Education	4,517.47	0.068420	0.079009	1,072.91	0.003324
Skills Workshops	14,728.50	0.279432	0.187594	10,529.20	0.045207
English Language	27,747.90	0.402294	0.204230	23,218.20	0.068469
Second Foreign Language	33,262.50	0.463334	0.231998	26,867.00	0.043972

Table 9.6. Predictive properties of different alternative GPAs

	2nd grade Students	%	3rd grade Students	%
GPA	64,507	76 %	59,194	69%
GPAχ^2	63,972	75 %	58,751	69 %
GPA$_{\text{inf.gain}}$	63,933	75 %	58,816	69 %
GPA$_{\text{gain ratio}}$	63,677	75 %	58,670	69 %
GPA$_{\text{anova}}$	63,858	75 %	58,420	68 %
GPA$_{\text{Relief F.}}$	64,780	76%	58,737	69 %

9.3 Findings

The predictive power of the student grade point average in relation to levels of academic achievement was studied after it was heavily used in the studies in this thesis. We discovered that Grade Point Average is a reliable predictor of future academic achievement, with only 1% of students performing significantly differently in third grade than in first grade. The first grade GPA predicts future high school GPA very accurately, with an error rate of no more than 0.6 in 20 for the second grade and 0.7 in 20 for the third grade. Finally, after testing a number of alternative approaches to calculating GPA, we discovered that conventional GPA performs as well as or better than any alternative tested. All three findings support the use of traditional GPA to evaluate academic potential in secondary school. The first two findings contend that GPA is a good predictor of future academic achievement, while the third contends that using an alternative form of GPA would be useless.

Our findings indicated that the use of GPA requires no changes in current practice. The findings are theoretically sound because they are based on a quantitative analysis of an entire country's student population. As a result, we no longer have an intuitive sense of GPA's value, but rather a very reliable (due to the sheer size of the dataset) quantitative confirmation of its predictive properties. Furthermore, our findings contradict an intuitive expectation, emphatically disproving the expectation that some alternative weighted average, emphasizing core subjects and less on subjects commonly considered minor, would be more representative of a student's potential.

The current study has been conducted using data exclusively obtained from the Greek education system. Hence, the outcomes of our study may be specific to the context of Greece. Further research is required to ascertain the generalizability of these findings to other nations, particularly those with educational frameworks that differ from Greece's. Examining GPA, as a predictor variable and across different school levels and educational systems, could confirm or raise questions about its predictive potential.

Chapter 10

Discussion

This thesis aimed to highlight the importance and potential of data analysis in the education system, addressing the prevailing subjectivity or lack of support for the views expressed. It examined dimensions of the educational system and specific educational interventions, expressing objective and data-based views. Beginning with a common issue in educational research—the academic achievement of students—it emphasized the potential for objective evaluation and decision support offered by the use of data in centralized educational systems. A key role was played by the concept of equal opportunity education, which ran throughout the thesis. The main conclusions of the thesis are presented next.

10.1 Steady levels of students' achievement

With student achievement as the main tool of study, it was necessary to objectively determine the levels of student achievement in order to then examine whether the school can function as an equal opportunity mechanism both overall and at the level of individual educational policies. Examining the possibility of objectively determining levels of achievement was the subject of the first research question.

Using unsupervised learning four levels of student achievement for primary and secondary schools emerged. The ratings ”very strong”, ”strong”, ”weak”, and ”very weak” result from the performance hierarchy and are stable over time. These ratings are based on the twenty-point linear scale for junior high school and on the ten-point scale for elementary school. The designation of a student as “strong” or ”weak” betrays an assessment that approximates a particular “picture” and is linked to some corresponding latent features.

A small overlap between achievement levels is evident when the lessons' grades are used as explanatory factors. In contrast, in the grading systems, there is a clear differentiation of levels based on GPA, as our study has also shown, using GPA as an explanatory variable. This is expected, as applying the clustering algorithm to all courses rather than the overall grade allows for grouping based on students' performance in subjects rather than the weighted average employed in grading systems. This ambiguity highlights the broadness of the concept of school achievement. On the basis of the initial grouping, a statistically significant difference emerged between the GPA scores of the next classes. Bigger differences have been found in the cases of students rated as ”very strong” or ”very weak”.

This grouping could lead to the development of new educational interventions more adapted to the level of student achievement. The aim of these policies should be to improve the upward mobility of students' achievement. The results of this research showed significantly higher mobility between levels “strong” and “weak”. These are those who are classified as “average” students. Improving this two-way mobility between levels in an upward direction should be an objective of education policy.

A clear picture also emerges with regard to students who fall into the categories of “very strong” and “very weak”. Although this is very positive for the “very strong” category, it raises major questions about the effectiveness of the education system for “very weak” students. Although the particular attributes of each student's profile in these groups are unknown, “weak” students are expected to improve faster than “very weak” students.

To improve the achievement level for students classified as “very weak”, the education system must implement targeted intervention programs. Additional support should be provided to help these students gain fundamental skills and close achievement gaps. Smaller class sizes and one-on-one tutoring could ensure “very weak” students receive individualized attention from teachers (Hawley et al., 1984). Specialized remedial teaching courses designed for their level of understanding are needed. Teachers require extensive training to effectively sup- port diverse learners with differentiated instruction methods (Grissmer et al., 2000). Increased monitoring and evaluation of student progress is also important to assess the effectiveness of interventions early on. Finally, engagement with parents/guardians is critical to develop support at home. With the right mix of academic, social-emotional and family support, more “very weak” students can significantly improve their achievement levels over time (Brigman & Campbell, 2003). But each intervention must be evaluated in terms of its degree of success. The role of data collection and analysis for this purpose is crucial.

10.2 Longitudinal dimension and stability in achievement levels

The second research question examined the longitudinal stability of student achievement over three years, using two datasets. The first dataset, which included data from primary and secondary schools. It became apparent that there were stable levels of achievement, particularly for ”very strong” and ”very weak” students. However, intermediate levels showed greater variation, with trends of falling performance suggesting increased difficulty in high school. The use of inferential statistical tests revealed statistically significant differences in GPA and clusters' centroids, based on initial achievement levels. According to the study most students tend to remain at the same relative level of achievement for several years. This stability indicates stability in the factors that influence achievement. Although stability provides predictability and consistency in learning, it also means that disadvantaged students are more difficult to improve without constant intervention and support.

As student achievement is stable over time, schools are expected to face difficulties in substantially improving the outcomes of students performing well below grade level only through current practices. Targeted, long-term interventions may be needed. From a policy perspective, stability indicates the importance of early intervention strategies. It also suggests that internal and external factors affecting students need to be considered (Karadaug, 2017). Ongoing assessment and personalization using data analysis are crucial for increasing support for students as their needs and environment change over the years. One-size-fits-all approaches are less effective.

Identifying common characteristics between students at each achievement level can provide the basis for different educational approaches. Adding more characteristics to the analysis, which are not available to us, will allow for better categorization of students and more targeted interventions. To initiate such a process requires the recognition of the stability of achievement over time as a challenge. In other words, it is about accepting the current situation as “undesirable” and identifying the need for improvement.

In essence, such an intention is the beginning of a new cycle of educational policy-making, which should follow the familiar steps: (1) agenda setting; (2) policy formulation; (3) decision-making; (4) implementation; (5) evaluation, and the monitoring (6) decision to keep, replace, as well as an ongoing evaluation of the policy cycle using data and analysis. The first and most important step is the identification of the problem. It is up to the political will and the intentions of other actors in education policy to consider whether the stability of student performance over time and the lack of an upward trend in student achievement is a desirable situation. If it is assumed that there is no reason or economic way to improve student performance over time, then our thesis finding confirms that the current situation is also the goal of educational policy.

In summary, the stability in achievement levels over time highlights the need for sustained, multi-tiered, personalized efforts to boost outcomes, particularly for disadvantaged students. It also highlights the important role of non-academic influences in student achievement.

10.3 An equal opportunities school

10.3.1 Guardian's occupation effect

Examining whether the school functions as an equal-opportunities institution, was the aim of third research question. This aim supported by studying the independence of students' achievement from non-academic factors. The available factors in the data sets were, the guardian occupation, the gender and the district of residence. Through the independence tests, we were able to test the established view of the role of education as a vehicle for socio-economic progress.

Prominent researchers and sociologists have expressed their perspectives on how the educational system operates. Some of them are clearly ideologically colored, ranging from liberal approaches that regard education as an individual right to Bourdieu's view (1973) that the education system simply reflects and reproduces social conditions.

Coleman's study (1968) provided the first empirical findings against the established post-war paradigm some years later. One outcome stands out above all others: teaching has little impact on student achievement, while the imbalances in achievement appear to be due to students houses, neighborhoods, and contacts with other children (H. C. Hill, 2017).

In theory, the guardian's job is linked to the family's socio-economic status, but other factors are taken into account, such as the family's income, the other parent's job, the area where the family lives, and so on (Kokkinos, 2019). For this reason, we are particularly careful in our use of the relevant terms. Despite the difficulty of examining the concept of socio-economic status as a whole, useful conclusions emerged in relation to the differentiation of school achievement by the occupation of the guardian. In order to limit the number of occupations, it was preferred to use a generally accepted classification (ILO, 1990).

This thesis contributes to educational research and educational policy by revealing the current situation. The differentiation of school achievement level by occupational category was clear. Students with guardians with manual occupations performed lower than expected, and occupations associated with freelancers or civil servants performed significantly better. Students from different backgrounds perform differently at school. It was therefore found that the school does not function as an equal opportunity school since there is no independence from non-academic, socio-economic status factors.

Occupation, as a component of socioeconomic status (SES), serves as a reflection of educational attainment, income, and social standing (Symeonaki et al., 2012). The educational level achieved by guardians, influenced by their occupation, correlates with their attitude towards education, the level of support they can provide to students, and the potential for informal educational support at home. This lack of resources or knowledge among manual workers becomes a source of inequality of opportunities for students. The widespread prevalence of private tutoring has resulted in multiple social, economic, and educational implications, including increased financial burden, social and spatial inequalities, and student burnout (Tsiplakides, 2018).

Research has also documented a correlation between the professions of guardians and students, particularly for highly skilled parents. These probabilities significantly contribute to the variations observed in student achievement. Professions categorized as ”highly skilled non-manual,” such as freelancers, engineers, doctors, and civil servants, typically require a higher education degree (Symeonaki et al., 2012). These occupations were found to be overrepresented among high-performing students in our study.

In contrast, ”manual” occupations do not directly connected with university degrees but emphasize practical skills . The importance placed on school and learning diminishes for children whose future careers are not reliant on theoretical knowledge (Symeonaki et al., 2012). The lower academic achievement of students with parents in manual or temporary jobs can be attributed to factors such as limited resources, job instability, and lower disposable incomes (Autor, 2014). Consequently, students in these categories face limited opportunities for additional education, a crucial component of secondary education in the country. Our research similarly revealed an overrepresentation of students with guardians in these professions among the low-achieving category.

Several policy recommendations could narrow the educational achievement gap for students from diverse socioeconomic backgrounds: First of all, schools need more government funding. This funding would enable smaller class sizes, tutoring, after school programs, and well-stocked libraries and labs (McGee, 2004). Increased support gives disadvantaged students equal chances to succeed. Career counseling and guidance should be improved, especially for working-class students. By providing detailed information about career education requirements, students can make informed decisions (Sosu & Ellis, 2014). This guidance can help them pursue their dreams regardless of socio-economic background. Working with community groups is also crucial. Collaboration can help students and families access vital social services and support. This collaborative effort addresses student challenges outside of school that may affect their academic achievement. Holistic support helps students overcome these challenges and succeed academically.

Teachers need implicit bias training to address socio-economic biases. By undergoing this specialized training, educators can ensure equal opportunities and support for all students, regardless of socio-economic background. Early childhood education programs in disadvantaged areas must also be expanded. These programs are crucial to student school readiness and academic success (Binning & Browman, 2020). Early intervention and educational support can help disadvantaged students start school on par with their peers. Finally, piloting programs that give families new pc's and internet access can boost remote learning and homeschooling. This initiative aims to close the digital divide and give students from all socio-economic backgrounds the tools they need to succeed academically.

These policies can help close the educational achievement gap between socio-economic groups. Addressing systemic barriers and providing targeted support can create a more equitable and inclusive educational system that empowers all students to succeed. The implementation of educational interventions to mitigate educational gaps is a reality in the country. An example of such an intervention is remedial teaching, whose effectiveness was also examined in this thesis. It seems that achieving an equal opportunity school is an ongoing challenge for our education system as a whole.

10.3.2 Gender effect

The current study found that girls have a higher-than-expected frequency in the category of "very strong" students and a lower frequency in the category of "very weak" students. This is a general finding where girls seem to do better than boys in general.

Our findings align with previous researches (Vandecandelaere et al., 2012; R. B. King & McInerney, 2014b). Bouiri et al. , (2022) found that girls outperformed boys in secondary education. Boys, in particular, tend to develop gender stereotypes and perceive themselves as academically superior in motivation, ability, performance, and self-regulation. However, the effects of gender were found to be small by Zell, Krizan, and Teeter (2015). In his study, Reilly, (2019) found that there are gender differences in reading and writing. Terrier (2016) found that teachers' gender biases significantly affected girls' progress, contributing to a lag in boys' performance. Siddiq and Scherer (2019) revealed a positive and significant effect in favor of girls in ICT, with varying results across studies. Breda and Napp (2019) using PISA competition data, also found that girls outperformed boys. In general, the studies agree with the findings of this thesis, indicating that girls perform better in school. Moreover, not only do boys exhibit lower motivation on (Butler, 2014), but they are also less engaged (Wong et al., 2012) and perform worse in secondary education compared to girls (Voyer & Voyer, 2014).

This thesis contributes to the research field by using complete country-level data. The findings of our research show an overachievement of girls at the country level over a period of three school years at the elementary and middle school levels. In contrast, other studies were limited to samples or competition data, with all that this implies for the generalizability of the findings. Previous research primarily looked at factors influencing scores, but this study strengthened understanding of the baseline reality: that girls in the country tend to achieve at higher levels than boys in primary and secondary grades. It helps contextualize factors within the overarching pattern revealed through population-level analysis.

Overall, through large-scale design and total data, this thesis advanced knowledge about gender disparities, from the level of samples to declarations that can be confidently applied to the country's students and education policies. This statewide perspective provides valuable insights to inform targeted interventions aiming to narrow gender gaps.

10.3.3 Region of residence effect

The effects of the area of residence on student performance did not allow the extraction of specific patterns. This was due to the large geographical area covered by the counties of the country, which constituted the variable provided. Consequently, each instance included many different characteristics that could not be identified. For example, there is a large difference in population density within the districts, which include urban, semi urban, and rural areas. The density of the student population also differs strongly between urban and rural areas, but unfortunately, such a distinction was not possible from the data we received. Mountainousness and insularity are also differentiating factors, but no such differentiation was obtained from our data. We think that there could be bigger differences between how students in cities and rural areas do in school.

Despite the broadness of the geographic area, the research showed the existence of three different performance groups, categorized as ”high”, ”middle”, and ”low”. The group of low performing areas shows overachievement of ”very weak” students and the group of high performing areas shows overachievement of ”very strong” students. A linearity was also found in the rates of overachievement and underachievement in the high and low performing categories. However, when examining the periods between the groups, no particular connection was found between them.

Another finding was the consistent presence of a particular geographical area in the very low-performing category. This is the region of West Attica, which is home to a significant number of minorities, such as Roma students. This is a minority that has been the target of many educational interventions. Even today, the Roma have a hard time fitting in at work and in society because of their unique traits and because of stereotypes and prejudices. In particular, the traits of poverty and social exclusion are particularly pronounced in the Roma population, where they are confronted on a daily basis with political, economic, social, and cultural inequalities. Levels of paid employment for Roma vary across Europe, but are usually significantly lower than those of the majority population. A high level of discrimination against job-seekers is also evident in many cases (Bartlett & Gordon, 2011). Research suggests that Roma employment is characterized by low skills, low pay, and precarious employment combined with limited opportunities for progression (Ahmed & Rogers, 2016).

The educational disadvantage of Roma students is an evident in many countries. In Greece showed that 54.7% of Roma did not go to school at all, 33.4% completed only some classes of primary school, 7% finished primary school, 3.4% attended some classes of secondary education, and 0.5% graduated from secondary education. Frazer and Marlier (2011) found that 54% of parents said their children have never been to school.

In our study, a well-known problem of educational policy has essentially reappeared. The education of Roma students has been a policy issue for decades. Many different efforts have been made at local and national level, with different results. The present study indicates, and through its performance, that the problem remains.

10.3.4 Equal opportunities and non - academic factors effects

For a school to function as an 'equal opportunities' school for all students, it is a prerequisite that the school succeed in providing teaching that is responsive to the internal and external factors that affect students' performance. These factors reflect on performance in different and interdependent ways. If differences in opportunities reflecting non-academic factors were addressed by the education system, then performance would not be affected by them. But what our research has shown is that there is a clear differentiation in performance by guardian occupation and gender. Bottom line, boys who attend secondary school and have parents in manual occupations are in the weakest position. In contrast, students with parents who are doctors, teachers, engineers, or lawyers are in an advantageous position. The lack of independence from demographic factors is a clear indication of the challenges facing education policy in trying to establish an education that functions as an 'equal opportunity' school. At present, the education system functions as a mechanism for reproducing the current SES.

This indicates that the schooling system is not always fulfilling its role in imparting equal opportunities for all students, no matter their own family backgrounds. Rather than assisting in improving social mobility, it appears to be entrenching present inequalities by reproducing the same consequences throughout generations. If a student's potential is predetermined by their mother and father's profession and economic level as opposed to their personal capabilities and efforts, it severely limits the prospects for upward mobility through training.

For the training gadget to in reality be characterized as an equal possibility school and provide social development, it must conquer the effect of demographic elements on performance. Schools want stronger assistance and sources to help level the playing field for college kids of lower SES. This should encompass projects like tutoring, improved early life programs, progressed profession counseling, and making sure all kids have access to educational resources at home. Teachers can also benefit from schooling to examine their personal unconscious biases (Hawley et al., 1984; Grissmer et al., 2000; Brigman & Campbell, 2003). Without concrete coverage measures and funding to deal with the socio-economic boundaries dealing with deprived children, the cycle of reproduced inequality is probably to be maintained. Establishing an equitable education system is vital for building a fairer society with opportunity for all.

10.4 Remedial teaching evaluation

The stability in the achievement levels showed a major weakness for improvement, particularly for "very weak" students. Essentially, students identified as very weak in the first grade show a little room for improvement. Essentially, they cannot improve performance beyond the category of "weak" students. It was also found that several students in the "very weak" category who initially improved their performance category in the following school year regressed to the original "D" performance category (Papadogiannis et al., 2021c). Obviously, performance is dependent on internal factors of students, which set limits on students' ability to do well in school (Tomlinson, 2018). Papadogiannis et. al., (2020) showed that the situation tends to be stable, which means that both internal and external factors that affect performance are stable.

A Greek educational public attempts to help disadvantaged students is remedial teaching. This policy has gained particular weight and importance in recent years as student populations are increasingly characterized by diversity. This policy is widely promoted and represents a significant financial commitment by the Greek government. Its primary objective is to promote equality of opportunity within the education system by providing assistance to students facing socio-economic disadvantages.

Applying a black-box approach and utilizing an extensive data set, we conducted an evaluation analysis of the educational policy of remedial teaching. The analysis of the data revealed several findings that challenge the prevailing view on the effectiveness of the policy.

The short- and medium-term effects of the policy were satisfied. We found that a remarkable 70% of students initially classified as "disadvantaged" are able to make sufficient progress to move into a higher academic achievement category in the following school year. This exceeds expected results. Moreover, a significant proportion of these students, namely one third, show continued improvement in academic achievement even after the end of remedial teaching in secondary school. This serves as a verification of the effectiveness of the policy in terms of significant improvement in students' academic achievement.

But looking at the findings from an equal opportunities perspective, this policy appears to disproportionately favor ”privileged” students, both in the short and medium term. In the short term, ”privileged” students demonstrate improved academic achievement, while in the medium term they return to the category of disadvantaged students at lower rates than non-privileged students.

The use of data to objectively evaluate an education policy shows its value in this case. Similarly, more education policies can be evaluated by comparing their goals to their results. The fact that remedial teaching policy results contradict the preexisting subjective, intuition-based assessment emphasizes the need for objective, data-based education policy evaluation. These findings should help us move toward a paradigm shift where policies are evaluated objectively, based on data, rather than subjectively, based on hope and intuition.

In conclusion, while remedial teaching has shown some success in improving achievement level in the short and medium term, particularly for ”privileged” students. But, the analysis highlights some limitations of this policy in promoting lasting equality of opportunity. A more holistic approach that addresses both academic and socioeconomic challenges may be needed to drive meaningful change.

10.5 GPA and its predictive power

The predictive power of the student Grade Point Average in relation to levels of academic achievement was studied after it was heavily used in the studies in this thesis. We discovered that Grade Point Average is a reliable predictor of future academic achievement, with only 1% of students performing significantly differently in third grade than in first grade. The first grade GPA predicts future Junior high school GPA very accurately, with an error rate of no more than 0.6 in 20 for the second grade and 0.7 in 20 for the third grade. Finally, after testing a number of alternative approaches to calculating GPA, we discovered that conventional GPA performs as well as or better than any alternative tested. All three findings support the use of traditional GPA to evaluate academic potential in secondary school. The first two broad findings contend that GPA is a good predictor of future academic achievement, while the third contends that using an alternative form of GPA would be useless.

Our findings indicated that the use of GPA requires no changes in current practice. The findings are theoretically sound and are based on a quantitative analysis of an entire country's student population. As a result, we no longer have an intuitive sense of GPA's value but rather a very reliable (due to the sheer size of the dataset) quantitative confirmation of its predictive properties. Furthermore, our findings contradict the expectation that some alternative weighted average, emphasizing core subjects and focusing less on subjects commonly considered minor, would be more representative of a student's potential.

10.6 Limitations

In this research, there were two main limitations related to the data. The first has to do with the range of demographic data received. Due to the need to comply with GDPR, the Greek Ministry of Education severely limited the number of characteristics provided. Although we received data for all students in the country, a significant number of attributes that were originally requested were not included in the final dataset. A typical example is the difficulty of geographically identifying the town to which each student belongs. This limited our analysis considerably in relation to the area of residence.

Another limitation related to the time span of the data we received Because the information system had been fully operational for three years at the time of data acquisition, our longitudinal analysis was limited to those three years. Although very important conclusions were drawn in relation to the transition of students from primary to secondary education and the progression of students to secondary education, the use of data on a longer time scale would have allowed more conclusions to be drawn.Finally, due to the lack of data on remedial teaching, the outcome of this educational intervention was studied through a black box approach. These limitations reduce to some extent the possibility of drawing conclusions, but without affecting the conclusions that are relevant to the main objective of the thesis.

10.7 Future research

There is a lot of space for study using educational data. Data analysis and all sub-fields of EDM such as learning and academic analytics, can be used. Some interesting areas for research emerged from this thesis.

A full longitudinal assessment of pupils throughout their school career is now possible, in which they are given a numerical assessment. Such research would more fully examine the changes in student achievement in primary and secondary education. Comparing these analyses between different data sets for each school year would offer new insights into changes in student achievement and the effectiveness of educational interventions.

A very important moment in students' lives is the transition from primary to secondary school. In our research, we found a clear differentiation of performance between the two school levels and a worsening of students' scores in secondary education. The study of the causes of this phenomenon may offer new explanations. Examining possible differences based on the non-academic characteristics of students can provide the basis for new educational interventions. Also, the differentiation of teaching between different educational levels can be examined using learning analytics and data analysis. The results may provide triggers for further study and the development of educational initiatives aimed at a smoother transition from one grade to the next.

We also consider it very important to examine the characteristics of students attending different kinds of schools, such as general and vocational education. Vocational education is offered in the second cycle of secondary education in Greece and other countries. The targeting of the curriculum focuses on the acquisition of practical skills by students, who, after graduation, can be employed professionally in the subject they have studied. The choice of vocational education should ideally be based on the interests and aptitudes of students and should be independent of social characteristics or student performance in secondary school. We believe a study in this field would provide an accurate depiction of these students' characteristics.

The move from secondary to tertiary education is perhaps the most significant milestone in students' scholastic experiences. This is done in Greece and other countries through general examinations that apply to all students. Shadow education is connected with students' attempts to pass extremely difficult entrance exams for higher education. Differential performance based on socioeconomic position is expected because access to shadow education is not free and the resources invested in it are connected to the financial capacities of families. In line with the thesis's dilemma, we believe that it is crucial to assess students' success in these exams in relation to their socioeconomic background. Such a study could lead to significant findings on the educational gap between rich and poor students in Greece.

Chapter 11

Conclusions

The primary objective of this thesis was to objectively evaluate the dimensions of the educational system and specific educational policies, with the aim of ensuring equal opportunities for all students. Using student achievement as the main research tool, it was critical to objectively identify student achievement levels in order to examine whether schools can function effectively as mechanisms of equal opportunity, both in general and in the context of individual educational policies.

Through the use of unsupervised learning, four distinct levels of student achievement were identified for primary and secondary schools. This categorization was derived from the students' achievement hierarchy that has remained constant over time. A clear differentiation between levels based on Grade Point Average was evident, highlighting the integrated nature of this indicator. A statistically significant difference between the centroids of achievement levels and the GPA scores confirms the findings.

A clear pattern also emerged in terms of students classified as 'very strong' and 'very weak', with three-fourths of students initially classified in these categories remaining unchanged over the period examined. While the lack of mobility for "very strong" students is a positive finding, the lack of improvement for "very weak" students raises questions about the effectiveness of the education system to help weak students. The stability of remaining at the same achievement level can be a challenge for the development of tailored educational interventions. To address this issue, targeted interventions for the 'very weak' students are needed.

Policy suggestions include shorter lessons, one-to-one tutoring, specialized remedial lessons, training of teachers in differentiated practice, monitoring of student progress, and parental involvement. The key aim is for 'very weak' students to raise achievement levels over the years through a combination of teaching, socio-emotional and family support. The ultimate goal of such policies could be to promote upward mobility in academic achievement.

This policy-making should follow an established policy pathway that includes agenda setting, policy formulation, decision-making, implementation, evaluation, monitoring, and decisions on retention or replacement, along with ongoing evaluation of the policy cycle using data and analysis.

As education has been seen as a means of socio-economic progress, with school performance playing an important role in the acquisition of knowledge, it has been examined whether schools function as institutions of equal opportunity. The examination was conducted through the independence of student achievement from non-academic factors. Factors such as guardian occupation, gender, and region of residence were examined.

Various scholars, mainly from the fields of politics and sociology, have expressed their views on the functioning of the education system. In this study, the connection between guardians' occupations and students achievements has been well documented, with highly skilled parents being more likely to have their children follow a similar path. Highly skilled, non-manual occupations, such as professionals, engineers, doctors, and civil servants, are associated with higher education degrees and tend to have children with ”very strong” achievement. On the other hand, manual occupations are not directly linked to university degrees but rather to the acquisition of practical skills, leading to children with lower achievement.

In order for a school to truly function as an institution of ”equal opportunities” for all students, it is necessary to be able to provide responsive teaching. This teaching takes into account both internal and external factors that affect students academic achievement. The dependence on SES factors highlights the challenges facing education policy in creating such a system. Data analysis on a deeper level can be valuable in order to study the specific challenges.

The findings of the current study also show that girls have a higher than expected frequency in the category of 'very strong' students and a lower frequency in the category of 'very weak' students. This is in line with numerous previous studies suggesting that girls tend to outperform boys in school settings. The thesis confirms the findings and contributes to the research field by using total country-level data and highlighting the overachievement of girls in five school years at primary and secondary school. In contrast, the data used in other studies are usually derived from samples and/or data from international competitions, which are not characterized by the application of sampling rules or the examination of all subjects.

The study also showed that it is not possible to draw specific patterns in the performance of students based on their region of residence at the region level (nuts3). Various characteristics, such as population density within districts, mountainous terrain, and insularity, cannot be specifically identified. However, two main conclusions have been drawn. The regions can be grouped into three achievement groups, categorized as ”high”, ”middle”, and ”low”. Furthermore, linearity was found in the rates of overachievement and underachievement in the high and low performing regions.

Also, West Attica shows a consistent presence in the very low-performing category for primary education. West Attica is home to a significant number of minorities, including Roma students. These minorities face unique challenges and stereotypes, such as poverty and social exclusion. Across Europe, Roma employment is typically characterized by low skills, low wages, precariousness, and limited opportunities for advancement. All the research shows that the social and economic status of Roma has a major impact on the performance of their children. The current research has simply confirmed this conclusion.

In order to address these inequalities, the implementation of educational interventions has become a reality in our country. Remedial teaching is as an example of such an intervention, with the aim of creating a school of equal opportunities. However, the achievement of this goal remains an ongoing challenge for the entire education system.

The thesis also focused on the effectiveness of remedial teaching, an educational policy to promote equality of opportunity in the education system. The short- and medium-term impact of the policy has been significant, with a large proportion of students making significant progress and moving into higher academic achievement levels in the following school year. Also, a significant proportion of these students showed improvement even after completing remedial teaching in secondary education. In contrast, when looking at improvement or handicap from an equal opportunity perspective, it is found that improvement benefits 'privileged' students with professional parents the most, and decline is borne more by 'disadvantaged' students with parents in manual occupations.

The results reveal that remedial teaching has limitations in promoting long-term equality of opportunity. Therefore, a more integrated approach that addresses both academic and socio-economic challenges may be necessary to promote meaningful and sustainable change. The use of data from EMIS to evaluate education policies can prove valuable in comparing their objectives with their actual outcomes.

Finally, the predictive power of Grade Point Average in forecasting future levels of academic achievement was studied. Using the original categorization as a basis for comparison, the results indicate that GPA is a reliable predictor of future academic achievement, with a small percentage of students showing significant differences in achievement levels between 2nd and 3rd grade in junior high school. After examining several alternative methods to calculating GPA, it was found that the conventional method of calculating GPA performed as well as or better than the alternatives tested. These results show that GPA is a good way to measure academic potential in high school. They also prove that the results and conclusions of this thesis, which also used GPA as an explanatory variable, are correct.

In conclusion, we believe that the aim of this thesis, which was to objectively evaluate the dimensions of the education system and policies to ensure equal opportunities for all students, has been achieved. Using educational data, this thesis sheds light on the critical role of education in promoting equality, both in a general sense and in the context of individual educational policies. The findings of this thesis contribute to the ongoing pursuit of a truly inclusive and equitable educational landscape.

References

Lasswell, H. D. (1956). *The decision process: Seven categories of functional analysis*. Bureau of Governmental Research, College of Business; Public∼…

Bourdieu, P. (1966). Lecole conservatrice: Les inegalites devant l'ecole et devant la culture. *Revue franaise de sociologie*, 325–347.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Ffth Berkeley symposium on mathematical statistics and probability*, 281–297. http://projecteuclid.org/euclid.bsmsp/1200512992

Coleman, J. S. (1968). Equality of educational opportunity. *Integrated education*, *6*(5), 19–28.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, *3*(1), 16–20. https://doi.org/10.1007/BF02322211

Bourdieu, P. (1973). Cultural reproduction and social reproduction. In *Knowledge, education, and cultural change* (pp. 71–112). Routledge.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Anderson, B. E., & Wallace, P. A. (1975). Public policy and Black economic Progress: A review of the evidence. *The American Economic Review*, *65*(2), 47–52.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of educational research*, *46*(3), 407–441.

Jenkins, W. I. (1978). *Policy analysis: A political and organisational perspective*. Wiley-Blackwell.

Schwarz, G. (1978). Estimating the dimension of a model Annals of Statistics, 6, 461-464. *MathSciNet zbMATH*.

Wildavsky, A. (1978). Policy analysis is what information systems are not. *Accounting, Organizations and Society*, *3*(1), 77–88. https://doi.org/10.1016/0361-3682(78)90008-9

Bronfenbrenner, U. (1979). Contexts of child rearing: Problems and prospects. *American psychologist*, *34*(10), 844.

Birman, B. F. (1981). Problems of Overlap Between Title I and P. L. 94–142: Implications for the Federal Role in Education. *Educational Evaluation and Policy Analysis*, *3*(3), 5–19. https://doi.org/10.3102/01623737003003005

Brewer, G. D., & DeLeon, P. (1983). *The foundations of policy analysis*. Dorsey Press.

Hawley, W. D., Rosenholtz, S., Goodstein, H. J., & Hasselbring, T. (1984). Good schools: What research says about improving student achievement. *Peabody Journal of Education*, *61*(4), iii–178.

Deci, E. L., & Ryan, R. (1985). Intrinsic motivation and self-determination in human behavior. berlin: Springer science & business media.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. https://doi.org/10.1037/0003-066X.41.10.1040

Keller, K. L., & Staelin, R. (1987). Effects of quality and quantity of information on decision effectiveness. *Journal of consumer research*, *14*(2), 200–213.

ILO. (1990). International standard classification of occupations: ISCO-88.

Banathy, B. (1991). Systems design of education: A journey to create the future.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, *84*(3), 261.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Machine learning proceedings 1992* (pp. 249–256). Elsevier.

Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. *American journal of public health*, *82*(6), 816–820.

Durm, M. W. (1993). An a is not an a is not an a: A history of grading. *The Educational Forum*, *57*(3), 294–297.

Greek National Printing House. (1995). Presidential decree 3/1995. http://grafis.sch.gr/public.php?service=files%20t=a52dde4a308910537aabab76857f2dc7

Christakis, K. (1996). Modern views on the education of people with special Needs. In: Proceedings in a training seminar for School Advisors of all Grades. *Training seminar for School Advisors of all Grades*, 528–538.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, *1*(1-4), 131–156.

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in us public health research: Concepts, methodologies, and guidelines. *Annual review of public health*, *18*(1), 341–378.

Currie, J., & Thomas, D. (1999). Early test scores, socioeconomic status and future outcomes.

Kogan, M. (1999). The impact of research on policy. *Speaking truth to power: Research and policy on lifelong learning*, *3*, 11.

Anderson, L. W., Jacobs, J., Schramm, S., & Splittgerber, F. (2000). School transitions: Beginning of the end or a new beginning? *International journal of educational research*, *33*(4), 325–339.

Freeman, S. F. N., & Alkin, M. C. (2000). Academic and social attainments of children with mental retardation in general education and special education settings. *Remedial and Special Education*, *21*(1), 3–26.

Grissmer, D. W., Flanagan, A., Kawata, J. H., Williamson, S., & LaTourrette, T. (2000). *Improving student achievement: What state naep test scores tell us*. Rand Corporation.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American economic review*, *90*(5), 1184–1208.

Heckman, J. J. (2000). Policies to foster human capital. *Research in economics*, *54*(1), 3–56.

Kochhar-Bryant, C. A., West, L. L., Taymans, J. M., & Kochhar-Bryant, C. A. (2000). *Successful Inclusion: Practical Strategies for a Shared Responsibility Subsequent Edition*. Prentice Hall.

Pelleg, D., & Moore, A. (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *In Proceedings of the 17th International Conf. on Machine Learning*, 727–734.

Wößmann. (2000). *Schooling Resources, Educational Institutions, and Student Performance: The International Evidence*, Kiel Institute of World Economics,

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13–39). Elsevier.

Crouch, L., Enache, M., & Supanc, P. (2001). Education Management Information Systems (EMIS): Guidelines for Design and Implementation. *Techknowlogia: International Journal of Technology for the Advancement of Knowledge and Learning*, *3*(1), 46–49.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification (2nd ed.).* John Wiley; Sons.

Laney, D. (2001, February). *3D Data Management: Controlling Data Volume, Velocity, and Variety* (tech. rep.). http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Langendorfer, S. J., Crawford, S. A., Nickels, J., & Mathews, A. (2001). Should the physical education grade be included in a high school student's gpa? *Journal of Physical Education, Recreation and Dance*, *72*(8), 9.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public administration*, *80*(1), 1–22.

Slavin, R. E. (2002). ”evidence-based education policies: Transforming educational practice and research.” *Educational researcher*, *31*(7), 15–21.

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. *Development of achievement motivation*, 91–120.

Brigman, G., & Campbell, C. (2003). Helping students improve academic achievement and school success behavior. *Professional school counseling*, 91–98.

Dimitrov, D., & Rumrill, P. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159–165. https://doi.org/10.3233/WOR-2003-1602

Everett, S. (2003). The policy cycle: democratic process or rational paradigm revisited? *Australian Journal of Public Administration*, *62*(2), 65–70.

Lehmbruch, G. (2003). Concertation and the structure of corporatist networks. In *Verhandlungsdemokratie* (pp. 103–128). Springer.

Linn, M. (2003). Technology and science education: Starting points, research programs, and trends. *International journal of science education*, *25*(6), 727–758.

Makwati, G., Audinos, B., & Lairez, T. (2003). The role of statistics in improving the quality of basic education in Sub-Saharan Africa. *ADEA Biennial Meeting*, 3–6.

Sirsch, U. (2003). The impending transition from primary to secondary school: Challenge or threat? *International Journal of Behavioral Development*, *27*(5), 385–395.

Wößmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford bulletin of economics and statistics*, *65*(2), 117–170.

Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. *Proceedings of the 20th international conference on machine learning (ICML-03)*, 856–863.

Zeedyk, M. S., Gallacher, J., Henderson, M., Hope, G., Husband, B., & Lindsay, K. (2003). Negotiating the transition from primary to secondary school: Perceptions of pupils, parents and teachers. *School Psychology International*, *24*(1), 67–79.

Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post 16 learning: A systematic and critical review*. Learning; Skills Research Centre.

Keefe, E. B., & Moore, V. (2004). The challenge of co-teaching in inclusive classrooms at the high school level: What the teachers told us. *American secondary education*, 77–88.

Lawrence-Brown, D. (2004). Differentiated instruction: Inclusive strategies for standards-based learning that benefit the whole class. *American secondary education*, 34–62.

McGee, G. W. (2004). Closing the achievement gap: Lessons from illinois' golden spike high-poverty high-performing schools. *Journal of education for students placed at risk*, *9*(2), 97–125.

OECD. (2004, December). *Pisa learning for tomorrow's world first results from pisa 2003*.

Pai, M., McCulloch, M., Gorman, J. D., Pai, N. P., Enanoria, W. T. A., Kennedy, G. E., Tharyan, P., & Colford, J. M. (2004). Systematic reviews and meta-analyses: an illustrated, step-by-step guide. *The National medical journal of India*, *17 2*, 86–95.

Cole, M., Cole, S. R., & Lightfoot, C. (2005). *The development of children*. Macmillan.

Colebatch, H. K. (2005). Policy analysis, policy practice and political science. *Australian Journal of Public Administration*, *64*(3), 14–23.

Crombie, G., Sinclair, N., Silverthorn, N., Byrne, B. M., & ... (2005). Predictors of young adolescents' math grades and course enrollment intentions: Gender similarities and differences. *Sex Roles*. https://doi.org/10.1007/s11199-005-2678-1

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of applied developmental Psychology*, *26*(3), 276–295.

John, E. P. S., & Parsons, M. D. (2005). *Public funding of higher education: Changing contexts and new rationales*. JHU Press.

Parsons Wayne. (2005). *Public Policy : An Introduction to the Theory and Practice of Policy Analysis*. E. Elgar.

Booth, T., Ainscow, M., & Kingston, D. (2006). *Index for Inclusion: Developing Play, Learning and Participation in Early Years and Childcare.* ERIC.

Cohen-Schotanus, J., Muijtjens, A. M. M., Reinders, J. J., Agsteribbe, J., Van Rossum, H. J. M., & Van Der Vleuten, C. P. M. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Medical Education*, *40*(10), 1012–1019.

Dunleavy, P., Margetts, H., Tinkler, J., Bastow, S., et al. (2006). *Digital era governance: It corporations, the state, and e-government*. Oxford University Press.

Feng, Y., & Hamerly, G. (2006). Pg-means: Learning the number of clusters in data. *Advances in neural information processing systems*, *19*.

Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan*, *87*(9), 670–675. https://doi.org/10.1177/003172170608700910

Abbott, L. (2007). Northern Ireland Special Educational Needs Coordinators creating inclusive environments: an epic struggle. *European journal of special needs education*, *22*(4), 391–407.

Álvarez, J., Moreno, V. G., & Patrinos, H. A. (2007). *Institutional effects as determinants of learning outcomes: Exploring state variations in Mexico* (Vol. 4286). World Bank Publications.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE review*, *42*(4), 40.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., et al. (2007). School readiness and later achievement. *Developmental psychology*, *43*(6), 1428.

Hanushek, E. A., & Woessmann, L. (2007). Education Quality and Economic Growth Washington. *DC: The World Bank*.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Scott, L., Ingels, S., & Owings, J. (2007). Interpreting 12th-graders' naep-scaled mathematics performance using high school predictors and postsecondary outcomes from the national education longitudinal study of 1988 (nels: 88). *Statistical Analysis Report. National Center for Education Statistics*.

Scruggs, T. E., Mastropieri, M. A., & McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A metasynthesis of qualitative research. *Exceptional children*, *73*(4), 392–416.

Greek_Parliament, T. (2008). *Constitution of Greece State.* The Greek Parliament.

Koutselini, M. (2008). Listening to students' voices for teaching in mixed ability classrooms: Presuppositions and considerations for differentiated instruction. *Learning and teaching*, *1*(1), 17–30.

Meelissen, M., & Luyten, H. (2008). The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. *Studies in Educational Evaluation*, *34*(2), 82–93.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American educational research journal*, *45*(2), 365–397.

Alcala-Fdez, J., Sanchez, L., Garcia, S., del Jesus, M. J., Ventura, S., Garrell, J. M., Otero, J., Romero, C., Bacardit, J., Rivas, V. M., et al. (2009). Keel: A software tool to assess evolutionary algorithms for data mining problems. *Soft Computing*, *13*, 307–318.

Anaya, A. R., & Boticario, J. G. (2009). Clustering learners according to their collaboration. *2009 13th International Conference on Computer Supported Cooperative Work in Design*, 540–545.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM| Journal of Educational Data Mining*, *1*(1).

Carrillo-de-la-Peña, M., Baillès, E., Caseras, X., Martínez, À., Ortet, G., & Pérez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education*, *14*(1), 61–67.

Holsinger, D. B., Jacob, W. J., & Mason, M. (Eds.). (2009). *Inequality in education: Comparative and international perspectives*. Springer Netherlands.

Maroy, C. (2009). Convergences and hybridization of educational policies around "post-bureaucratic" models of regulation. *Compare*, *39*(1), 71–84.

Moore, C., & Shulock, N. (2009). Student progress toward degree completion: Lessons from the research literature. *Institute for Higher Education Leadership Policy*, 8–10.

Nikolaou, G. (2009). Teacher training on Roma education in Greece: a discussion about the results of INSETRom experience in two Greek schools. *Intercultural Education*, *20*(6), 549–557. https://doi.org/10.1080/14675980903448585

Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. *Handbook of motivation at school*, *35*, 54.

Andrew-Kaliandis, S. (2010). *Differentiation of teaching / learning processes and results in mixed capacity classes* [Doctoral dissertation, University of Cyprus, School of Social Sciences, Education Sciences / University of Cyprus, Faculty of Social Sciences, and Education].

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching & Learning*, *4*(2).

Bijker, W. E. (2010). How is technology made?—That is the question! *Cambridge journal of economics*, *34*(1), 63–76.

Braun, A., Maguire, M., & Ball, S. J. (2010). Policy enactments in the UK secondary school: examining policy, practice and school positioning. *Journal of Education Policy*, *25*(4), 547–560. https://doi.org/1080/02680931003698544

Chang, W.-C., Wang, T.-H., & Li, M.-F. (2010). Learning ability clustering in collaborative learning. *J. Softw.*, *5*(12), 1363–1370.

Cobo, G., García-Solórzano, D., Santamaría, E., Morán, J. A., Melenchón, J., & Monzo, C. (2010). Modeling students' activity in online discussion forums: A strategy based on time series and agglomerative hierarchical clustering. *Educational data mining 2011*.

Friend, M., Cook, L., Hurley-Chamberlain, D., & Shamberger, C. (2010). Co-teaching: An illustration of the complexity of collaboration in special education. *Journal of educational and psychological consultation*, *20*(1), 9–27.

Halverson, R. (2010). School Formative Feedback Systems. *Peabody Journal of Education*, *85*(2), 130–146. https://doi.org/10.1080/01619561003685270

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2010). Systematic literature reviews in software engineering–a tertiary study. *Information and software technology*, *52*(6).

McQuarrie, L. M., & McRae, P. (2010). A provincial perspective on differentiated instruction: The Alberta Initiative for School Improvement (AISI). *Journal of applied research on learning*, *3*(4), 1–18.

Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Schildkamp, K., & Kuiper, W. (2010). Data-driven curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*, 482–496. https://doi.org/https://doi.org/10.1016/j.tate.2009.06.007

Zhao, J., & Wang, T. (2010). A general framework for medical data mining. *2010 International Conference on Future Information Technology and Management Engineering*, *2*, 163–165.

Badie, B., Berg-Schlosser, D., & Morlino, L. (2011). *International encyclopedia of political science* (Vol. 4). Sage.

Bartlett, W., & Gordon, C. (2011). Measures to promote the situation of Roma EU citizens in the European Union. *EPRS: European Parliamentary Research Service*.

Bruns, B., Filmer, D., & Patrinos, H. A. (2011). *Making schools work: New evidence on accountability reforms*. World Bank Publications.

Frazer, H., & Marlier, E. (2011). Promoting the social inclusion of Roma. *CEPS/INSTEAD*.

Grissom, J. A., & Loeb, S. (2011). Triangulating principal effectiveness: How perspectives of parents, teachers, and assistant principals identify the central importance of managerial skills. *American Educational Research Journal*, *48*(5), 1091–1123.

Kuh, G., Kinzie, J., Schuh, J., & Whitt, E. (2011). *Student success in college: Creating conditions that matter*. John Wiley; Sons.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Hung Byers, A., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

Mattern, K., & Patterson, B. (2011). Validity of the sat for predicting second-year grades: 2006 sat validity sample. *The College Board*, 1–30.

Spinthourakis, J., & Karakatsanis, D. (2011). The Multicultural Class in Greece. In J. Spinthourakis, J. Lalor, & W. Berg (Eds.), *Cultural diversity in the classroom*. VS Verlag für Sozialwissenschaften.

Tresch, A., Sciarini, P., & Varone, F. (2011). A policy cycle perspective on the media's political agenda-setting power. *6th ECPR General Conference, University of Iceland*, 25–27.

Tzanakis, M. (2011). Bourdieu's social reproduction thesis and the role of cultural capital in educational attainment: A critical review of key empirical studies. *Educate*, *11*(1), 76–90.

Valiande, S., Kyriakides, L., & Koutselini, M. (2011). Investigating the Impact of Differentiated Instruction in Mixed Ability Classrooms: It's impact on the Quality and Equity Dimensions of Education Effectiveness Valiande A. Stavroula Cyprus Pedagogical Institute. *Paper presented at the*, *2011*.

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*, 383–398.

Bradley, S., & Migali, G. (2012). An Evaluation of Multiple Overlapping Policies in a Multi-Level Framework: The Case of Secondary Education Policies in the UK. *Education Economics*, *20*(3), 322–342.

Eranki, K. L., & Moudgalya, K. M. (2012). Evaluation of web based behavioral interventions using spoken tutorials. *2012 IEEE Fourth International Conference on Technology for Education*, 38–45.

Gayles, J. (2012). Race, late bloomers and first-year gpa: Predicting beyond the freshman year. *Educational Research Quarterly*, *36*(1), 13–29.

Gobert, J. D., Sao Pedro, M. A., Baker, R. S., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, *4*(1), 104–143.

Han, J., & Kamber, M. (2012). *Data mining: concepts and techniques*. Elsevier.

Heward, W. L. (2012). *Exceptional children: an introduction to special education*. Pearson.

ILO. (2012). *International Standard Classification of Occupations 2008: ISCO-08*. International Labour Office.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71–85.

Pinar, W. F. (2012). *What is curriculum theory?* Routledge.

Symeonaki, M., Stamatopoulou, G., & Michalopoulou, C. (2012). Intergenerational occupational mobility in Greece: evidence from EU-SILC. *Demographic Analysis and Research International Conference, Chania, Greece*, 5–8.

Vandecandelaere, M., Speybroeck, S., Vanlaar, G., & ... (2012). Learning environment and students' mathematics attitude. *Studies in Educational* ... https://www.sciencedirect.com/science/article/pii/S0191491X12000417

Waters, S. K., Lester, L., Wenden, E., & Cross, D. (2012). A theoretically grounded exploration of the social and emotional outcomes of transition to secondary school. *Journal of Psychologists and Counsellors in Schools*, *22*(2), 190–205.

Webb, P. T., & Gulson, K. N. (2012). Policy prolepsis in education: Encounters, becomings, and phantasms. *Discourse: studies in the cultural politics of education*, *33*(1), 87–99.

Wong, S. W. L., McBride-Chang, C., Lam, C., Chan, B., Lam, F. W. F., & Doo, S. (2012). The joint effects of risk status, gender, early literacy and cognitive skills on the presence of dyslexia among a group of high-risk Chinese children. *Dyslexia*, *18*(1), 40–57.

Anagnostopoulos, D., Rutledge, S., & Jacobsen, R. (2013). Mapping the information infrastructure of accountability. *The infrastructure of accountability: Data use and the transformation of American education*, 1–20.

Anastasiou, D., & Mavropalias, T. (2013). Teachers' perceptions of the Greek co-teaching model. In *The bridge from segregation to inclusion ...a long journey* (pp. 107–108). https://doi.org/10.13140/2.1.2611.8081

Aturupane, H., Glewwe, P., & Wisniewski, S. (2013). The impact of school quality, socioeconomic factors, and child health on students' academic performance: evidence from Sri Lankan primary schools. *Education Economics*, *21*(1), 2–37.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2013). School leaders matter. *Education Next*, *13*(1), 62–69.

Henig, J. R. (2013). *The infrastructure of accountability: Data use and the transformation of American education*. Harvard Education Press.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American political science Review*, *107*(2), 326–343.

Lingard, B., Martino, W., & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: Commensurate global and national developments.

Margetts, H., & Dunleavy, P. (2013). The second wave of digital-era governance: a quasi-paradigm for government on the Web. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1987), 20120382.

Martino, W., & Rezai-Rashti, G. (2013). 'Gap talk'and the global rescaling of educational accountability in Canada. *Journal of Education Policy*, *28*(5), 589–611.

McCombs, M. E., Shaw, D. L., & Weaver, D. H. (2013). *Communication and democracy: Exploring the intellectual frontiers in agenda-setting theory*. Routledge.

Miller, K. J., Morfidi, E., & Soulis, S. (2013). Teachers' perceptions of greek special education policies and practices. *Journal of International Special Needs Education*, *16*(1), 53–65.

Miller, P. (2013). School leadership in the caribbean: Perceptions, practices, paradigms.

Ming-Hsueh Tsai, F. L. (2013). Multigroup structural equation approach: Examing the relationship among family socioeconomic status, parent-child interaction, and academic achievement using TASA samples. *International Journal of Intelligent Technologies and Applied Statistics*, *6*, 353–373. https://doi.org/10.6148/IJITAS.2013.0604.03

Moutavelis, A. (2013). Comparative study and evaluation on building Department of Integration Programs. *3rd Panhellenic Special Education Conference with international participation. Dilemmas and Prospects in Special Education*, 13–27.

Park, S., & Holloway, S. D. (2013). No parent left behind: Predicting parental involvement in adolescents' education within a sociodemographically diverse population. *The Journal of Educational Research*, *106*(2), 105–119.

Patrinos, H. A., & Psacharopoulos, G. (2013). Education: the income and equity loss of not having a faster rate of human capital accumulation. *How Much Have Global Problems Cost the World*.

Sanchez, E. (2013). *Differential effects of using ACT® college readiness assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups* (tech. rep. No. 4). ACT Research Report Series.

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. Remedial and Special Education. *Remedial and Special Education*, *34*(1), 9–19. https://doi.org/10.1177/0741932512440730

Thomas, M., & Burnett, N. (2013). Exclusion from education: the economic cost of out of school children in 20 countries. *Washington, DC: Results for Development*.

Wang, X. (2013). Why students choose stem majors: Motivation, high school learning, and postsecondary context of support. *American Educational Research Journal*, *50*(5), 1081–1121.

Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. *arXiv*.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, *344*(6186), 843–851.

Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In *Learning analytics* (pp. 61–75, Vol. 10.). Springer New York. https://doi.org/10.1007/978-1-4614-3305-7_4

Butler, R. (2014). Motivation in educational contexts: Does gender matter? *Advances in child development and behavior*, *47*, 1–41.

Camargo, B., Camelo, R., Firpo, S., & Ponczek, V. P. (2014). Information, market incentives, and student performance. *Market Incentives, and Student Performance*.

Darolia, R., Potochnick, S., & Menifield, C. E. (2014). Assessing admission criteria for early and mid-career students: Evidence from a us mpa program. *Education Policy Analysis Archives*, *22*, 101–101.

Donné, N. L. (2014). European Variations in Socioeconomic Inequalities in Students' Cognitive Achievement: The Role of Educational Policies. *European Sociological Review*, *30*(3), 329–343. https://doi.org/10.1093/esr/jcu040

for Fundamental Rights, E. U. A., de los Derechos Fundamentales de la Unión Europea, A., für Grundrechte, E. U. A., & for Fundamental Rights, F.-E. U. A. (2014). *Violence against women: An eu-wide survey: Main results*. FRA, European Union Agency for Fundamental Rights.

Gargani, J., & Strong, M. (2014). Can we identify a successful teacher better, faster, and cheaper? Evidence for innovating teacher observation systems. *Journal of Teacher Education*, *65*(5), 389–401.

Hill, P., & Barber, M. (2014). *Preparing for a renaissance in assessment*. Pearson London.

Jayaprakash, S. M., Moody, E. W., Lauria, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47.

King, R. B., & McInerney, D. M. (2014a). Mapping changes in students' English and math self-concepts: A latent growth model study. *Educational Psychology*. https://doi.org/10.1080/01443410.2014.909009

King, R. B., & McInerney, D. M. (2014b). Culture's consequences on student motivation: Capturing cross-cultural universality and variability through personal investment theory. *Educational psychologist*, *49*(3), 175–198.

Knowles, J. (2014). Of Needles and Haystacks: Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *JEDM - Journal of Educational Data Mining*, *7*(3), 1–52. https://doi.org/doi:10.6084/m9.figshare.1142580

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203–1205.

Papamitsiou, Z., & Economides, A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology and Society*, *17*(4), 49–64.

Pardo, M., & Woodrow, C. (2014). Improving the quality of early childhood education in Chile: Tensions between public policy and teacher discourses over the schoolarisation of early childhood education. *International Journal of Early Childhood*, *46*(1), 101–115.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, *41*(4), 1432–1462.

Prieto, L. P., Wen, Y., Caballero, D., Sharma, K., & Dillenbourg, P. (2014). Studying teacher cognitive load in multi-tabletop classrooms using mobile eye-tracking. *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, 339–344.

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest L. Rev.*, *49*, 393.

Russell Neuman, W., Guggenheim, L., Mo Jang, S. a., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of communication*, *64*(2), 193–214.

Sosu, E., & Ellis, S. (2014). *Closing the attainment gap in scottish education*. Joseph Rowntree Foundation York.

Tsiplakides, I. (2014). *The English language course in public Secondary education and the appeal to the tutorial: an empirical research on the impact of social class In its learning and certification* [Doctoral dissertation, Univerisity of Ioannina].

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological bulletin*, *140*(4), 1174.

Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. *Contemporary educational psychology*, *39*(2), 145–155.

Ahadi, A., Lister, R., Haapala, H., & Vihavainen, A. (2015). Exploring machine learning methods to automatically identify students in need of assistance. *ICER 2015 - Proceedings of the 2015 ACM Conference on International Computing Education Research*, 121–130. https://doi.org/10.1145/2787622.2787717

Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, *9*(129), 6415–6426. https://doi.org/10.12988/ams.2015.53289

Al Mutawah, M. A. (2015). The Influence of Mathematics Anxiety in Middle and High School Students Math Achievement. *International Education Studies*, *8*(11), 239. https://doi.org/10.5539/ies.v8n11p239

Al-Saleem, M., Al-Kathiry, N., Al-Osimi, S., & Badr, G. (2015). Mining Educational Data to Predict Students' Academic Performance. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (pp. 403–414, Vol. 9166). https://doi.org/10.1007/978-3-319-21024-7_28

Amaya, Y., Barrientos, E., & Heredia, D. (2015). Student Dropout Predictive Model Using Data Mining Techniques. *Ieee Latin America Transactions*, *13*(9), 3127–3134.

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015). Preprocessing and analyzing educational data set using X-API for improving student's performance. *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2015*. https://doi.org/10.1109/AEECT.2015.7360581

Beatty, A. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice*, *34*(4), 31–40.

Bergin, S., Mooney, A., Ghent, J., & Quille, K. (2015). Using Machine Learning Techniques to Predict Introductory Programming Performance. *International Journal of Computer Science and Software Engineering*, *4*(12), 323–328.

Biesta, G. (2015). What is education for? On good education, teacher judgement, and educational professionalism. *European Journal of education*, *50*(1), 75–87.

Brinton, C. G., & Chiang, M. (2015). MOOC performance prediction via clickstream data and social learning networks. *Proceedings - IEEE INFOCOM*, *26*, 2299–2307. https://doi.org/10.1109/INFOCOM.2015.7218617

Brook, C. A., & Willoughby, T. (2015). The Social Ties That Bind: Social Anxiety and Academic Achievement Across the University Years. *Journal of Youth and Adolescence*, *44*(5), 1139–1152. https://doi.org/10.1007/s10964-015-0262-8

Bydžovská, H. (2015). Are Collaborative Filtering Methods Suitable for Student Performance Prediction? In *Progress in artificial intelligence. epia 2015* (pp. 425–430). https://doi.org/10.1007/978-3-319-23485-4_42

Corrigan, O., Smeaton, A. F., Glynn, M., & Smyth, S. (2015). Using Educational Analytics to Improve Test Performance Owen. In *Design for teaching and learning in a networked world. lecture notes in computer science* (pp. 42–55). https://doi.org/10.1109/DICTAP.2015.7113173

Hall, D., Grimaldi, E., Gunter, H. M., Møller, J., Serpieri, R., & Skedsmo, G. (2015). Educational reform and modernisation in Europe: The role of national contexts in mediating the new public management. *European Educational Research Journal*, *14*(6), 487–507.

He, J., Baileyt, J., Rubinstein, B. I., & Zhang, R. (2015). Identifying at-risk students in massive open online courses. *Proceedings of the National Conference on Artificial Intelligence*, *3*, 1749–1755.

Hwang, C.-S., & Su, Y.-C. (2015). Unified clustering locality preserving matrix factorization for student performance prediction. *IIAENG - nternational Journal of Computer Science*, *43*(2), 245–253.

Iam-On, N., & Boongoen, T. (2015). Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. *International Journal of Machine Learning and Cybernetics*, *8*(2), 497–510. https://doi.org/10.1007/s13042-015-0341-x

Jishan, S. T., Rashu, R. I., Haque, N., & Rahman, R. M. (2015). Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, *2*(1), 1–25. https://doi.org/10.1186/s40165-014-0010-2

Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). How do you connect? Analysis of Social Capital Accumulation in connectivist MOOCs. *Proceedings of the fifth international conference on learning analytics and knowledge*, 64–68.

Kalaycioglu, D. B. (2015). The influence of socioeconomic status, self-efficacy, and anxiety on mathematics achievement in England, Greece, Hong Kong, the Netherlands, Turkey, and the USA. *Educational Sciences: Theory and Practice*, *15*(5), 1391–1401.

Kamateri, E., Panopoulou, E., Tambouris, E., Tarabanis, K., Ojo, A., Lee, D., & Price, D. (2015). A comparative analysis of tools and technologies for policy making. In *Policy practice and digital science* (pp. 125–156). Springer.

Klusener, M., & Fortenbacher, A. (2015). Predicting students' success based on forum activities in MOOCs. *Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2015*, *2*(September), 925–928. https://doi.org/10.1109/IDAACS.2015.7341439

Knowles, J. (2015). Building an Accurate Statewide Dropout Early Warning System in Wisconsin. *Journal of Educational Data Mining*, *7*(3), 18–67.

Koprinska, I., Stretton, J., & Yacef, K. (2015a). Predicting Student Performance from Multiple Data Sources. *Artificial Intelligence in Education*, 678–681. https://doi.org/10.1007/978-3-319-19773-9

Koprinska, I., Stretton, J., & Yacef, K. (2015b). Students at Risk : Detection and Remediation. *Proceeding of the 8th International Conference on Educational Data Mining*, 512–515.

Kostopoulos, G., Kotsiantis, S., & Pintelas, P. (2015). Predicting Student Performance in Distance Higher Education Using Semi-supervised Techniques. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (pp. 259–270, Vol. 9344). https://doi.org/10.1007/978-3-319-23781-7_21

Lahanas, A; Efstathiou, M. (2015). Why Inclusive Education, a Different Background – A Different Way of Thought. *Special Education Issues*, *69*, 3–28.

Lopez Guarin, C. E., Guzman, E. L., & Gonzalez, F. A. (2015). A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining. *Revista Iberoamericana de Tecnologias del Aprendizaje*, *10*(3), 119–125. https://doi.org/10.1109/RITA.2015.2452632

Luo, J., E. Sorour, S., Goda, K., & Mine, T. (2015). Predicting Student Grade based on Free-style Comments using Word2Vec and ANN by Considering Prediction Results Obtained in Consecutive Lessons. *Proceedings of the 8th International Conference on Educational Data Mining*, 396–399.

Morningstar, M. E., Shogren, K. A., Lee, H., & Born, K. (2015). Preliminary lessons about supporting participation and learning in inclusive classrooms. *Research and Practice for Persons with Severe Disabilities*, *40*(3), 192–210.

Naser, S. A., Zaqout, I., Ghosh, M. A., Atallah, R., & Alajrami, E. (2015). Predicting Student Performance Using Artificial Neural Network: in the Faculty of Engineering and Information Technology. *International Journal of Hybrid Information Technology*, *8*(2), 221–228. https://doi.org/10.14257/ijhit.2015.8.2.20

Navamani, J. M. A., & Kannammal, A. (2015). Predicting performance of schools by applying data mining techniques on public examination results. *Research Journal of Applied Sciences, Engineering and Technology*, *9*(4), 262–271. https://doi.org/10.19026/rjaset.9.1403

Nelson, J. D., Stahl, G., & Wallace, D. O. (2015). Race, class, and gender in boys' education: Repositioning intersectionality theory. *Culture, Society and Masculinities*, *7*(2), 171.

OECD. (2015, March). *The ABC of Gender Equality in Education*. https://doi.org/10.1787/9789264229945-en

Parmar, K., Vaghela, D., & Sharma, P. (2015). Performance Prediction of Students Using distributed Data Mining. *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, (1000), 1–5. https://doi.org/10.1109/ICIIECS.2015.7192860

Pradeep, A., Das, S., & Kizhekkethottam, J. J. (2015). Students dropout factor prediction using EDM techniques. *Proceedings of the IEEE International Conference on Soft-Computing and Network Security, ICSNS 2015*. https://doi.org/10.1109/ICSNS.2015.7292372

Rezaeinejad, M., Azizifar, A., & Gowhary, H. (2015a). The study of learning styles and its relationship with educational achievement among iranian high school students. *Procedia - Social and Behavioral Sciences*, 218–224.

Rezaeinejad, M., Azizifar, A., & Gowhary, H. (2015b). The Study of Learning Styles and its Relationship with Educational Achievement Among Iranian High School Students. *Procedia - Social and Behavioral Sciences*, *199*, 218–224. https://doi.org/10.1016/j.sbspro.2015.07.509

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12.

Ruby, J., & David, K. (2015). Analysis of Influencing Factors in Predicting Students Performance Using MLP -A Comparative Study. *International Journal of Innovative Research in Computer and Communication Engineering (An ISO Certified Organization)*, *3297*(2), 1085–1092. https://doi.org/10.15680/ijircce.2015.0302070

S. Abu-Oda, G., & M. El-Halees, A. (2015). Data Mining in Higher Education : University Student Dropout Case Study. *International Journal of Data Mining and Knowledge Management Process*, *5*(1), 15–27. https://doi.org/10.5121/ijdkp.2015.5102

Santana, M. A., Costa, E. B., Neto, B. F. S., Silva, I. C. L., & Rego, J. B. A. (2015). A predictive model for identifying students with dropout profiles in online courses. *CEUR Workshop Proceedings*, *1446*.

Sara, N.-B., Halland, R., Igel, C., & Alstrup, S. (2015). High-school dropout prediction using machine learning High-School Dropout Prediction Using Machine Learning : A Danish Large-scale Study. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 319–324.

Snellman, K., Silva, J. M., & Putnam, R. D. (2015). Inequity outside the Classroom: Growing Class Differences in Participation in Extracurricular Activities. *Voices in urban education*, *40*, 7–14.

Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., & Abreu, R. (2015). A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. *Proceedings of the 8th International Conference on Educational Data Mining*, 392–395. http://www.educationaldatamining.org/EDM2015/proceedings/short392-395.pdf

Strogilos, V., & Stefanidis, A. (2015). Contextual antecedents of co-teaching efficacy: Their influence on students with disabilities' learning progress, social participation and behaviour improvement. *Teaching and Teacher Education*, *47*, 218–229. https://doi.org/10.1016/j.tate.2015.01.008

Sukhija, K., Jindal, M., & Aggarwal, N. (2015). The recent state of educational data mining: A survey and future visions. *2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE)*, 354–359. https://doi.org/10.1109/MITE.2015.7375344

Tang, J. K. T., Xie, H., & Wong, T.-L. (2015). A Big Data Framework for Early Identification of Dropout Students in MOOC. In J. Lam, K. K. Ng, S. K. S. Cheung, T. L. Wong, K. C. Li, & F. L. Wang (Eds.), *Communications in computer and information science* (pp. 127–132, Vol. 559). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-48978-9_12

Thakar, P. (2015). Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. *arXiv*, *110*(15), 60–68. http://arxiv.org/abs/1509.05176

Yehuala, M. A. (2015). Application Of Data Mining Techniques For Student Success And Failure Prediction The Case Of Debre Markos University. *International Journal of Scientific and Technology Research*, *4*(4), 91–94.

Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American psychologist*, *70*(1), 10.

Zhou, Q., Zheng, Y., & Mou, C. (2015). Predicting students' performance of an offline course from their online behaviors. *2015 5th International Conference on Digital Information and Communication Technology and Its Applications, DICTAP 2015*, 70–73. https://doi.org/10.1109/DICTAP.2015.7113173

Agaoglu, M. (2016). Predicting Instructor Performance Using Data Mining Techniques in Higher Education. *IEEE Access*, *4*(100), 2379–2387. https://doi.org/10.1109/ACCESS.2016.2568756

Ahmed, A., & Rogers, M. (2016). *Working with marginalised groups: from policy to practice*. Macmillan International Higher Education.

Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. *International Journal of Information and Education Technology*, *6*(7), 528–533. https://doi.org/10.7763/ijiet.2016.v6.745

Alfaro, C., Cano-Montero, J., Gmez, J., Moguerza, J. M., & Ortega, F. (2016). A multistage method for content classification and opinion mining on weblog comments. *Annals of Operations Research*, *236*(1), 197–213.

Alharbi, Z., Cornford, J., Dolder, L., & De La Iglesia, B. (2016). Using data mining techniques to predict students at risk of poor performance. *Proceedings of 2016 SAI Computing Conference, SAI 2016*, 523–531. https://doi.org/10.1109/SAI.2016.7556030

Altujjar, Y., Altamimi, W., Al-Turaiki, I., & Al-Razgan, M. (2016). Predicting Critical Courses Affecting Students Performance: A Case Study. *Procedia Computer Science*, *82*(March), 65–71. https://doi.org/10.1016/j.procs.2016.04.010

Amornsinlaphachai, P. (2016). Efficiency of data mining models to predict academic performance and a cooperative learning model. *2016 8th International Conference on Knowledge and Smart Technology, KST 2016*, 66–71. https://doi.org/10.1109/KST.2016.7440483

Anuradha, C., & Velmurugan, T. (2016). Performance Evaluation of Feature Selection Algorithms in Educational Data Mining. *International Journal of Data Mining Techniques and Applications*, *5*(2), 131–139. https://doi.org/10.20894/IJDMTA.102.005.002.007

Bablekou, Z., & Kazi, S. (2016). Intellectual assessment of children and adolescents: The case of Greece. *International Journal of School and Educational Psychology*, *4*(4), 225–230. https://doi.org/10.1080/21683603.2016.1163655

Baker, R. S., Martin, T., & Rossi, L. M. (2016). Educational data mining and learning analytics. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, 379–396.

Bhegade, K. B., & Shinde, S. V. (2016). Student Performance Prediction System with Educational Data Mining. *International Journal of Computer Applications*, *146*(5), 975–8887. https://doi.org/10.17148/IJARCCE.2017.63195

Brookhart, S., et al. (2016). A century of grading research: Meaning and value in the most common educational measure. *Educational Research*, *86*(4), 803–848.

Buniyamin, N., Mat, U. B., & Arshad, P. M. (2016). Educational data mining for prediction and classification of engineering students achievement. *2015 IEEE 7th International Conference on Engineering Education, ICEED 2015*, 49–53. https://doi.org/10.1109/ICEED.2015.7451491

Bydžovská, H. (2016). A comparative analysis of techniques for predicting student performance. *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016*, *2*, 306–311.

Cinque, M. (2016). “lost in translation”. Soft skills development in european countries. *Tuning Journal for Higher Education*, *3*(2), 389–427. https://doi.org/https://doi.org/10.18543/tjhe-3(2)-2016pp389-427

Comendador, B. E. V., Rabago, L. W., & Tanguilig, B. T. (2016). An educational model based on Knowledge Discovery in Databases (KDD) to predict learners behavior using classification techniques. *2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. https://doi.org/10.1109/icspcc.2016.7753623

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28.

Day, C., Gu, Q., & Sammons, P. (2016). The Impact of Leadership on Student Outcomes. *Educational Administration Quarterly*, *52*(2), 221–258. https://doi.org/10.1177/0013161X15616863

European_Union. (2016). Eur-lex - 32016r0679 - en - eur-lex. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Fan, Z., & Yan, S. (2016). Clustering of College Students Based on Improved K-Means Algorithm. *2016 International Computer Symposium (ICS)*, 676–679. https://doi.org/10.1109/ICS.2016.0139

Gershenfeld, S., Hood, D. W., & Zhan, M. (2016). The role of first-semester gpa in predicting graduation rates of underrepresented students. *Journal of College Student Retention: Research, Theory and Practice*, *17*(4), 469–488.

Gray, C. M., & Boling, E. (2016). Inscribing ethics and values in designs for learning: a problematic. *Educational technology research and development*, *64*(5), 969–1001.

Greek National Printing House. (2016). Presidential decree 126/2016. https://www.et.gr/api/DownloadFeksApi/?fek_pdf=20160100211

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2016). Predicting Students Performance in Educational Data Mining. *Proceedings - 2015 International Symposium on Educational Technology, ISET 2015*, 125–128. https://doi.org/10.1109/ISET.2015.33

Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. *Procedia Technology*, *25*, 326–332. https://doi.org/10.1016/j.protcy.2016.08.114

Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. *Proceedings of the sixth international conference on learning analytics & knowledge*, 198–207.

Hoogland, I., Schildkamp, K., der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and teacher education*, *60*, 377–386.

Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. *Government Information Quarterly*, *33*(1), 89–104.

Kaur, G., & Singh, W. (2016). Prediction of Student Performance Using Weka Tool. *International Journal of Engineering Sciences*, *17*(1), 8–16.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? *arXiv*.

Klievink, B., Bharosa, N., & Tan, Y.-H. (2016). The collaborative realization of public values and business goals: Governance and infrastructure of public–private information platforms. *Government information quarterly*, *33*(1), 67–79.

Kumar, S. A. (2016). Edifice an Educational Framework using Educational Data Mining and Visual Analytics. *International Journal of Education and Management Engineering*, *6*(2), 24–30. https://doi.org/10.5815/ijeme.2016.02.03

Lai, C.-L., & Hwang, G.-J. (2016). A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course. *Computers and Education*, *100*, 126–140. https://doi.org/10.1016/j.compedu.2016.05.006

Lee, J. (2016). Attitude toward school does not predict academic achievement. *Learning and Individual Differences*, *52*, 1–9. https://doi.org/10.1016/j.lindif.2016.09.009

Liang, J., Li, C., & Zheng, L. (2016). Machine learning application in MOOCs: Dropout prediction. *2016 11th International Conference on Computer Science and Education (ICCSE)*, 52–57. https://doi.org/10.1109/ICCSE.2016.7581554

Livieris, I., Mikropoulos, T., & Pintelas, P. (2016). A decision support system for predicting students' performance. *Themes in Science and Technology Education*, *9*(1), 43–57.

Livingstone, S., Carr, J., & Byrne, J. (2016). One in three: Internet governance and children's rights.

Meedech, P., Iam-On, N., & Boongoen, T. (2016). Prediction of Student Dropout Using Personal Profile and Data Mining Approach. In *Intelligent and evolutionary systems* (pp. 143–155, Vol. 5). https://doi.org/10.1007/978-3-319-27000-5_12

Merchan Rubiano, S. M., & Duarte Garcia, J. A. (2016). Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance. *IEEE Latin America Transactions*, *14*(6), 2783–2788. https://doi.org/10.1109/TLA.2016.7555255

Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. *International Journal of Modern Education and Computer Science*, *8*(11), 36–42. https://doi.org/10.5815/ijmecs.2016.11.05

Mundada, O. (2016). Mining Educational Data from Student ' s Management System. *International Journal of Advanced Research in Computer Science*, *7*(3), 244–249.

Ohlson, M., Swanson, A., Adams-Manning, A., & Byrd, A. (2016). A Culture of Success—Examining School Culture and Student Outcomes via a Performance Framework. *Journal of Education and Learning*, *5*(1), 114. https://doi.org/10.5539/jel.v5n1p114

Paidousi, C. (2016, February). https://www.eiead.gr/wp-content/uploads/2012/01/publications_docs__%CE%95%CE%9B%CE%9A%CE%A5%CE%A3%CE%A4%CE%99%CE%9A%CE%9F%CE%A4%CE%97%CE%A4%CE%91__%CF%84%CE%B7%CF%82_%CE%95%CF%80%CE%B1%CE%B3%CE%B3%CE%B5%CE%BB%CE%BC%CE%B1%CF%84%CE%B9%CE%BA%CE%AE%CF%82_%CE%95%CE%BA%CF%80%CE%B1%CE%AF%CE%B4%CE%B5%CF%85%CF%83%CE%B7%CF%82.pdf

Pandey, M., & Taruna, S. (2016). Towards the integration of multiple classifier pertaining to the Student's performance prediction. *Perspectives in Science*, *8*, 364–366. https://doi.org/10.1016/j.pisc.2016.04.076

Reardon, S. F. (2016). School segregation and racial academic achievement gaps. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(5), 34–57.

Rio, D., C., Insuasti, P., & J. (2016). *Predicting academic performance in traditional environments at higher-education institutions using data mining: A review*. Ecos de la Academia(4).

Satyanarayana, A., & Nuckowski, M. (2016). *Data Mining using Ensemble Classifiers for Improved Prediction of Student Academic Performance Data Mining using Ensemble Classifiers for Improved Prediction of Student Academic Performance Ashwin*

Satyanarayana, Mariusz Nuckowski (tech. rep.). New York. https://academicworks.cuny.edu/ny_pubs

Sivakumar, S., Venkataraman, S., & Selvaraj, R. (2016). Predictive Modeling of Student Dropout Indicators in Educational Data Mining using Improved Decision Tree. *Indian Journal of Science and Technology*, *9*(January). https://doi.org/10.17485/ijst/2016/v9i4/87032

Skourtou, E. (2016). Education of Roma children. In *Bilingualism and teaching greek as a second language* (pp. 97–112). Kallipos, Open Academic Editions.

Slater, J. J., & Mudryj, A. N. (2016). Self-perceived eating habits and food skills of Canadians. *Journal of nutrition education and behavior*, *48*(7), 486–495.

Strogilos, V., Stefanidis, A., & Tragoulia, E. (2016). Co-teachers' attitudes towards planning and instructional activities for students with disabilities. *European Journal of Special Needs Education*, *31*(3), 344–359.

Sumitha, R., Vinothkumar, E. S., & Scholar, P. G. (2016). Prediction of Students Outcome Using Data Mining Techniques. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, *2*(6), 132–139. www.ijseas.com

Terrier, C. (2016). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review*.

Union, E. (2016). Eur-lex - 32016r0679 - en - eur-lex. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Vlachou, A., & Fyssa, A. (2016). 'Inclusion in practice': Programme practices in mainstream preschool classrooms and associations with context and teacher characteristics. *International Journal of Disability, Development and Education*, *63*(5), 529–544.

Williamson, B. (2016). Digital education governance: An introduction.

Wilson, K. H., Xiong, X., Khajah, M., Lindsey, R., Zhao, S., Karklin, Y., Van Inwegen, E. G., Han, B., Beck, J. E., Heffernan, N., & Mozer, M. (2016). Estimating student proficiency: Deep learning is not the panacea. *Neural Information Processing Systems, Workshop on Machine Learning for Education*, (Nips), 3.

Abdul-Hamid, H. (2017). *Value of Data: Better Data, Better Education*. Worldbank.

Agasisti, T., & Bowers, A. J. (2017). Data analytics and decision making in education: towards the educational data scientist as a key actor in schools and higher education institutions. In *Johnes, g., johnes, j., agasisti, t., lpez-torres, l* (pp. 184–210). Edward Elgar Publishing.

Alhosani, A. A., Singh, S. K., & Al Nahyan, M. T. (2017). Role of school leadership and climate in student achievement. *International Journal of Educational Management*, *31*(6), 843–851. https://doi.org/10.1108/IJEM-05-2016-0113

Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., Alhiyafi, J., & Olatunji, S. O. (2017). Student performance prediction using Support Vector Machine and K-Nearest Neighbor. *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. https://doi.org/10.1109/ccece.2017.7946847

Arens, A. K., Marsh, H. W., Pekrun, R., Lichtenfeld, S., Murayama, K., & vom Hofe, R. (2017). Math self-concept, grades, and achievement test scores: Long-term reciprocal effects across five waves and three achievement tracks. *Journal of Educational Psychology*, *109*(5), 621–634. https://doi.org/10.1037/edu0000163

Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, *113*, 177–194. https://doi.org/10.1016/j.compedu.2017.05.007

Athani, S. S., Kodli, S. A., Banavasi, M. N., & Hiremath, P. G. S. (2017). Student academic performance and social behavior predictor using data mining techniques. *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 170–174. https://doi.org/10.1109/CCAA.2017.8229794

Bauer, A., Flatten, J., & Popovic, Z. (2017). Analysis of Problem-Solving Behavior in Open-Ended Scientific-Discovery Game Challenges. *International Educational Data Mining Society*.

Ben Shahar, T. H. (2017). Educational justice and big data. *Theory and Research in Education*, *15*(3), 306–320.

Berman, G., & Albright, K. (2017). Children and the data cycle: Rights and ethics in a big data world. *arXiv*.

Castro-Wunsch, K., Ahadi, A., & Petersen, A. (2017). Evaluating neural networks as a method for identifying students in need of assistance. *Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE*, 111–116. https://doi.org/10.1145/3017680.3017792

CEME, N. (2017). Student performance prediction and risk analysis by using data mining approach. *Journal of Intelligent Computing Volume*, *8*(2), 49.

Charles Lang, George Siemens, Alyssa Wise, & Dragan Gašević. (2017). *The Handbook of Learning Analytics*. Society for Learning Analytics Research (SoLAR).

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017a). Evaluating the effectiveness of educational data mining techniques for early prediction of students academic failure in introductory programming courses. *Computers in Human Behavior*, *73*, 247–256. https://doi.org/10.1016/j.chb.2017.01.047

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017b). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, *73*, 247–256. https://doi.org/10.1016/j.chb.2017.01.047

Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017). Predicting Student Performance using Advanced Learning Analytics. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*. https://doi.org/10.1145/3041021.3054164

Domnech-Betoret, F., Abellan-Rosello, L., & Gomez-Artiga, A. (2017a). Self-efficacy, satisfaction, and academic achievement: The mediator role of students' expectancy-value beliefs. *Frontiers in Psychology*, *8*, 1664–1078.

Domnech-Betoret, F., Abellan-Rosello, L., & Gomez-Artiga, A. (2017b). Self-Efficacy, Satisfaction, and Academic Achievement: The Mediator Role of Students' Expectancy-Value Beliefs. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01193

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. *Ieee Access*, *5*, 15991–16005.

Enli, G. (2017). Twitter as arena for the authentic outsider: exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *European journal of communication*, *32*(1), 50–61.

Epp, C. D., Phirangee, K., & Hewitt, J. (2017). Student actions and community in online courses: The roles played by course length and facilitation method. *Online Learning Journal*, *21*(4).

Gkofa, P. (2017). Promoting Social Justice and Enhancing Educational Success: Suggestions from Twenty Educationally Successful Roma in Greece. *Urban Rev*, *49*, 443–462. https://doi.org/10.1007/s11256-016-0393-6

Govindasamy, K., & Velmurugan, T. (2017). A Study on Classi fi cation and Clustering Data Mining Algorithms based on Students Academic Performance Prediction. *International Journal of control theory and Applications*, *10*(23), 147–160.

Hall, J. D., O'Connell, A. B., & Cook, J. G. (2017). Predictors of student productivity in biomedical graduate school applications. *PLoS One*, *12*(1), e0169121.

Head, A., Glassman, E., Soares, G., Suzuki, R., Figueredo, L., D'Antoni, L., & Hartmann, B. (2017). Writing reusable code feedback at scale with mixed-initiative program synthesis. *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, 89–98.

Hill, H. C. (2017). The Coleman Report, 50 years on: What do we know about the role of schools in academic inequality? *The ANNALS of the American Academy of Political and Social Science*, *674*(1), 9–26.

Hung, J.-L., Wang, M. C., Wang, S., Abdelrasoul, M., Li, Y., & He, W. (2017). Identifying At-Risk Students for Early Interventions—A Time-Series Clustering Approach.

IEEE Transactions on Emerging Topics in Computing, *5*(1), 45–55. https://doi.org/10.1109/TETC.2015.2504239

Janssen, M., Konopnicki, D., Snowdon, J. L., & Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information systems frontiers*, *19*(2), 189–195.

Karadaug, E. (2017). *The factors effecting student achievement: Meta-analysis of empirical studies*. Springer.

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media and Society*, *19*(5), 780–794. https://doi.org/10.1177/1461444816686328

Mahboob, T., Irfan, S., & Karamat, A. (2017). A machine learning approach for student assessment in E-learning using Quinlan's C4.5, Naive Bayes and Random Forest algorithms. *Proceedings of the 2016 19th International Multi-Topic Conference, INMIC 2016*. https://doi.org/10.1109/INMIC.2016.7840094

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *Aera Open*, *3*(4), 2332858417743754.

Mehboob, B., Liaqat, R. M., & Nazar, A. (2017). Student Performance Prediction and Risk Analysis by Using Data Mining Approach. *Journal of Intelligent Computing*, *8*(2), 49–57.

Méndez, G. G., Hinrichs, U., & Nacenta, M. A. (2017). Bottom-up vs. top-down: Trade-offs in efficiency, understanding, freedom and creativity with infovis tools. *Proceedings of the 2017 CHI conference on human factors in computing systems*, 841–852.

Mhetre, V., & Nagar, M. (2017). Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA. *2017 International Conference on Computing Methodologies and Communication (ICCMC)*. https://doi.org/10.1109/iccmc.2017.8282735

Okubo, F., Shimada, A., Yamashita, T., & Ogata, H. (2017). A neural network approach for students' performance prediction. *ACM International Conference Proceeding Series*, (March), 598–599. https://doi.org/10.1145/3027385.3029479

Orhan Özen, S. (2017). The Effect of Motivation on Student Achievement. In *The factors effecting student achievement* (pp. 35–56). Springer International Publishing. https://doi.org/10.1007/978-3-319-56083-0_3

Ozen, S. (2017). The effect of motivation on student achievement. In *The factors affecting student achievement* (pp. 35–56). Springer International Publishing.

Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(1), e1187.

Scheihing, E., Vernier, M., Guerra, J., Born, J., & Crcamo, L. (2017). Understanding the role of micro-blogging in B-learning activities: Kelluwen experiences in Chilean public schools. *IEEE Transactions on Learning Technologies*, *11*(3), 280–293.

Sebastian, J., Moon, J.-M., & Cunningham, M. (2017). The relationship of school-based parental involvement with student achievement: a comparison of principal and parent survey reports from PISA 2012. *Educational Studies*, *43*(2), 123–146. https://doi.org/10.1080/03055698.2016.1248900

Shrivas, A. K., & Tiwari, P. (2017). Comparative Analysis of Models for Student Performance with Data Mining Tools. *International Journal of Computer Trends and Technology*, *46*(1), 42–46. https://doi.org/10.14445/22312803/ijctt-v46p109

Simon, C. A. (2017). *Public policy: Preferences and outcomes*. Routledge.

Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardized effect sizes. *Journal of Education Policy*, *32*(4), 450–466. https://doi.org/10.1080/02680939.2017.1280183

Sorić, I., Penezić, Z., & Burić, I. (2017). The Big Five personality traits, goal orientations, and academic achievement. *Learning and Individual Differences*, *54*, 126–134. https://doi.org/10.1016/j.lindif.2017.01.024

Srnicek, N. (2017). *Platform capitalism*. John Wiley & Sons.

Toth, K., Rolke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills.

Viennet, R., & Pont, B. (2017). Education policy implementation: A literature review and proposed framework. *OECD Education Working Papers*, (162). https://doi.org/10.1787/fc467a64-en

Wang, C., Hu, Q., Wang, X., Chen, D., Qian, Y., & Dong, Z. (2017). Feature selection based on neighborhood discrimination index. *IEEE transactions on neural networks and learning systems*, *29*(7), 2986–2999.

Widyahastuti, F., & Tjhin, V. U. (2017). Predicting students performance in final examination using linear regression and multilayer perceptron. *Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017*, 188–192. https://doi.org/10.1109/HSI.2017.8005026

Williamson, B. (2017, September). *Big data in education*.

Yao, C., Liu, Y.-F., Jiang, B., Han, J., & Han, J. (2017). Lle score: A new filter-based unsupervised feature selection method based on nonlinear manifold embedding and its application to image recognition. *IEEE Transactions on Image Processing*, *26*(11), 5257–5269.

Abdulrahman Al-Noshan, A., Abdullah Al-Hagery, M., Abdulaziz Al-Hodathi, H., & Sulaiman Al-Quraishi, M. (2018). Performance Evaluation and Comparison of Classification Algorithms for Students at Qassim University. *International Journal of Science and Research*. https://doi.org/10.21275/ART20202907

Allen, W. R., McLewis, C., Jones, C., & Harris, D. (2018). From Bakke to Fisher: African American students in US higher education over forty years. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *4*(6), 41–72.

Anand, V., Rahiman, S. A., George, E. B., & Huda, A. (2018). Recursive clustering technique for students' performance evaluation in programming courses. *2018 Majan International Conference (MIC)*, 1–5.

Atapattu, T., & Falkner, K. (2018). Impact of lecturer's discourse for students' video engagement: Video learning analytics case study of moocs. *Journal of Learning Analytics*, *5*(3), 182–197.

Boles, L. (2018). Predicting graduate school success in a speech-language pathology program. *Teaching and Learning in Communication Sciences and Disorders*, *2*(2).

Bradshaw, S., & Howard, P. N. (2018). Challenging truth and trust: A global inventory of organized social media manipulation. *The computational propaganda project*, *1*, 1–26.

Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, *66*, 541–556. https://doi.org/10.1016/j.compeleceng.2017.03.005

Camargo, B., Camelo, R., Firpo, S., & Ponczek, V. (2018). Information, market incentives, and student performance evidence from a regression discontinuity design in Brazil. *Journal of Human Resources*, *53*(2), 414–444.

Chanlekha, H., & Niramitranon, J. (2018). Student performance prediction model for early-identification of at-risk students in traditional classroom settings. *Proceedings of the 10th International Conference on Management of Digital EcoSystems - MEDES '18*, 239–245. https://doi.org/10.1145/3281375.3281403

Chaturapruek, S., Dee, T. S., Johari, R., Kizilcec, R. F., & Stevens, M. L. (2018). How a data-driven course planning tool affects college students' GPA: evidence from two field experiments. *Proceedings of the fifth annual ACM conference on learning at scale*, 1–10.

Cilliers, J., Kasirye, I., Leaver, C., Serneels, P., & Zeitlin, A. (2018). Pay for locally monitored performance? a welfare analysis for teacher attendance in ugandan primary schools. *Journal of Public Economics*, *167*, 69–90.

Crossley, S. A., & Kyle, K. (2018). Analyzing spoken and written discourse: A role for natural language processing tools. In *The palgrave handbook of applied linguistics research methodology* (pp. 567–594). Springer.

Crues, R. W., Henricks, G. M., Perry, M., Bhat, S., Anderson, C. J., Shaik, N., & Angrave, L. (2018). How do gender, learning goals, and forum participation predict persistence in a computer science MOOC? *ACM Transactions on Computing Education (TOCE)*, *18*(4), 1–14.

Custer, S., King, E. M., Atinc, T. M., Read, L., & Sethi, T. (2018). Toward Data-Driven Education Systems: Insights into Using Information to Measure Results and Manage Change. *Center for Universal Education at The Brookings Institution*.

Cvencek, D., Fryberg, S. A., Covarrubias, R., & Meltzoff, A. N. (2018). Self-Concepts, Self-Esteem, and Academic Achievement of Minority and Majority North American Elementary School Children. *Child Development*, *89*(4), 1099–1109. https://doi.org/10.1111/cdev.12802

Davis, D., Triglianos, V., Hauff, C., & Houben, G.-J. (2018). SRLx: A personalized learner interface for MOOCs. *European Conference on Technology Enhanced Learning*, 122–135.

de Boer, H., Timmermans, A., & van der Werf, M. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: Narrative review and meta-analysis. *Educational Research and Evaluation*, *4*(3), 180–200.

de Boer, H., Timmermans, A. C., & van der Werf, M. P. C. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: narrative review and meta-analysis. *Educational Research and Evaluation*, *24*(3-5), 180–200. https://doi.org/10.1080/13803611.2018.1550834

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education*, *28*(2), 152–193.

de Sousa, A. C. C.-R., de Oliveira, C. A. B., Borges, J. L. C. M., de Sousa, A. C. C.-R., de Oliveira, C. A. B., & Borges, J. L. C. M. (2018). Using Academic Performance to Predict College Students Dropout: a case study. *Educação e Pesquisa*, *44*(0), 1–17. https://doi.org/10.1590/s1678-4634201844180590

Easton, J. Q., Johnson, E., & Sartain, L. (2018). *The predictive power of ninth-grade GPA* (tech. rep. No. 2018-10). University of Chicago Consortium on School Research.

Fancsali, C., Tigani, L., Toro Isaza, P., & Cole, R. (2018). A landscape study of computer science education in nyc: Early findings and implications for policy and practice. *Proceedings of the 49th acm technical symposium on computer science education*, 44–49.

Gampa, R., & Baynes, A. (2018). Visual Analytic Workflow to Understand Students' Performance in Computer Science Courses. *2018 IEEE Frontiers in Education Conference (FIE)*, 1–5. https://doi.org/10.1109/FIE.2018.8658790

Geller, J., Toftness, A. R., Armstrong, P. I., Carpenter, S. K., Manz, C. L., Coffman, C. R., & Lamm, M. H. (2018). Study strategies and beliefs about learning as a function of academic achievement and achievement goals. *Memory*, *26*(5), 683–690. https://doi.org/10.1080/09658211.2017.1397175

Hasheminejad, S. M., & Sarvmili, M. (2018). S3PSO: Students' Performance Prediction Based on Particle Swarm Optimization. *Journal of AI and Data Mining*, *0*(0). https://doi.org/10.22044/jadm.2018.5506.1662

Hussain, M., Hussain, S., Zhang, W., Zhu, W., Theodorou, P., & Abidi, S. M. R. (2018). Mining moodle data to detect the inactive and low-performance students during the moodle course. *ACM International Conference Proceeding Series*, 133–140. https://doi.org/10.1145/3291801.3291828

Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447–459. https://doi.org/10.11591/ijeecs.v9.i2.pp447-459

Islam, S., Baharun, H., Muali, C., Ghufron, M. I., Bali, M. e. I., Wijaya, M., & Marzuki, I. (2018). To Boost Students' Motivation and Achievement through Blended Learning. *Journal of Physics: Conference Series*, *1114*, 012046. https://doi.org/10.1088/1742-6596/1114/1/012046

Kiu, C. C. (2018). Data Mining Analysis on Student's Academic Performance through Exploration of Student's Background and Social Activities. *Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018*, 1–5. https://doi.org/10.1109/ICACCAF.2018.8776809

Kohn, A. (2018). *Punished by rewards: Twenty-fifth anniversary edition: The trouble with gold stars, incentive plans, a's, praise and other bribes*. HMH Books.

Kokkinos, P., C. M.and Kiprianos. (2018). Remedial teaching and student motivation: A controlled field study. *Journal of Educational and Developmental Psychology*, *8*(2), 35–37.

Konold, T., Cornell, D., Jia, Y., & Malone, M. (2018). School Climate, Student Engagement, and Academic Achievement: A Latent Variable, Multilevel Multi-Informant Examination. *AERA Open*, *4*(4), 233285841881566. https://doi.org/10.1177/2332858418815661

Li, K., Yu, M., Liu, L., Li, T., & Zhai, J. (2018). Feature selection method based on weighted mutual information for imbalanced data. *International Journal of Software Engineering and Knowledge Engineering*, *28*(08), 1177–1194.

Livieris, I. E., Drakopoulou, K., Mikropoulos, T. A., Tampakas, V., & Pintelas, P. (2018). An Ensemble-Based Semi-Supervised Approach for Predicting Students' Performance. *Research on e-Learning and ICT in Education*, 25–42. https://doi.org/10.1007/978-3-319-95059-4_2

Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Lin, A. J. Q., Ogata, H., & Yang, S. J. H. (2018). Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology and Society*, *21*(2), 220–232.

Mahzoon, M. J., Maher, M. L., Eltayeby, O., Dou, W., & Grace, K. (2018). A framework for interactive exploratory learning analytics. *International Conference on Learning and Collaboration Technologies*, 319–331.

Marsh, H., Pekrun, R., Murayama, K., Arens, K., Parker, P., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*(2), 263–280.

Marsh, H. W., Pekrun, R., Murayama, K., Arens, A. K., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over 6 years. *Developmental Psychology*, *54*(2), 263–280. https://doi.org/10.1037/dev0000393

Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, *115*, 36–51.

Namomsa, G., & Sharma, D. P. (2018). Comparative Analytics and Predictive Modeling of Student Performance Through Data Mining Techniques. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3395126

Nemomsa, G., & Mulugeta, A. (2018). COMPARATIVE ANALYTICS AND PREDICTIVE MODELING OF. *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3395126

Nuñez, H., Maldonado, G., & Astudillo, C. A. (2018). Semi-supervised regression based on tree SOMs for predicting students performance. *IET Conference Publications*, *2018*(CP745), 65–71. https://doi.org/10.1049/cp.2018.1288

Park, S. K., Lee, C. K., & Kim, H. (2018). Suicide mortality and marital status for specific ages, genders, and education levels in South Korea: using a virtually individualized dataset from national aggregate data. *Journal of affective disorders*, *237*, 87–93.

Polyzou, A., & Karypis, G. (2018). Feature extraction for classifying students based on their academic performance. *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018*.

Pristyanto, Y., Pratama, I., & Nugraha, A. F. (2018). Data level approach for imbalanced class handling on educational data mining multiclass classification. *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, *2018-Janua*, 310–314. https://doi.org/10.1109/ICOIACT.2018.8350792

Schenke, W., & Meijer, J. (2018). Datagebruik in Het Onderwijs. Problematiek Uiteengezet [Data Use in Education: An Overview of the Problems].

Singh, J., & Kaur, E. H. (2018). Implementation of Classification Based Algorithms to Prediction of Students Performance Using Educational System. *International Journal of Engineering Science and Computing*, *8*(11), 19233–19238.

Sivakumar, S., & Selvaraj, R. (2018). Predictive Modeling of Students Performance Through the Enhanced Decision Tree. https://doi.org/10.1007/978-981-10-4765-7_3

Tegegne, A. K., & Alemu, T. A. (2018). Educational data mining for students' academic performance analysis in selected Ethiopian universities. *Information Impact: Journal of Information and Knowledge Management*, *9*(2), 1. https://doi.org/10.4314/iijikm.v9i2.1

Timarán Pereira, R., Hidalgo Troya, A., Caicedo Zambrano, J., & Benavides, J. (2018). Discovering Factors Associated with Academic Performance of High School Students in Saber 11th Test using Educational Data Mining Techniques. *Association for the Advancement of Computing in Education (AACE)*, (March 2019).

Tomlinson, S. (2018). Special education and minority ethnic young people in England: continuing issues. In *Special education and globalisation* (pp. 25–40). Routledge.

Tsiplakides, I. (2018). Shadow education and social class inequalities in secondary education in Greece: The case of teaching English as a foreign language. *International Journal of Sociology of Education*, *7*(1), 71–93.

Ulloa-Cazarez, R. L., López-Martín, C., Abran, A., & Yáñez-Márquez, C. (2018). Prediction of online students performance by means of genetic programming. *Applied Artificial Intelligence*, *32*(9-10), 858–881. https://doi.org/10.1080/08839514.2018.1508839

van Rens, M., Haelermans, C., Groot, W., & Maassen van den Brink, H. (2018). Facilitating a successful transition to secondary school:(how) does it work? a systematic literature review. *Adolescent Research Review*, *3*(1), 43–56.

Vulperhorst, J., Lutz, C., de Kleijn, R., & van Tartwijk, J. D. (2018). Disentangling the predictive validity of high school grades for academic success in university. *Assessment & Evaluation in Higher Education*, *43*(3), 399–414.

Wang, K., Singh, R., & Su, Z. (2018). Search, align, and repair: data-driven feedback generation for introductory programming exercises. *Proceedings of the 39th ACM SIGPLAN conference on programming language design and implementation*, 481–495.

Zhang, X., Sun, G., Pan, Y., Sun, H., He, Y., & Tan, J. (2018). Students performance modeling based on behavior pattern. *Journal of Ambient Intelligence and Humanized Computing*, *9*(5), 1659–1670. https://doi.org/10.1007/s12652-018-0864-6

Abdulazeez, Y., & Abdulwahab, L. (2019). Application of classification models to predict students' academic performance using classifiers ensemble and synthetic minority over sampling techniques. *Bayero Journal of Pure and Applied Sciences*, *11*(2), 142. https://doi.org/10.4314/bajopas.v11i2.17

Adekitan, A. I., & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining. *Heliyon*, *5*(2), e01250. https://doi.org/10.1016/j.heliyon.2019.e01250

Afeni, B. O., Oloyede, I. A., & Okurinboye, D. (2019). Students' Performance Prediction Using Classsification Algorithms. *Journal of Advances in Mathematics and Computer Science*, *30*(2), 1–9. https://doi.org/10.9734/jamcs/2019/45438

Aggarwal, D., & Sharma, D. (2019). Application of clustering for student result analysis. *Int J Recent Technol Eng*, *7*(6), 50–53.

Aggarwal, S., & Singh, P. (2019). Cuckoo and krill herd□based k□means++ hybrid algorithms for clustering. *Expert Systems*, *36*(4). https://doi.org/10.1111/exsy.12353

Akanni, O. R. (2019). Citizen Participation and Policy Process in a Fragile State: Experience from South-West Nigeria. *Social Science and La w Journal of Policy Review and Development Strategies*, *7*(1), 158–175.

Almasri, A., Celebi, E., & Alkhawaldeh, R. (2019). MISDataset: Management information systems dataset for predicting undergraduate students' performance. *Proceedings - 2019 4th International Conference on Computational Intelligence and Applications, ICCIA 2019*, 54–57. https://doi.org/10.1109/ICCIA.2019.00017

Almutairi, S., Shaiba, H., & Bezbradica, M. (2019). Predicting Students' Academic Performance and Main Behavioral Features Using Data Mining Techniques. In *Advances in data science, cyber security and it applications* (pp. 245–259). https://doi.org/10.1007/978-3-030-36365-9_21

Altaf, S., Soomro, W., & Rawi, M. I. M. (2019). Student Performance Prediction using Multi-Layers Artificial Neural Networks. *Proceedings of the 2019 3rd International Conference on Information System and Data Mining - ICISDM 2019*. https://doi.org/10.1145/3325917.3325919

Amirhajlou, L., Sohrabi, Z., Alebouyeh, M. R., Tavakoli, N., Haghighi, R. Z., Hashemi, A., Asoodeh, A., et al. (2019). Application of data mining techniques for predicting residents' performance on pre-board examinations: A case study. *Journal of Education and Health Promotion*, *8*(1), 108.

Angiani, G., Ferrari, A., Fornacciari, P., Mordonini, M., & Tomaiuolo, M. (2019). *Real marks analysis for predicting students' performance* (Vol. 804). Springer International Publishing. https://doi.org/10.1007/978-3-319-98872-6_5

Anthony, C. J., & Ogg, J. (2019). Parent involvement, approaches to learning, and student achievement: Examining longitudinal mediation. *School Psychology*, *34*(4), 376.

Anwer, F. (2019). Activity-based teaching, student motivation and academic achievement. *Journal of Education and Educational Development*, *6*(1), 154–170.

Barberá, P., Casas, A., Nagler, J., Egan, P. J., Bonneau, R., Jost, J. T., & Tucker, J. A. (2019). Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data. *American Political Science Review*, *113*(4), 883–901.

Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, *12*(2), 158–170.

Breda, T., & Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences*, *116*(31), 15435–15440.

Brimi, H. R. (2019). Reliability of grading high school work in english. *Practical Assessment, Research, and Evaluation*, *16*(17), 1–12.

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. *Sustainability*, *11*(10), 2833. https://doi.org/10.3390/su11102833

Canagareddy, D., Subarayadu, K., & Hurbungs, V. (2019). A Machine Learning Model to Predict the Performance of University Students. In *Smart and sustainable engineering for next generation applications* (pp. 313–322). Springer International Publishing. https://doi.org/10.1007/978-3-030-18240-3_29

Chango, W., Cerezo, R., & Romero, C. (2019). Predicting academic performance of university students from multi-sources data in blended learning. *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems - DATA '19*, 1–5. https://doi.org/10.1145/3368691.3368694

Christodoulou, A., Kyriakides, L., & Koutselini, M. (2019). The experiences of students receiving remedial teaching: A qualitative study in the greek educational context. *European Journal of Psychology of Education*, *36*(2), 335–354.

Creemers, R. (2019). Artificial Intelligence, China, Russia, and the Global Order. In *The international and foreign policy impact of china's artificial intelligence and big-data strategies* (pp. 129–135). Air University Press.

Daniel, B. K. (2019). Big Data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, *50*(1), 101–113.

Eklund, J., & Isaksson, F. (2019). Identifying and Evaluating SystemComponents for Cognitive Trustin AI-Automated Service Encounters: Trusting a Study-and Vocational Chatbot.

Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. V. (2019). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. *Journal of Business Research*, *94*, 335–343. https://doi.org/10.1016/j.jbusres.2018.02.012

Fesler, L., Dee, T., Baker, R., & Evans, B. (2019). Text as data methods for education research. *Journal of Research on Educational Effectiveness*, *12*(4), 707–727.

Francis, B. K., & Babu, S. S. (2019). Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. *Journal of Medical Systems*, *43*(6), 162. https://doi.org/10.1007/s10916-019-1295-4

González-Marcos, A., Olarte-Valentín, R., Ordieres-Meré, J., & Alba-Elías, F. (2019). Predicting students' performance in a virtual experience for project management learning. *CSEDU 2019 - Proceedings of the 11th International Conference on Computer Supported Education*, *1*(Csedu), 665–673. https://doi.org/10.5220/0007843506650673

Harris, M. K., & Williams, L. T. (2019). Evidence-based policymaking in auditing regulation: An historical analysis of academic research citation in PCAOB standard-setting practices. *Journal of Accounting and Public Policy*, *38*(3), 238–251.

Hassan, H., Anuar, S., & Ahmad, N. B. (2019). Students' performance prediction model using meta-classifier approach. *International Conference on Engineering Applications of Neural Networks*, 221–231.

Hooshyar, D., Pedaste, M., & Yang, Y. (2019). Mining educational data to predict students' performance through procrastination behavior. *Entropy*, *22*(1), 12.

Huang, A. Y. Q., Lu, O. H. T., Huang, J. C. H., Yin, C. J., & Yang, S. J. H. (2019). Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments*, 1–25. https://doi.org/10.1080/10494820.2019.1636086

Hussain, S., Muhsin, Z. F., Salal, Y. K., Theodorou, P., Kurtoğlu, F., & Hazarika, G. C. (2019). Prediction model on student performance based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning*, *14*(8), 4–22. https://doi.org/10.3991/ijet.v14i08.10001

Hutt, S., Gardner, M., Duckworth, A. L., & D'Mello, S. K. (2019). Evaluating Fairness and Generalizability in Models Predicting On-Time Graduation from College Applications. *International Educational Data Mining Society*.

Inyang, U. G., Umoh, U. A., Nnaemeka, I. C., & Robinson, S. A. (2019). Unsupervised Characterization and Visualization of Students' Academic Performance Features.

Computer and Information Science, *12*(2), 103. https://doi.org/10.5539/cis.v12n2p103

Kamal, P., & Ahuja, S. (2019). An ensemble-based model for prediction of academic performance of students in undergrad professional course. *Journal of Engineering, Design and Technology*, *17*(4), 769–781. https://doi.org/10.1108/JEDT-11-2018-0204

Kasthuriarachchi, K. T. S., & Liyanage, S. R. (2019). Predicting Students' Academic Performance Using Utility Based Educational Data Mining. In *Lecture notes in electrical engineering* (pp. 29–39). Springer Singapore. https://doi.org/10.1007/978-981-13-3648-5_4

Ketui, N., Wisomka, W., & Homjun, K. (2019). Using Classification Data Mining Techniques for Students Performance Prediction. *2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON)*. https://doi.org/10.1109/ecti-ncon.2019.8692227

Kokkinos, P., C. M.and Kiprianos. (2019). *Education policy outlook: Greece*. OECD Publishing.

Kumar, M. (2019). Performance Analysis of Students Using Machine Learning and Data Mining Approach. *International Journal of Engineering and Advanced Technology (IJEAT)*, *8*(3), 75–79. https://doi.org/10.13140/RG.2.2.21634.53444

Kumar, T. R., Vamsidhar, T., Harika, B., Kumar, T. M., & Nissy, R. (2019). Students Performance Prediction Using Data Mining Techniques. *2019 International Conference on Intelligent Sustainable Systems (ICISS)*, 407–411. https://doi.org/10.1109/ISS1.2019.8907945

Kurlaender, M., & Cohen, K. (2019). *Predicting college success: How do different high school assessments measure up?* (Tech. rep. No. 1). Policy Analysis for California Education.

Kyritsi, K. H., Zorkadis, V., Stavropoulos, E. C., & Verykios, V. S. (2019). The pursuit of patterns in educational data mining as a threat to student privacy. *Journal of Interactive Media in Education*, *2019*(1), 1–10. https://doi.org/10.5334/jime.502

Livieris, I. E., Kotsilieris, T., Tampakas, V., & Pintelas, P. (2019). Improving the evaluation process of students' performance utilizing a decision support software. *Neural Computing and Applications*, *31*(6), 1683–1694. https://doi.org/10.1007/s00521-018-3756-y

Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., & Pintelas, P. (2019). Predicting Secondary School Students' Performance Utilizing a Semi-supervised Learning Approach. *Journal of Educational Computing Research*, *57*(2), 448–470. https://doi.org/10.1177/0735633117752614

Livieris, I. E., Tampakas, V., Kiriakidou, N., Mikropoulos, T., & Pintelas, P. (2019). Forecasting Students' Performance Using an Ensemble SSL Algorithm. In *Communications in computer and information science* (pp. 566–581). Springer International Publishing. https://doi.org/10.1007/978-3-030-20954-4_43

M., A., & Salama, S. (2019). Discovering Performance Evaluation Features of faculty Members using Data Mining Techniques to Support Decision Making. *International Journal of Computer Applications*, *178*(49), 25–29. https://doi.org/10.5120/ijca2019919417

Mai, T. L., Do, P. T., Chung, M. T., & Thoai, N. (2019). An Apache Spark-Based Platform for Predicting the Performance of Undergraduate Students. *2019 IEEE 21st International Conference on High Performance Computing and Communications IEEE 17th International Conference on Smart City IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCit*. https://doi.org/10.1109/hpcc/smartcity/dss.2019.00041

Moscoso-Zea, O., Saa, P., & Luján-Mora, S. (2019). Evaluation of algorithms to predict graduation rate in higher education institutions by applying educational data mining. *Australasian Journal of Engineering Education*, 1–10. https://doi.org/10.1080/22054952.2019.1601063

Norman, L. D. (2019). *School culture: A correlational study of school culture and student achievement of central mississippi schools*. Delta State University.

Orben, A., Dienlin, T., & Przybylski, A. K. (2019). Social media's enduring effect on adolescent life satisfaction. *Proceedings of the National Academy of Sciences*, *116*(21), 10226–10228.

Pardos, Z. A., Fan, Z., & Jiang, W. (2019). Connectionist recommendation in the wild: on the utility and scrutability of neural networks for personalized course guidance. *User modeling and user-adapted interaction*, *29*(2), 487–525.

Patrinopoulos, M., & Iatrou, P. (2019). Implementation of STEM tinkering approaches in primary school education in greece. *Sino-US English Teaching*, *16*(12), 510–516.

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, *74*(4), 445.

Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Rodríguez, S., & Corchado, J. M. (2019). Students Performance Analysis Based on Machine Learning Techniques. In *Springer* (pp. 428–438, Vol. 2). https://doi.org/10.1007/978-3-030-20798-4_37

Rodrigues, R. L., Ramos, J. L. C., Silva, J. C. S., Dourado, R. A., & Gomes, A. S. (2019). Forecasting Students' Performance Through Self-Regulated Learning Behavioral

Analysis. *International Journal of Distance Education Technologies*, *17*(3), 52–74. https://doi.org/10.4018/ijdet.2019070104

Rodriguez, L. G., Rodriguez, C. A., Niguidula, J. D., & Calibo, D. I. (2019). Analysis of Students Performance: Input to Program Enhancement of Students in Computing. *International Journal of Information and Education Technology*, *9*(6), 429–432. https://doi.org/10.18178/ijiet.2019.9.6.1240

Santoso, L. W., & Yulia. (2019). The Analysis of Student Performance Using Data Mining. In *Advances in computer communication and computational sciences* (pp. 559–573). https://doi.org/10.1007/978-981-13-6861-5_48

Saravanan, M., & Jyothi, V. L. (2019). A NOVEL APPROACH FOR PERFORMANCE AND ESTIMATION OF STUDENTS BY USING HYBRID MODEL. *IIOAB Journal*, *10*(1), 9–12. www.iioab.org

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational research*, *61*(3), 257–273.

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, *6*(3), 11–19.

Shum, S. J. B., & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. *British journal of educational technology*, *50*(6), 2785–2793.

Siddiq, F., & Scherer, R. (2019). Is there a gender gap? A meta-analysis of the gender differences in students' ICT literacy. *Educational research review*, *27*, 205–217.

Son, L. H., & Fujita, H. (2019). Neural-fuzzy with representative sets for prediction of student performance. *Applied Intelligence*, *49*(1), 172–187. https://doi.org/10.1007/s10489-018-1262-7

Tasnim, N., Kumar Paul, M., & Sarowar Sattar, A. H. M. (2019). Performance Analysis of Different Decision Tree Based Methods for Identifying Drop Out Students. *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 1–6. https://doi.org/10.1109/ICASERT.2019.8934518

Van Hofwegen, L., Eckfield, M., & Wambuguh, O. (2019). Predicting nursing program success for veterans: Examining the importance of TEAS and pre-admit science GPA. *Journal of Professional Nursing*, *35*(3), 209–215.

Widyaningsih, Y., Fitriani, N., & Sarwinda, D. (2019). A Semi-Supervised Learning Approach for Predicting Student ' s Performance : First-Year. *2019 12th International Conference on Information and Communication Technology and System (ICTS)*, 291–295.

Williamson, B., Potter, J., & Eynon, R. (2019). New research problems and agendas in learning, media and technology: the editors' wishlist.

Yang, A., Li, W., & Yang, X. (2019). Short-term electricity load forecasting based on feature selection and least squares support vector machines. *Knowledge-Based Systems*, *163*, 159–173.

Yang, Q., Tian, L., Huebner, E. S., & Zhu, X. (2019). Relations among academic achievement, self-esteem, and subjective well-being in school among elementary school students: A longitudinal mediation model. *School Psychology*, *34*(3), 328–340. https://doi.org/10.1037/spq0000292

Yuan, C., & Yang, H. (2019). Research on k-value selection method of k-means clustering algorithm. *J*, *2*(2), 226–235. https://doi.org/10.3390/j2020016

Zhang, Y., Ren, X., & Zhang, J. (2019). Intrusion detection method based on information gain and relieff feature selection. *2019 International Joint Conference on Neural Networks (IJCNN)*, 1–5.

Adrian, J. A., Bakeman, R., Akshoomoff, N., & Haist, F. (2020). Cognitive functions mediate the effect of preterm birth on mathematics skills in young children. *Child Neuropsychology*, *26*(6), 834–856.

Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: learner choice and fundamental rights. *Learning, Media and Technology*, *45*(3), 312–324.

Binning, K. R., & Browman, A. S. (2020). Theoretical, ethical, and policy considerations for conducting social–psychological interventions to close educational achievement gaps. *Social Issues and Policy Review*, *14*(1), 182–216.

Chamuah, A., & Ghildiyal, H. (2020). AI and Education in India. *Tandem Research*.

Costanza-Chock, S. (2020). *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443.

Dai, D., & Xia, X. (2020). Whether the school self-developed e-learning platform is more conducive to learning during the covid-19 pandemic? *Best Evid Chin Edu*, *5*(1), 569–580.

Demetroulis, E., & Wallace, M. (2020). Educational robotics as a tool for the development of collaboration skills. In S. Papadakis & M. Kalogiannakis (Eds.), *Using educational robotics to facilitate student learning*. IGI Global.

D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT press.

Dixon-Román, E., Nichols, T. P., & Nyame-Mensah, A. (2020). The racializing forces of/in AI educational technologies. *Learning, Media and Technology*, *45*(3), 236–250.

Golden, G. (2020). Education policy evaluation: Surveying the OECD landscape. *OECD Education Working Papers*, (227).

Gubbins, V., & Otero, G. (2020). Parental involvement and low-ses children's academic achievement in early elementary school: New evidence from chile. *Educational Studies*, *46*(5), 548–569.

Henderson, J., & Corry, M. (2020). Data literacy training and use for educational professionals. *Journal of Research in Innovative Teaching & Learning*.

Hooshyar, D., Yang, Y., Pedaste, M., & Huang, Y.-M. (2020). Clustering algorithms in an educational context: An automatic comparative approach. *IEEE Access*, *8*, 146994–147014. https://doi.org/10.1109/ACCESS.2020.3014948

Knox, J. (2020). Artificial intelligence and education in China. *Learning, Media and Technology*, *45*(3), 298–311.

Kuhl, N., Lobana, J., & Meske, C. (2020). Do you comply with AI?–Personalized explanations of learning algorithms and their impact on employees' compliance behavior. *arXiv*, 13–14.

Leithwood, K., Sun, J., & Schumacker, R. (2020). How School Leadership Influences Student Learning: A Test of "The Four Paths Model". *Educational Administration Quarterly*, *56*(4), 570–599. https://doi.org/10.1177/0013161X19878772

Lipnevich, A., Guskey, T., Murano, D., & Smith, J. (2020). What do grades mean? variation in grading criteria in american college and university courses. *Assessment in Education*, *27*(5), 480–500.

Mašić, A., Polz, E., & Becirovic, S. (2020). The relationship between learning styles, GPA, school level and gender. *European Researcher*, *11*(1), 51–60.

McStay, A. (2020). Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy. *Big Data & Society*, *7*(1), 2053951720904386.

Molnar, A., & Boninger, F. (2020). The commercial transformation of America's schools. *Phi Delta Kappan*, *102*(2), 8–13.

Papadimitriou, P., & Tzivinikou, S. (2020). Integration classes in Secondary Education: Critical consideration of the procedures and means of evaluation and intervention adopted. *Panhellenic Conference of Education Sciences*, *9*, 565–578.

Papadogiannis, I., Poulopoulos, V., & Wallace, M. (2020). A Critical Review of Data Mining for Education: What has been done, what has been learnt and what remains to be seen. *International Journal of Educational Research Review*, 353–372. https://doi.org/10.24331/ijere.755047

Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration*, *35*(1), 24–44. https://doi.org/10.1177/0952076718780537

Perrotta, C., & Selwyn, N. (2020). Deep learning goes to school: Toward a relational understanding of AI in education. *Learning, Media and Technology*, *45*(3), 251–269.

Piattoeva, N., & Boden, R. (2020). Escaping numbers? The ambiguities of the governance of education through data.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3). https://doi.org/10.1002/widm.1355

Saa, A. A., Al-Emran, M., & Shaalan, K. (2020). Mining Student Information System Records to Predict Students' Academic Performance. *Advances in Intelligent Systems and Computing*, *921*, 229–239. https://doi.org/10.1007/978-3-030-14118-9_23

Skourtou, E. (2020). Investigating Literacy Issues on Roma Education. In *Language diversity in greece* (pp. 157–167). Springer.

Spyropoulou, C., Wallace, M., Vassilakis, C., & Poulopoulos, V. (2020). Examining the use of STEAM education in preschool education. *European Journal of Engineering and Technology Research*. https://doi.org/https://doi.org/10.24018/ejeng.2020.0.CIE.2309□

Valle-Cruz, D., Criado, J. I., Sandoval-Almazán, R., & Ruvalcaba-Gomez, E. A. (2020). Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, *37*(4), 101509.

Vonneilich, N., Lüdecke, D., & von dem Knesebeck, O. (2020). Educational inequalities in self-rated health and social relationships – analyses based on the European Social Survey 2002-2016. *Social Science and Medicine*, *267*, 112379. https://doi.org/10.1016/j.socscimed.2019.112379

Warner, J. W. E. (2020). Coyote, the hero of ungrading. In S. D. Blum (Ed.), *Ungrading: Why rating students undermines learning (and what to do instead)* (pp. 204–218). West Virginia University Press.

Watts, T. W. (2020). Academic achievement and economic attainment: Reexamining associations between test scores and long-run earnings. *AERA Open*, *6*(2), 2332858420928985. https://doi.org/10.1177/2332858420928985

Webb, P. T., Sellar, S., & Gulson, K. N. (2020). Anticipating education: Governing habits, memories and policy-futures. *Learning, Media and Technology*, *45*(3), 284–297.

Yoon, E.-S. (2020). School choice research and politics with Pierre Bourdieu: New possibilities. *Educational Policy*, *34*(1), 193–210.

Allen, L. K., Creer, S. D., & Poulos, M. C. (2021). Natural language processing as a technique for conducting text-based research. *Language and Linguistics Compass*, *15*(7), e12433.

Averchenkova, A., Fankhauser, S., & Finnegan, J. J. (2021). The influence of climate change advisory bodies on political debates: Evidence from the uk committee on climate change. *Climate Policy*, *21*(9), 1218–1233.

Berkovich, I. (2021). The Policy Process: The Design of Education Policy. In *Education policy, theories, and trends in the 21st century* (pp. 21–39). Springer.

Chaidi, E., Kefalis, C., Papagerasimou, Y., & Drigas, A. (2021). Educational robotics in primary education. a case in greece. *Research, Society and Development*, *10*(9), e17110916371. https://doi.org/10.33448/rsd-v10i9.16371

Dong, Y., Yin, H., Du, S., & Wang, A. (2021). The effects of flipped classroom characterized by situational and collaborative learning in a community nursing course: A quasi-experimental design. *Nurse Education Today*, *105*, 105037.

Easterbrook, M. J., & Hadden, I. R. (2021). Tackling Educational Inequalities with Social Psychology: Identities, Contexts, and Interventions. *Social Issues and Policy Review*, *15*(1), 180–236. https://doi.org/10.1111/sipr.12070

Eriksson, K., Lindvall, J., Helenius, O., & Ryve, A. (2021). Socioeconomic status as a multidimensional predictor of student achievement in 77 societies. *Frontiers in Education*, *6*, 464.

Eynon, R., & Young, E. (2021). Methodology, legend, and rhetoric: The constructions of AI by academia, industry, and policy groups for lifelong learning. *Science, Technology, and Human Values*, *46*(1), 166–191.

Fateel, M. J., Mukallid, S., & Arora, B. (2021). The interaction between socioeconomic status and preschool education on academic achievement of elementary school students. *International Education Studies*, *14*(8), 60–66.

Hermassi, S., Chelly, M. S., Michalsik, L. B., Sanal, N. E. M., Hayes, L. D., & Cadenas-Sanchez, C. (2021). Relationship between fatness, physical fitness, and academic performance in normal weight and overweight schoolchild handball players in qatar state. *Plos One*, *16*(2), e0246476.

Karountzou, G., Papadogiannis, I., & Wallace, M. (2021). Achieving Intercultural Education and Language Awareness and Intercultural Communication In School of Inclusion Using ICT. *IAIE 2021 Intercultural Conference*, June 27–30, 2021.

Longobardi, C., Settanni, M., Lin, S., & Fabris, M. A. (2021). Student–teacher relationship quality and prosocial behaviour: The mediating role of academic achievement and a positive attitude towards school. *British Journal of Educational Psychology*, *91*(2), 547–562.

Martin, A. J., Collie, R. J., & Nagy, R. P. (2021). Adaptability and high school students' online learning during covid-19: A job demands-resources perspective. *Frontiers in Psychology*, *12*, 702163.

Mehmetaj, N., & Alili, M. Z. (2021). Employment of economics graduates: Does gpa matter? *Interdisciplinary Description of Complex Systems: INDECS*, *19*(2), 210–226.

Mejía-Rodríguez, Luyten, H., & Meelissen, M. R. M. (2021). Gender differences in mathematics self-concept across the world: An exploration of student and parent data of TIMSS 2015. *International Journal of Science and Mathematics Education*, *19*(6), 1229–1250.

Miller, M. T. (2021). Do learning organizations learn? higher education institutions and pandemic response strategies. *The Learning Organization*.

Papadogiannis, I., Wallace, M., Poulopoulos, V., Karountzou, G., & Ekonomopoulos, D. (2021a). A First Ever Look into Greece's Vast Educational Data: Interesting Findings and Policy Implications. *Educ. Sci.*, *11*(9), 489. https://doi.org/10.3390/educsci11090489

Papadogiannis, I., Wallace, M., Poulopoulos, V., Karountzou, G., & Ekonomopoulos, D. (2021b). A first ever look into greece's vast educational data: Interesting findings and policy implications. *Educ. Sci.*, *11*, 489. https://doi.org/10.3390/educsci11090489

Papadogiannis, I., Wallace, M., Poulopoulos, V., Karountzou, G., & Ekonomopoulos, D. (2021c). A First Ever Look into Greece's Vast Educational Data: Interesting Findings and Policy Implications. *Education Sciences*, *11*(9), 489. https://doi.org/10.3390/educsci11090489

Paulus, L., Spinath, F. M., & Hahn, E. (2021). How do educational inequalities develop? The role of socioeconomic status, cognitive ability, home environment, and self-efficacy along the educational path. *Intelligence*, *86*, 101528. https://doi.org/10.1016/j.intell.2021.101528

Verostek, M., Miller, C. W., & Zwickl, B. (2021). Analyzing admissions metrics as predictors of graduate gpa and whether graduate gpa mediates ph.d. completion. *Physical Review Physics Education Research*, *17*(2).

Zhu, X., Dai, J., Wei, H., Yang, D., Huang, W., & Yu, Z. (2021). Application of the fuzzy optimal model in the selection of the startup hub. *discrete dynamics in nature and society*, *2021*.

Bouiri, O., Lotfi, S., & Talbi, M. (2022). Effect of Student, Family-Related Features on Academic Achievement. *Education Research International*, *2022*.

Escaño, C., Mesías-Lema, J. M., & Mañero, J. (2022). Empowerment of the refugee migrant community through a cooperation project on art education in Greece. *Development in Practice*, *32*(7), 912–927. https://doi.org/10.1080/09614524.2021.1944986

European_Union. (2022, May). *What is gdpr, the eu's new data protection law?* [https://gdpr.eu/what-is-gdpr/%20%E2%86%97](https://gdpr.eu/what-is-gdpr/)

Gambini, A., Desimoni, M., & Ferretti, F. (2022). Predictive tools for university performance: An explorative study. *International Journal of Mathematical Education in Science and Technology*, 1–27.

Irwin, V., De La Rosa, J., Wang, K., Hein, S., Zhang, J., Burr, R., Roberts, A., Barmer, A., Bullock Mann, F., Dilig, R., et al. (2022). Report on the condition of education 2022. nces 2022-144. *National Center for Education Statistics*.

Karacabey, M. F., Bellibaş, M. Ş., & Adams, D. (2022). Principal leadership and teacher professional learning in turkish schools: Examining the mediating effects of collective teacher efficacy and teacher trust. *Educational studies*, *48*(2), 253–272.

Mubarok, H., Dinangsit, D., & Lengkana, A. S. (2022). The relationship of self esteem and physical fitness to learning achievement in jabal toriq boarding school students. *JUARA: Jurnal Olahraga*, *7*(3), 512–525.

National Assessment for educational Statistics. (2022). How is Grade Point Average Calculated? https://nces.ed.gov/nationsreportcard/hsts/howgpa.aspx

Papadogiannis, I., Wallace, M., & Poulopoulos, V. (2022). Examining Pupils' Achievement in Primary and Secondary Schools in Greece. *European Journal of Engineering and Technology Research*, 10–18. https://doi.org/10.24018/ejeng.2022.1.cie.2958

Pearson. (2022). Tutoring apps to get affordable, personalized study help. https://www.pearson.com/en-us/tutoring-apps.html

Rajbhandari-Thapa, J., Metzger, I., Ingels, J., Thapa, K., & Chiang, K. (2022). School climate-related determinants of physical activity among high school girls and boys. *Journal of adolescence*, *94*(4), 642–655.

Stamatopoulou, G. (2022). *Intergenerational mobility trends in greece and a comparative study amongst southern european countries* [Doctoral dissertation, Panteion University].

Villa Alvarez, D. P., Auricchio, V., & Mortati, M. (2022). Mapping design activities and methods of public sector innovation units through the policy cycle model. *Policy Sciences*, *55*(1), 89–136. https://doi.org/10.1007/s11077-022-09448-4

von der Embse, N. P., Kilgus, S. P., Eklund, K., Zahn, M., Peet, C., & Durango, S. (2022). Promoting effective decision making: Training educators to collect and use social–emotional skill assessment data to inform tier 2 interventions. *School Psychology Review*, *51*(5), 574–588.

Bellei, C., & Munoz, G. (2023). Models of regulation, education policies, and changes in the education system: a long-term analysis of the Chilean case. *J Educ Change*, 49–76. https://doi.org/10.1007/s10833-021-09435-1

Pisinger, V. S. C., Andersen, S., & Tolstrup, J. S. (2023). Perceived parental alcohol problems and later dropout and grade point average in high school: A register-based follow-up study. *Drug and Alcohol Review*.

Safarov, F., Kutlimuratov, A., Abdusalomov, A. B., Nasimov, R., & Cho, Y.-I. (2023). Deep learning recommendations of e-education based on clustering and sequence. *Electronics*, *12*(4), 809. https://doi.org/10.3390/electronics12040809

Wallace, M., Poulopoulos, V., Papavramidou, Z., Tzioras, I., & Liaperdos, J. (n.d.). ANTS: From History of Science to Future of Science. *8th International Euro-Mediterranean Conference (EuroMed)*, 2–5 November 2020, Online. Version September 1, 2023 submitted to Knowledge 15.