UNIVERSITY OF THESSALY

School of Economics and Administrative Sciences

Department of Business Administration

Title

Evaluating and Enhancing Academic Excellence: A Competency-based Model for Quality of Teaching

Charalampos Dervenis

Larissa, February 2025

Faculty advisor

Professor Dr. Panagiotis Fitsilis

Doctoral Dissertation submitted to the faculty in fulfilment of the requirements for the doctoral degree of the Doctoral Program in Business Administration of the University of Thessaly.

Approved by the Faculty Committee:

1. Professor Dr. Panagiotis Fitsilis, University of Thessaly, Faculty advisor

2. Professor Dr. Athanasios Koustelios, University of Thessaly

3. Assistant Professor Dr. Omiros Iatrellis, University of Thessaly

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Examination Committee:

Professor Dr. Panagiotis Fitsilis, University of Thessaly

Professor Dr. Athanasios Koustelios, University of Thessaly

Assistant Professor Dr. Omiros Iatrellis, University of Thessaly

Professor Dr. Leonidas Anthopoulos, University of Thessaly

Associate Professor Dr. Dimitris Tselios, University of Thessaly

Assistant Professor Dr. Alexandra Bousia, University of Thessaly

Professor Dr. Achilleas Kameas, Hellenic Open University

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

"The approval of the Doctoral Dissertation by the Department of Business Administration of the University of Thessaly does not imply acceptance of the author's opinions (Law 5343/1932, Article 202, Paragraph 2)."

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

"With my personal responsibility and being fully aware of the possibility of sanctions imposed by the Department Assembly, I hereby declare that, in the writing of my Doctoral Dissertation, I have not used, in whole or in part, the work of another creator, or the ideas and perceptions of another creator, without citing the source of origin (book, newspaper or journal article, website, etc.)."

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Acknowledgement

The completion of this dissertation would not have been possible without the contribution and unwavering support of many individuals, each of whom stood by me in their own way, and to whom I owe my deepest gratitude.

First and foremost, I would like to express my heartfelt thanks to my supervisor, Professor Dr. Panagiotis Fitsilis, for his invaluable guidance, scientific support, and the trust he placed in me throughout the course of my research. He was always there for me with continuous support, a tireless willingness for scientific dialogue, and constant encouragement. Through his insightful interventions, he managed to shed light on every dark corner of this journey. This guidance has been pivotal, not only for my academic but also for my personal development. At the same time, I would like to thank the other members of my advisory committee, Professor Dr. Athanasios Koustelios and Assistant Professor Dr. Omiros Iatrellis, for the time they dedicated and for their constructive comments and suggestions throughout my dissertation.

I would also like to extend my heartfelt gratitude to my wife, Evangelia, and my children, Thanos and Giorgos, who wholeheartedly accepted and supported my absence from the onset so that this goal could be achieved! Countless family moments were sacrificed… even then, they stood by me, encouraging me and offering me their unwavering support throughout this long journey that at times seemed to have no end.

I would also like to express my warmest thanks to Vasilis Kyriatzis, first as a friend and then as a colleague, who, with his support, assistance, and unique sense of humor, pushed me toward my goal. Finally, I would like to express my sincere gratitude to Eleftheria Skretta, who stood by my side from day one, offering encouragement until the completion of this work.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table of Contents

List of Figures 9
List of Tables 11
Abstract 13
Keywords 14
Περίληψη 15
1. Introduction 19
1.1. Introduction to the subject 19
1.2. Research problem 20
1.3. Research aims and objectives 21
1.4. Contribution 22
1.5. Thesis outline 24
1.6. Summary 25
2. Literature review 26
2.1. Part 1: Teacher competencies model 26
2.1.1. Introduction 26
2.1.2. Methodology 27
2.1.3. A multi-dimensional approach of teacher competencies in higher education. 30
2.1.3.1. Competencies per dimension 38
2.1.3.2. The influence of dimensions in the education process 49
2.1.4. Limitations 51
2.1.5. Practical and social implications 52
2.1.6. Concluding remarks 52
2.2. Part 2: Assessing teacher competencies through student feedback 53
2.2.1. Introduction 53
2.2.2. The value of student feedback in teaching progress 54
2.2.3. Educational data analysis 55
2.2.3.1. Methods and approaches for analyzing educational data 56
2.2.4. SA in education 57
2.2.5. Concluding remarks 60
3. Research Design and Methodology 62
3.1. Introduction 62
3.2. Research philosophy 63
3.2.1. The nature of reality 63
3.2.2. Epistemology 64
3.2.3. Axiology 65

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

3.2.4. Positivism vs. Interpretivism 66
3.2.5. Pragmatism 67
3.3. Research design and methodology 69
3.3.1. Detailed framework of the initial research process: From objectives to identifying gaps (phase 1) 69
3.3.1.1. Defining research objectives 71
3.3.1.2. Gathering previous research 71
3.3.1.3. Analysing and synthesizing previous research 71
3.3.2. Methodological framework for analysing educational text data: Approaches and techniques (phase 2) 72
3.3.2.1. Investigating and analysing educational data 73
3.3.2.2. Apply text data analysis techniques 74
3.3.2.3. Compare different approaches 74
3.3.3. Model development and validation: Methodological framework (phase 3) 74
3.3.3.1. Model creation and implementation 76
3.3.3.2. Development of a validation framework 76
3.3.3.3. Extraction and analysis of results 78
3.3.3.4. Final presentation and conclusions 79
3.4. Methods and tools 79
3.4.1. Methods and tools used for the literature review 80
3.4.2. Data collection methods 81
3.4.3. SA method 83
3.4.3.1. Algorithms and techniques used in SA 84
3.4.4. Topic modeling method 85
3.5. Summary 86
4. Implementation and Results 88
4.1. Introduction 88
4.2. Implementation 88
4.2.1. Study 1: Predicting student performance using ML algorithms 89
4.2.1.1. Data collection 89
4.2.1.2. Model design 91
4.2.1.3. Testing algorithms and classification 92
4.2.1.4. Model evaluation 93
4.2.1.5. Results and discussion 95
4.2.1.6. Conclusions 98
4.2.2. Study 2: Assessing teacher competencies in higher education: A SA of student feedback 98

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

4.2.2.1. SA approach 98
4.2.2.2. Method 101
4.2.2.3. Implementation 104
4.2.2.4. Results and discussion 107
4.2.2.5. Limitations 111
4.2.2.6. Conclusions 111
4.2.3. Study 3: Chat GPT vs. Typical data mining process: A comparative case study on teaching competencies via SA 112
4.2.3.1. ChatGPT and traditional lexicon-based approach 112
4.2.3.2. Method 114
4.2.3.3. Results and discussion 118
4.2.3.4. Limitations 126
4.2.3.5. Conclusions 127
4.2.4. Study 4: Comparative SA using lexicon and ML techniques 127
4.2.4.1. Method 127
4.2.4.2. Implementation 134
4.2.4.3. Results and discussion 138
4.2.4.4. Conclusions 142
4.3. Validation 143
4.3.1. Method 144
4.3.2. Results and discussion 146
5. Discussion, Conclusions and Recommendations 160
5.1. Introduction 160
5.2. Conclusions 160
5.3. Limitations and implications 162
5.4. Recommendations and future research 164
References 166
Appendix A – Publications 181
Appendix B – Generation of a 3x3 Confusion Matrix using Python Code 185
Appendix C – BERTopic Implementation Using Python in Google Colab 190
Appendix D – Turnitin report 193

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

List of Figures

Figure 1. Relevant bibliography search process flowchart based on PRISMA statement (source: Authors' own work). 29
Figure 2. The study process of the 39 eligible articles (source: Authors' own work). 30
Figure 3. Percentage appearance Apr(x) and coverage level Cvr(x) of dimensions in the eligible papers (source: Authors' own work). 49
Figure 4. Pragmatic research framework (source: Authors' own work). 63
Figure 5. Knowledge acquisition in research (source: Authors' own work). 65
Figure 6. Values in educational research (source: Authors' own work). 66
Figure 7. Positivism vs. Interpretivism (source: Authors' own work). 67
Figure 8. Pragmatic research framework (source: Authors' own work). 68
Figure 9. Framework for initial research process: from objectives to identifying gaps (phase 1) (source: Authors' own work). 70
Figure 10. Methodological framework for analyzing educational text data (phase 2) (source: Authors' own work). 73
Figure 11. Model development and validation (phase 3) (source: Authors' own work). 75
Figure 12. Model validation (source: Authors' own work). 77
Figure 13. PRISMA statement flowchart (Page et al., 2021). 81
Figure 14. Predicting student performance using machine learning algorithms: The procedure followed (source: Authors' own work). 89
Figure 15. Model designed (source: Authors' own work). 92
Figure 16. K-fold cross-validation with k = 10 (source: Authors' own work). 93
Figure 17. KNN, SVM and Random forest confusion matrix (2-class implementation) (source: Authors' own work). 96
Figure 18. KNN, SVM and Random Forest confusion matrix (5-class implementation) (source: Authors' own work). 97
Figure 19. Text preprocessing phase (Stage 2) (source: Authors' own work). 102
Figure 20. Workflow of model designed for SA (source: Authors' own work). 106
Figure 21. Sample comments from students with their corresponding scores (source: Authors' own work). 107
Figure 22. Sentiment intensity of student comments for prof_1 in course_1, prof_2 in course_1 and prof_2 in course_2 (source: Authors' own work). 108
Figure 23. Mean sentiment score of student comments received, regarding the competencies of 2 different instructors of the same course (course_1) (source: Authors' own work). 109
Figure 24. Data clustering using K-means algorithm (source: Authors' own work). 110

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Figure 25. Workflow of model designed for SA (source: Authors' own work). 115
Figure 26. Text preprocessing phase (stage 2) (source: Authors' own work). 116
Figure 27. Sample comments from students with their corresponding scores (source: Authors' own work). 117
Figure 28. The steps in our engagement with ChatGPT (source: Authors' own work). 118
Figure 29. Sentiment score of student comments about a teacher of a particular course (source: Authors' own work). 119
Figure 30. The mean sentiment score of student comments received, regarding their teacher's competencies of a specific course (source: Authors' own work). 120
Figure 31. The overall procedure followed. 128
Figure 32. Lexicon-based approach (source: Authors' own work). 135
Figure 33. Sentiment score of student feedback comments (source: Authors' own work). 136
Figure 34. Random forest based approach (source: Authors' own work). 137
Figure 35. The overall procedure followed (source: Authors' own work). 145
Figure 36. Topic coherence across different k values (source: Authors' own work). 147

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

List of Tables

Table 1. A multi-dimensional approach of teacher competencies in HE (source: Authors' own work). 34
Table 2. The list of teaching staff competencies in HE and the papers in which they were addressed (source: Authors' own work). 41
Table 3. Influence of each competency dimension (source: Authors' own work). 50
Table 4. Example words from VADER lexicon (valence-based) (source: Authors' own work). 59
Table 5. Contribution of special characteristics to the intensity of polarity (source: Authors' own work). 60
Table 6. Model variables and description. 90
Table 7. Confusion matrix. 93
Table 8. Classification evaluation metrics. 94
Table 9. Evaluation results (2-class implementation) (source: Authors' own work). 95
Table 10. Categories of students' grades (for Math) (source: Authors' own work). 96
Table 11. Evaluation results (5-class implementation) (source: Authors' own work). 97
Table 12. Text data challenges (source: Authors' own work). 103
Table 13. Teacher competencies that refer to the Communication dimension with the corresponding open-ended questions asked to students (source: Authors' own work). 104
Table 14. Challenges faced; ChatGPT and traditional lexicon-based method comparison (source: Authors' own work). 113
Table 15. Teacher competencies related to the "Communication" dimension, along with the corresponding open-ended questions posed to students (source: Authors' own work). 114
Table 16. The results received from ChatGPT based on student feedback regarding each teacher competence (source: Authors' own work). 121
Table 17. The score assigned by ChatGPT regarding each teacher competence (source: Authors' own work). 124
Table 18. The score assigned by the sentiment analysis model, converted to a scale from 0 to 100 (source: Authors' own work). 125
Table 19. The distribution of samples across the sentiment classes (source: Authors' own work). 129
Table 20. Preprocessing steps applied in lexicon-based algorithm (source: Authors' own work). 130
Table 21. Confusion matrix for binary classification (source: Authors' own work). 132
Table 22. Metrics for evaluating classification performance (source: Authors' own work). 133
Table 23. ANN hyperparameters (source: Authors' own work). 137
Table 24. Evaluation results over all classes (source: Authors' own work). 139

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 25. Evaluation results by class (source: Authors’ own work)..139
Table 26. Lexicon confusion matrix (source: Authors’ own work). 139
Table 27. RF confusion matrix (source: Authors’ own work). 140
Table 28. ANN confusion matrix (source: Authors’ own work). 140
Table 29. Pros and Cons of each method examined in our study (source: Authors’ own work). ..141
Table 30. Identified topics in student feedback through LDA (source: Authors’ own work). 148
Table 31. Representative student comments for each topic (source: Authors’ own work). 150
Table 32. Mapping results: Topic modeling (LDA) insights vs. literature-derived competency framework (source: Authors’ own work). 153
Table 33. Identified topics in student feedback through BERTopic (source: Authors’ own work).154
Table 34. Mapping results: Topic modeling (BERTopic) insights vs. literature-derived competency framework (source: Authors’ own work). 156
Table 35. Comparison of teacher competencies identified through LDA and BERTopic with the Literature-based framework (source: Authors’ own work)...157

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Abstract

In contemporary educational contexts, the importance of academic excellence and the essential skills of teachers is becoming increasingly critical. The demand for continuous professional development has become paramount, as educators must constantly adapt to new technologies, teaching methods, and student needs. Consequently, understanding which competencies are critical for effective teaching and how they can be cultivated remains a major issue in modern educational systems. However, a key challenge facing academic institutions is the lack of a clear and comprehensive framework for defining and assessing teacher competencies in Higher Education (HE). Many studies have attempted to identify a competency model by examining both the conceptual framework exploring educational concepts as well as learning theories focusing on teaching staff roles and competencies in higher education. However, many of these studies remain purely theoretical, while others concentrate on narrow categories, failing to provide a comprehensive and practical competency framework tailored to academics. Without such a framework, educational institutions struggle to establish consistent and reliable methods for developing and assessing their teaching staff.

At the same time, a gap remains in how specific teaching competencies can be effectively identified and measured. Traditional methods for evaluating teacher performance, such as student surveys and peer assessments, often rely on standardized, quantitative data, which may overlook nuanced insights embedded in open-ended student feedback. Furthermore, the rapid advancement in Machine Learning (ML) techniques and Sentiment Analysis (SA) offers new opportunities for exploring educational data in more sophisticated ways. However, there is a lack of research investigating how these methods can be utilized for assessing teacher competencies and how they compare to traditional techniques. Consequently, despite the acknowledged significance of teaching skills in achieving high-quality education, the absence of internationally recognized standards, along with the limited use of innovative data analysis techniques—including methods employing ML algorithms—restricts the meaningful evaluation and development of teaching practices.

These gaps—both in defining a clear framework for teacher competencies and in measuring them—represent a key area that this research aims to address. By establishing a detailed framework for teacher competencies and exploring innovative data analysis techniques, this study will provide valuable insights into how student feedback can be leveraged to assess and improve teaching quality.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

This research aims to address these challenges by identifying, evaluating, and developing a framework of competencies tailored to higher education, with the ultimate goal of improving teaching quality. More specifically, the initial aim of this study is to do extensively assemble, analyze and synthesize previous research in order to investigate and identify teaching staff competencies derived from the roles and tasks attributed to university professors. Through an original in-depth review, an attempt is made to identify and classify the teaching staff competencies in higher education, revealing the gap between the perceived and actual importance of various competencies. Subsequently, by employing SA and ML methods, this study analyzes qualitative data from student feedback to gain deeper insights into their perceptions of teacher competencies.

The findings of this research advocate for the integration of these innovative data analysis methods into teacher evaluation and development programs, offering essential tools for enhancing teaching quality. These findings are compared to the competency framework that emerged from our previous research, providing a unique opportunity to validate and expand our understanding of what defines effective teaching in higher education.

Through our aims and objectives, this research seeks not only to define the competencies necessary for effective teaching but also to explore innovative techniques for measuring them using student feedback data. By bridging the gap in defining educator competencies and developing methods for their assessment, this research aims to offer valuable insights that will improve both teaching quality and student learning experiences in higher education. Finally, the results underscore the importance of continuously adapting teaching skills to meet the demands of technological and educational advancements, ensuring that teachers are well-equipped to address the contemporary challenges in higher education.

Keywords

Teacher quality; Teacher competencies; Student feedback analysis; Open-ended data; Topic modeling in education; Sentiment analysis in education; Machine learning; Teaching evaluation; Learning analytics; Data mining

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Περίληψη

Στα σύγχρονα εκπαιδευτικά πλαίσια, η σημασία της ακαδημαϊκής αριστείας και των βασικών δεξιοτήτων των εκπαιδευτικών γίνεται ολοένα και πιο κρίσιμη. Η απαίτηση για συνεχή επαγγελματική εξέλιξη έχει καταστεί πρωταρχική, καθώς οι εκπαιδευτικοί πρέπει να προσαρμόζονται συνεχώς στις νέες τεχνολογίες, τις μεθόδους διδασκαλίας και τις ανάγκες των φοιτητών. Κατά συνέπεια, η κατανόηση των ικανοτήτων που είναι κρίσιμες για την αποτελεσματική διδασκαλία και του τρόπου με τον οποίο μπορούν να καλλιεργηθούν παραμένει μείζον ζήτημα στα σύγχρονα εκπαιδευτικά συστήματα. Ωστόσο, μια βασική πρόκληση που αντιμετωπίζουν τα ακαδημαϊκά ιδρύματα είναι η έλλειψη ενός σαφούς και ολοκληρωμένου πλαισίου για τον καθορισμό και την αξιολόγηση των ικανοτήτων των εκπαιδευτικών στην Τριτοβάθμια Εκπαίδευση (ΑΕ). Πολλές μελέτες έχουν προσπαθήσει να προσδιορίσουν ένα μοντέλο ικανοτήτων των εκπαιδευτικών, εξετάζοντας τόσο το εννοιολογικό πλαίσιο που διερευνά τις εκπαιδευτικές έννοιες όσο και τις θεωρίες μάθησης που εστιάζουν στους ρόλους και τις ικανότητες του διδακτικού προσωπικού στην τριτοβάθμια εκπαίδευση. Από αυτές τις μελέτες άλλες παραμένουν καθαρά θεωρητικές και άλλες επικεντρώνονται σε στενές κατηγορίες, αποτυγχάνοντας έτσι να παράσχουν ένα ολοκληρωμένο και πρακτικό πλαίσιο ικανοτήτων προσαρμοσμένο στους ακαδημαϊκούς. Η απουσία ενός τέτοιου πλαισίου αναγκάζει τα εκπαιδευτικά ιδρύματα να αγωνίζονται να καθιερώσουν συνεπείς και αξιόπιστες μεθόδους για την ανάπτυξη και την αξιολόγηση του διδακτικού τους προσωπικού.

Ταυτόχρονα, παραμένει ένα κενό στον τρόπο με τον οποίο μπορούν να εντοπιστούν και να μετρηθούν αποτελεσματικά συγκεκριμένες εκπαιδευτικές ικανότητες. Οι παραδοσιακές μέθοδοι για την αξιολόγηση της απόδοσης των εκπαιδευτικών, όπως οι έρευνες φοιτητών και οι αξιολογήσεις από ομότιμους, βασίζονται συχνά σε τυποποιημένα, ποσοτικά δεδομένα, τα οποία ενδέχεται να παραβλέπουν διαφοροποιημένες ιδέες που ενσωματώνονται στα σχόλια ανοιχτού τύπου ερωτήσεων των φοιτητών. Από την άλλη, η ταχεία πρόοδος στις τεχνικές Μηχανικής Μάθησης (ΜΜ) και στην Ανάλυση Συναισθήματος (ΑΣ) προσφέρει νέες ευκαιρίες για την εξερεύνηση των εκπαιδευτικών δεδομένων με πιο εξελιγμένους τρόπους. Ωστόσο, υπάρχει έλλειψη έρευνας που να διερευνά πώς αυτές οι μέθοδοι μπορούν να χρησιμοποιηθούν για την αξιολόγηση των ικανοτήτων των εκπαιδευτικών και πώς συγκρίνονται με τις παραδοσιακές τεχνικές. Κατά συνέπεια, παρά την αναγνωρισμένη σημασία των διδακτικών δεξιοτήτων για την επίτευξη εκπαίδευσης υψηλής ποιότητας, η απουσία διεθνώς αναγνωρισμένων προτύπων, μαζί με την περιορισμένη χρήση καινοτόμων τεχνικών ανάλυσης δεδομένων - συμπεριλαμβανομένων μεθόδων που χρησιμοποιούν αλγόριθμους ML - περιορίζουν τη ουσιαστική αξιολόγηση και ανάπτυξη διδακτικών πρακτικών.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Αυτά τα κενά -τόσο στον καθορισμό ενός σαφούς πλαισίου για τις ικανότητες των εκπαιδευτικών όσο και στη μέτρησή τους- αντιπροσωπεύουν έναν βασικό τομέα που στοχεύει να αντιμετωπίσει αυτή η έρευνα. Με τη δημιουργία ενός λεπτομερούς πλαισίου για τις ικανότητες των εκπαιδευτικών και τη διερεύνηση καινοτόμων τεχνικών ανάλυσης δεδομένων, αυτή η μελέτη παρέχει πολύτιμες γνώσεις για το πώς μπορεί να αξιοποιηθεί η ανατροφοδότηση των φοιτητών για την αξιολόγηση και τη βελτίωση της ποιότητας διδασκαλίας.

Η παρούσα έρευνα στοχεύει να αντιμετωπίσει αυτές τις προκλήσεις εντοπίζοντας, αξιολογώντας και αναπτύσσοντας ένα πλαίσιο ικανοτήτων προσαρμοσμένο στην τριτοβάθμια εκπαίδευση, με απώτερο στόχο τη βελτίωση της ποιότητας της διδασκαλίας. Πιο συγκεκριμένα, ο αρχικός στόχος αυτής της μελέτης είναι η εκτενής συγκέντρωση, ανάλυση και σύνθεση προηγούμενων ερευνών με σκοπό τη διερεύνηση και τον προσδιορισμό των ικανοτήτων του διδακτικού προσωπικού που απορρέουν από τους ρόλους και τα καθήκοντα που αποδίδονται σε καθηγητές πανεπιστημίου. Μέσα από μια πρωτότυπη σε βάθος ανασκόπηση, επιχειρείται να εντοπιστούν και να ταξινομηθούν οι ικανότητες του διδακτικού προσωπικού στην τριτοβάθμια εκπαίδευση, αποκαλύπτοντας το χάσμα μεταξύ της αντιληπτής και της πραγματικής σημασίας των διαφόρων ικανοτήτων. Στη συνέχεια, με τη χρήση μεθόδων ΑΣ και ΜΜ και την ανάλυση ποιοτικών δεδομένων από την ανατροφοδότηση των φοιτητών αποκτώνται βαθύτερες γνώσεις σχετικά με τις αντιλήψεις των εκπαιδευομένων για τις ικανότητες των εκπαιδευτικών τους.

Τα ευρήματα αυτής της έρευνας συνηγορούν υπέρ της ενσωμάτωσης αυτών των καινοτόμων μεθόδων ανάλυσης δεδομένων σε προγράμματα αξιολόγησης και ανάπτυξης εκπαιδευτικών, προσφέροντας βασικά εργαλεία για τη βελτίωση της ποιότητας της διδασκαλίας. Αυτά τα ευρήματα συγκρίνονται με το πλαίσιο ικανοτήτων που προέκυψε από την αρχική έρευνά μας που στηρίχθηκε στην βιβλιογραφία, παρέχοντας μια μοναδική ευκαιρία να επικυρώσουμε και να διευρύνουμε την κατανόησή μας για το τι ορίζει την αποτελεσματική διδασκαλία στην τριτοβάθμια εκπαίδευση.

Πιο συγκεκριμένα, η παρούσα έρευνα εστιάζοντας στον ορισμό, την αξιολόγηση και την ενίσχυση των ικανοτήτων των εκπαιδευτικών συμβάλλει πολύπλευρα στον χώρο της τριτοβάθμιας εκπαίδευσης. Με τη σύνθεση θεωρητικών πλαισίων, καινοτόμων μεθοδολογιών και πρακτικών εφαρμογών, προάγει την κατανόηση για τις αποτελεσματικές διδακτικές πρακτικές και τον ρόλο τους στην προώθηση της συμμετοχής και της ικανοποίησης των μαθητών στην εκπαιδευτική διαδικασία. Οι βασικές συνεισφορές αυτής της διατριβής περιγράφονται ως εξής:

Ανάπτυξη πλαισίου ικανοτήτων εκπαιδευτικού: Σημαντική συμβολή αυτής της έρευνας είναι η διαμόρφωση ενός ολοκληρωμένου πλαισίου ικανοτήτων εκπαιδευτικών στην τριτοβάθμια εκπαίδευση, όπως δημοσιεύτηκε στην εργασία μας με τίτλο «Ανασκόπηση της έρευνας για τις

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

ικανότητες των εκπαιδευτικών στην τριτοβάθμια εκπαίδευση» (Dervenis et al., 2022). Αυτό το πλαίσιο χρησιμεύει ως βασικό σημείο αναφοράς για τη διασφάλιση της ποιότητας στην πρόσληψη, την επαγγελματική ανάπτυξη και την αξιολόγηση του διδακτικού προσωπικού. Πέρα από την πρακτική του αξία, το πλαίσιο διευκολύνει τον ουσιαστικό διάλογο μεταξύ των ενδιαφερομένων, συμπεριλαμβανομένων των υπευθύνων χάραξης πολιτικής, των εκπαιδευτικών, των ερευνητών και των διευθυντών εκπαίδευσης, ενισχύοντας μια συλλογική προσέγγιση για τη Διασφάλιση Ποιότητας (ΔΠ) και την εκπαιδευτική βελτίωση. Δίνοντας έμφαση στον εξέχοντα ρόλο αυτών των ικανοτήτων όχι μόνο στη διαδικασία διδασκαλίας-μάθησης αλλά και στα εργασιακά και στα ευρύτερα κοινωνικά περιβάλλοντα, το πλαίσιο υπογραμμίζει τον ουσιαστικό ρόλο τους στην αποτελεσματική απόκριση στις προκλήσεις και στην επίτευξη βέλτιστων αποτελεσμάτων.

Καινοτόμος εφαρμογή της ΜΜ στην εκπαίδευση: Αυτή η διατριβή γεφυρώνει το χάσμα μεταξύ των παραδοσιακών μεθόδων εκπαιδευτικής ανάλυσης και των αναδυόμενων τεχνικών ΜΜ, όπως δημοσιεύτηκε σε μία από τις μελέτες μας με τίτλο "Sentiment analysis of student feedback: A comparative studying employing lexicon and machine learning techniques" (Dervenis et al., 2024). Διεξάγοντας μια συγκριτική ανάλυση συμβατικών και ΜΜ προσεγγίσεων, αξιολογεί τα αντίστοιχα δυνατά σημεία και περιορισμούς στην κατανόηση των εκπαιδευτικών δεδομένων. Η έρευνα εντοπίζει βέλτιστες πρακτικές για την ανάλυση δεδομένων κειμένου, ιδιαίτερα την ανατροφοδότηση των μαθητών, παρέχοντας αξιόπιστες γνώσεις σχετικά με τις ικανότητες των εκπαιδευτικών και τους παράγοντες που επηρεάζουν την ικανοποίηση των μαθητών.

Ενσωμάτωση τεχνικών ΑΣ στην αξιολόγηση ικανοτήτων: Μια επιπλέον συμβολή αυτής της έρευνας είναι η εφαρμογή των τεχνικών ΑΣ στην αξιολόγηση των ικανοτήτων των εκπαιδευτικών, που ανοίγει νέους δρόμους για την αξιολόγηση της αποτελεσματικότητας της διδασκαλίας και τον εντοπισμό περιοχών προς βελτίωση, όπως δημοσιεύτηκε σε άλλη μελέτη με τίτλο «Assessing Teacher Competencies in Higher Education: A Sentiment Analysis of Student Feedback» (Dervenis et al., 2024). Η σύγκριση των τεχνικών που βασίζονται σε λεξικό και σε ΜΜ μέσα στο πλαίσιο της εκπαίδευσης προσφέρει μια ισχυρή βάση για μελλοντική έρευνα πάνω στις μεθόδους αξιολόγησης ικανοτήτων.

Προώθηση του ρόλου της τεχνητής νοημοσύνης στην εκπαιδευτική ανάλυση: Αυτή η έρευνα υπογραμμίζει επίσης τις δυνατότητες των εργαλείων που υποστηρίζονται από τεχνητή νοημοσύνη, όπως τα chatbots, στην απλοποίηση και τη βελτίωση της ανάλυσης των σχολίων των μαθητών, όπως αναλύεται στο χειρόγραφό μας που βρίσκεται επί του παρόντος υπό αναθεώρηση προς δημοσίευση. Επιδεικνύοντας τη σκοπιμότητα χρήσης chatbot για ΑΣ και εξόρυξη γνώμης, η μελέτη υπογραμμίζει την ικανότητά τους να εξορθολογίζουν τις διαδικασίες συλλογής και ερμηνείας δεδομένων. Αυτή η συνεισφορά όχι μόνο απλοποιεί τη ροή αναλυτικών εργασιών, αλλά και

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

μεγιστοποιεί το εύρος των πρακτικών πληροφοριών που είναι διαθέσιμες σε εκπαιδευτικούς και διαχειριστές.

Πρακτική εφαρμογή: Τα ευρήματα αυτής της διατριβής έχουν άμεσες συνέπειες στην εκπαιδευτική πολιτική και πρακτική. Παρέχοντας μια δομημένη προσέγγιση για τον καθορισμό και την αξιολόγηση των ικανοτήτων των εκπαιδευτικών, αυτή η έρευνα υποστηρίζει τη λήψη αποφάσεων βάσει στοιχείων σε θεσμικό και εθνικό επίπεδο. Επιπλέον, οι μεθοδολογίες και οι ιδέες που παρουσιάζονται σε αυτή τη μελέτη προσφέρουν πολύτιμα εργαλεία για τη βελτίωση της ποιότητας της διδασκαλίας και, κατ' επέκταση, των μαθησιακών εμπειριών των εκπαιδευομένων.

Συνοπτικά, αυτή η διατριβή συνεισφέρει ουσιαστικά στον τομέα της ανώτατης εκπαίδευσης ορίζοντας ένα ισχυρό πλαίσιο των ικανοτήτων των εκπαιδευτικών και αναπτύσσοντας μεθόδους για την αξιολόγησή τους, με τη χρησιμοποίηση καινοτόμων τεχνικών ανάλυσης δεδομένων και με τη διερεύνηση των παραγόντων που επηρεάζουν την αποτελεσματικότητα της διδασκαλίας. Μέσω αυτών των συνεισφορών, η έρευνα όχι μόνο εμπλουτίζει την ακαδημαϊκή κατανόηση, αλλά διασφαλίζει επίσης ότι η ΑΕ παραμένει συμβατή με τις παγκόσμιες εκπαιδευτικές προόδους, ενισχύοντας πρακτικές που αντιμετωπίζουν τις εξελισσόμενες ανάγκες των μαθητών, των εκπαιδευτικών και των υπευθύνων χάραξης πολιτικής και προσφέροντας λύσεις για τη βελτίωση τόσο των διδακτικών πρακτικών και της ποιότητας της διδασκαλίας όσο και των μαθησιακών αποτελεσμάτων στην τριτοβάθμια εκπαίδευση. Τέλος, τα αποτελέσματα υπογραμμίζουν τη σημασία της συνεχούς προσαρμογής των δεξιοτήτων διδασκαλίας, ώστε να ανταποκρίνονται στις απαιτήσεις των τεχνολογικών και εκπαιδευτικών προόδων, διασφαλίζοντας ότι οι εκπαιδευτικοί είναι καλά εξοπλισμένοι για να αντιμετωπίσουν τις σύγχρονες προκλήσεις στην τριτοβάθμια εκπαίδευση.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

1. Introduction

1.1. Introduction to the subject

In today's rapidly evolving educational landscape, the need for competent educators has never been greater. The quality of teaching, which directly impacts student learning outcomes, is increasingly recognized as a critical factor in the development of intellectual and professional capital—not just on an individual level, but even at the level of an entire nation. This growing recognition has driven educational reforms worldwide, with a focus on identifying, developing, and enhancing teacher competencies, particularly in higher education.

Competencies, broadly defined as the combination of knowledge, skills, and abilities that enable individuals to perform effectively in a given context (European Commission, 2008), are essential for educators. For example, understanding effective teaching methods, being proficient in the use of digital tools for learning, and creating an inclusive and encouraging classroom atmosphere are vital competencies for educators in the modern era. For teachers, these competencies include not only expertise in a specific subject area but also pedagogical skills, social abilities, and personal development strategies that foster an engaging and effective learning environment. The demand for continuous professional development has become paramount, as educators must constantly adapt to new technologies, teaching methods, and student needs. As a result, understanding which competencies are critical for effective teaching and how they can be cultivated remains a key challenge in our educational systems.

At the same time, the role of data, particularly student feedback, in evaluating and improving teaching practices has drawn significant attention in recent years. Academic institutions, with access to vast amounts of educational data, are increasingly turning to data mining and SA methods to extract valuable insights from student feedback. This data, often collected through traditional closed-ended questionnaires, can reveal crucial information about various aspects of the teaching and learning process, such as the effectiveness of teaching methods, the competence of instructors, and the alignment of course's content with learning objectives (Rybinski, 2023; Mayordomo et al., 2022). However, despite the availability of large datasets, traditional methods often fail to fully capture the richness of student perspectives, especially in open-ended feedback. This highlights the need for more advanced data-driven approaches to analyse textual data and uncover deeper insights. Contemporary approaches and techniques that use ML algorithms, drawn from other fields, such as …. can be utilized to analyse and categorize subjective information within text, offering a promising solution to this challenge. By leveraging these computational methods to analyse and categorize subjective information, institutions can gain a more nuanced understanding

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

of how students perceive their instructors and the overall learning experience. These approaches will help not only identify strengths and weaknesses in teaching practices but also provide reliable data that can inform curriculum design and teaching strategies (McDonald et al., 2019). The integration of such advanced analytical techniques can assist institutions in making data-informed decisions to improve teaching quality, ensuring that educators are better equipped to meet the diverse needs of their students.

In conclusion, the need for educators to possess a wide range of competencies is central to improving the quality of education. As institutions strive to meet the challenges of a constantly changing educational environment, the integration of data analysis techniques, such as SA, topic modeling and machine learning, provide powerful tools for evaluating and improving teaching practices. By harnessing the wealth of student feedback data, universities can make informed decisions that not only enhance teaching effectiveness but also contribute to the overall improvement of student learning experiences. This ongoing process of reflection, analysis, and adjustment is essential to creating an educational system that is both adaptive and responsive to the needs of its learners. All of the above served as the foundation and starting point for our research, guiding us in the search of a specific framework for teacher competencies and methods for measuring them.

1.2. Research problem

One of the primary challenges in HE is the lack of a clear and comprehensive framework for defining and measuring the competencies of educators. While it is widely accepted that teacher effectiveness directly impacts student learning outcomes, a significant gap remains in understanding which specific competencies are essential for effective teaching. This gap is particularly evident in the absence of an internationally recognized framework for identifying the critical competencies that educators should possess. Without such a framework, educational institutions struggle to establish consistent and reliable methods for assessing and developing their teaching staff.

At the same time, there is still a gap in how specific teaching competencies can be effectively identified and measured. Traditional methods for evaluating teacher performance, such as student surveys and peer assessments, often rely on standardized, quantitative data, which may overlook nuanced insights embedded in open-ended student feedback. Furthermore, the rapid advancement in ML techniques and SA offers new opportunities for exploring educational data in more sophisticated ways. However, there is a lack of research investigating how these methods can be utilized for assessing teacher competencies and how they compare to traditional techniques.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

These gaps—both in defining a clear framework for teacher competencies and in measuring them—represent a key area that this research aims to address. By establishing a detailed framework for teacher competencies and exploring innovative data analysis techniques, this study will provide valuable insights into how student feedback can be leveraged to assess and improve teaching quality.

1.3. Research aims and objectives

The primary goal of this research is to address the existing gap in HE regarding the definition and measurement of teacher competencies. Specifically, this study seeks to identify and create a comprehensive framework of the competencies required for effective teaching. The absence of such a framework provides an invaluable opportunity to evaluate and improve teaching practices, as it is crucial to understand which competencies are fundamental to fostering student learning and enhancing overall teaching quality.

To achieve this goal, the research will pursue the following objectives:

- Identifying and defining the core competencies of teaching staff higher education: This objective aims to create a clear framework outlining the competencies necessary for effective teaching, drawing on both theoretical foundations and practical insights from existing literature.

- Exploring the extent to which these competencies are addressed in the current literature: This objective will involve a review of existing studies and frameworks related to teacher competencies, identifying what has been covered and highlighting gaps that remain in understanding the full spectrum of competencies required for successful teaching.

- Exploring the use of student feedback data, particularly written text, to measure teacher competencies: This objective will examine how SA techniques and ML algorithms can be applied to student feedback in order to assess specific competencies and gain insights into effective teaching.

- Comparing and evaluating different data analysis techniques for measuring teacher competencies: This objective will examine various methods, including ML approaches and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

lexicon-based techniques, to determine their effectiveness in analyzing student feedback and providing accurate assessment of teacher competencies.

- Examining how the nature of a course affects student perceptions of teacher competencies: This objective will explore whether student perceptions of teacher competencies vary based on factors such as the difficulty of the course content.

- Exploring the relationship between competencies identified through Topic Modeling (TM) and those documented in the literature: This objective will compare competencies derived from student feedback analysis using appropriate algorithms, with those proposed in existing educational frameworks, assessing alignment and potential discrepancies.

- Examining the effectiveness of chatbots as data analysis tools in student satisfaction surveys: This objective will focus on whether chatbots can be used effectively to analyze student feedback, particularly in terms of their ability to measure sentiment and provide useful insights into teaching quality.

Through these aims and objectives, the research seeks not only to define the competencies necessary for effective teaching but also to explore innovative techniques for measuring them using student feedback data. By bridging the gap in defining educator competencies and developing methods for their assessment, this research aims to offer valuable insights that will improve both teaching quality and student learning experiences in higher education.

1.4. Contribution

Our research provides a multifaceted contribution to the field of higher education, focusing on the definition, evaluation, and enhancement of teacher competencies. By synthesizing theoretical frameworks, innovative methodologies, and practical applications, the research advances our understanding of effective teaching practices and their role in fostering student engagement and satisfaction. The key contributions of this dissertation are outlined as follows:

- Development of a Teacher Competencies Framework: A significant contribution of this research is the formulation of a comprehensive framework of teacher competencies in higher education, as published in our paper titled “A review of research on teacher competencies in higher education” (Dervenis et al., 2022). This framework serves as a pivotal reference point for ensuring quality in the recruitment, professional development, and evaluation of teaching staff. Beyond its practical value, the framework facilitates

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

meaningful dialogue among stakeholders, including policymakers, educators, researchers, and education managers, fostering a collaborative approach to Quality Assurance (QA) and educational improvement. By emphasizing the prominent role of these competencies in the teaching-learning process, workplace environments, and broader societal contexts, the framework highlights their essential role in responding effectively to challenges and achieving optimal outcomes.

- Innovative Application of ML in education: This dissertation bridges the gap between traditional educational analysis methods and emerging ML techniques, as published in one of our studies titled "Sentiment analysis of student feedback: A comparative study employing lexicon and machine learning techniques" (Dervenis et al., 2024). By conducting a comparative analysis of conventional and ML approaches, it evaluates their respective strengths and limitations in understanding educational data. The research identifies best practices for analyzing textual data, particularly student feedback, providing actionable insights into teacher competencies and the factors that influence student satisfaction.

- Integration of SA techniques in competency assessment: a novel contribution of this research is the application of SA techniques in assessing teacher competencies, opening new pathways for evaluating teaching effectiveness and identifying areas for improvement, as published in another study titled “Assessing Teacher Competencies in Higher Education: A SA of Student Feedback” (Dervenis et al., 2024). The comparison of lexicon-based and ML techniques in the context of education offers a robust foundation for future research, paving the way for more precise and scalable methods of competency assessment.

- Advancing the role of AI in educational analysis: This research also highlights the potential of AI-powered tools, such as chatbots, in simplifying and enhancing the analysis of student feedback, as discussed in our manuscript currently under review for publication. By demonstrating the feasibility of using chatbots for SA and opinion mining, the study underscores their capacity to streamline data collection and interpretation processes. This contribution not only simplifies the analytical workflow but also expands the scope of actionable insights available to educators and administrators.

- Practical Application for Policy and Practice: The findings of this dissertation have direct implications for educational policy and practice. By providing a structured approach to defining and assessing teacher competencies, this research supports evidence-based decision-making at institutional and national levels. Additionally, the methodologies and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

insights presented in this study offer valuable tools for enhancing teaching quality and, by extension, student learning experiences.

In summary, this dissertation makes a substantial contribution to the field of HE by defining a robust framework of teacher competencies, employing innovative data analysis techniques, and exploring the contextual factors that influence teaching effectiveness. Through these contributions, the research not only enriches academic understanding but also ensures that HE remains in step with global educational advancements, fostering practices that address the evolving needs of students, educators, and policymakers and offers practical solutions for improving teaching practices and student outcomes.

1.5. Thesis outline

Chapter 1: Introduction

This chapter provides an overview of the research, detailing its aims and objectives. Additionally, it highlights the fundamental research problem and presents the structure of the thesis.

Chapter 2: Literature review - part I (Teacher competencies model)

This section presents an in-depth exploration of the evolving global landscape of skills and competencies, particularly in the context of higher education. Through a thorough literature review, we examine both the conceptual framework, exploring educational concepts, and learning theories, focusing on the roles and competencies of teaching staff in higher education.

Chapter 2: Literature review - part II (Assessing teacher competencies through student feedback)

This section presents an overview of the use of data mining techniques in universities today. It highlights how institutions leverage vast amounts of data, often sourced from student feedback, to enhance teaching quality. It also discusses the challenges posed by the sheer volume and complexity of this data and explores how it encompasses various aspects of student experiences, from course objectives to instructor competence.

Chapter 3: Research design and methodology

This chapter provides a comprehensive overview of the stages involved in the research process we followed. Through a detailed diagrammatic representation, this chapter aims to describe each phase of our methodology in a focused and detailed manner.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Chapter 4: Implementation and results

This chapter aims to provide a comprehensive overview of the implementation process and results derived from the analysis of student feedback, aligning with the overarching goal of this dissertation: "Evaluating and Enhancing Academic Excellence: A Competency-Based Model for the Quality of Teaching". The methods and tools discussed herein are the result of a systematic application of advanced data analysis techniques across various phases of the research. Specifically, this chapter integrates insights from five of our published studies, each contributing distinctively to the analysis of student feedback and the evaluation of teaching quality. While some studies focus directly on identifying and validating teacher competencies through SA and topic modeling, others demonstrate the applicability of ML and advanced tools for extracting meaningful insights from educational data.

Chapter 5: Discussion, conclusions and recommendations

This chapter begins by presenting the conclusions of this research through a detailed discussion of its key findings. It then highlights the limitations encountered during the study, as well as the implications for both academia and society. Finally, suggestions for future research are provided to inspire the research community.

1.6. Summary

This chapter introduces the research topic, focusing on the research problem and the motivation behind the preparation of this thesis. It defines the aims and objectives of the research, as well as the overall direction of the methodology followed. Additionally, the outline of the thesis is presented, identifying the main sections and the structure of the work.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

2. Literature review

2.1. Part 1: Teacher competencies model

This section presents an in-depth exploration of the evolving landscape of skill and competence, particularly within the context of higher education. It delves into the complexities of defining these terms in contemporary society, highlighting the various dimensions such as soft and hard skills, interpersonal skills, and emotional intelligence. Additionally, it sheds light on the crucial role of educators in fostering learning outcomes and the continuous development of their competencies, especially within HE. This part of the literature review examines both the conceptual framework, exploring educational concepts, and the learning theories, focusing on teaching staff roles and competencies in higher education. Thirty-nine (39) scientific papers were studied in detail from a total of 102 results, which were eligible based on the preferred reporting items for systematic reviews and meta-analyses statement. A multi-dimensional approach to teacher competencies in HE was revealed that consists of six main dimensions with their respective characteristics. Thirty-three (33) discrete teaching staff competencies were identified and distributed in the aforementioned dimensions.

2.1.1. Introduction

Skill and competence have always been concepts difficult to define. In the past, however, their definition seemed a much simpler matter than it is today. Today, the semantics of these terms is as complex as ever. There are soft and hard skills, skills that are generic and transferable, interpersonal skills, emotional and aesthetic skills. According to the European education and training policy (European Centre for the Development of Vocational Training., 2008), "skill" is defined as the ability to apply knowledge and use know-how to complete tasks and solve problems, while "competence" is the ability to apply learning outcomes adequately in a defined context (education, work, personal or professional development), to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development.

Today, more than ever, the challenge of a country's competitiveness is based on highly-skilled work, across all professional profiles. The European Union has adopted the tendency to upgrade the skills of its workforce, calling on all its Member States to join this strategic direction by “Rethinking Education: Investing in skills for better socio-economic outcomes”. In the long-term, skills can trigger innovation and growth, move production up the value chain, stimulate the concentration of higher level skills in the EU and shape the future labour market (OECD, 2018; UE, 2012). In addition, as highlighted by the European Commission (Caena, 2013) regarding educational

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

competencies, one of the keys to increasing levels of learning is to make sure that teachers have the necessary competencies for effective learning. In this context, every education system must adopt the necessary reforms needed for high quality teachers in terms of what competencies are needed, how these competencies can be identified and developed and what policies can help in this regard. It is not only necessary to have special competencies such as professional knowledge, social skills and ability in scientific research but it also necessary to continuously strengthen them, to inspire students in their future development (Caena, 2013). It is vital that teachers be constantly encouraged to develop, broaden and cultivate their competencies. In this direction, especially in the last decade, many countries are moving to review (Caena and Redecker, 2019; Krumsvik, 2014) and strengthen the profile of their teachers.

Despite the wide variety of approaches to defining the competencies that teachers are generally required to be able to deploy based on the different national educational policies (Caena, 2013) as well as the complexity of the concept of 'competency' that is used in differing ways, this literature review attempts to present the competencies of teaching staff in higher education. This chapter endeavors to collect and synthesize previous research thoroughly in order to identify teaching staff competencies derived from the roles and tasks attributed to university professors. It also endeavors to identify the competencies of teachers in HE horizontally, in order to maintain a homogeneous range, avoiding the inclusion of any domains of expertise that may have additional specialized skills. Furthermore, an effort is made to examine their impact and importance. Therefore, the research questions that guided our initial research are:

1. What are the competencies of the teaching staff in HE and what is their impact?
2. To what extent are these competencies addressed in the relevant literature?

In the following subsections of this part, we describe in detail the methodology we followed in order to collect the relevant literature. Then, relevant tables are used to present the results and information in a clear and detailed manner. In the last subsection of this chapter, closing remarks are provided.

2.1.2. Methodology

To get an overall view of the related field, a systematic review was carried out based on PRISMA methodology. We chose the PRISMA methodology as it represents the gold standard for systematic reviews in academic research. Its structured approach ensures comprehensive coverage and minimizes bias in the literature selection process (Page et al., 2021). PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Initially, it was developed as an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

(Moher et al., 2009). The aim of the PRISMA statement was to help authors improve the reporting of systematic reviews. PRISMA can also be used as a basis for reporting literature reviews for other types of research, as in the present study. A search process was thoroughly performed on ACM Digital Library, IEEE Xplore, Springer, Elsevier and Scholar databases to cover the relevant field. Boolean operators were used to combine different search queries and proximity operators (Hammerstrøm et al., 2009) to find words within a specified range. For the purposes of our search we used the following query in universal format: ((teacher OR "academic staff" OR professor) NEAR/3 competencies) AND ("higher education" OR "tertiary education"). The NEAR/3 operator in this search syntax ensures that the terms 'teacher', 'academic staff', or 'professor' appear within 3 words of 'competencies' in either direction. Unlike the simpler AND operator, NEAR/3 captures more relevant results by requiring these terms to be in close proximity, which typically indicates a stronger conceptual relationship between them. This proximity search helps filter out articles where these terms might appear in the same document but are not actually discussing teacher competencies specifically. The '/3' parameter defines this maximum word distance, making the search more precise than a broad keyword combination. We adjusted it according to syntax requirements of the above databases and executed appropriate combinations in order to penetrate the relevant literature and find the appropriate papers for the needs of our research. Between February and March 2021, the search process was conducted, followed by an examination as to, which of the 102 retrieved results should be included as eligible or excluded. The examination criteria had to do with whether they were duplicates, the written language, the title, the abstract and whether or not the full text was relevant to the purpose of our research. More specifically, in addition to the 102 retrieved results, another 3 papers were added because of their relevance to the field although they had not appeared in the search. In the next phase, the initial results were limited to 93 after having identified duplicates. Next, we screened the titles and abstracts for articles relevant to our research questions and with a complete text in English. The results were limited to 69.

Subsequently, the research focused exclusively on the complete content of each of the 69 results in order to review their eligibility according to:

i) whether the retrieved papers described skills or competencies that were fully or partially related to the teaching staff in HE and,

ii) whether theories, models and concepts related to the specific field were explicitly mentioned.

Following the above restrictions, 23 papers were removed. In addition, 7 papers were judged necessary to use for future research purposes.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Figure 1 illustrates thoroughly the process that was performed based on the template of PRISMA 2020 flow diagram (PRISMA 2020, n.d.), where in the final phase, our bibliographic research included 39 papers that finally met the research criteria.

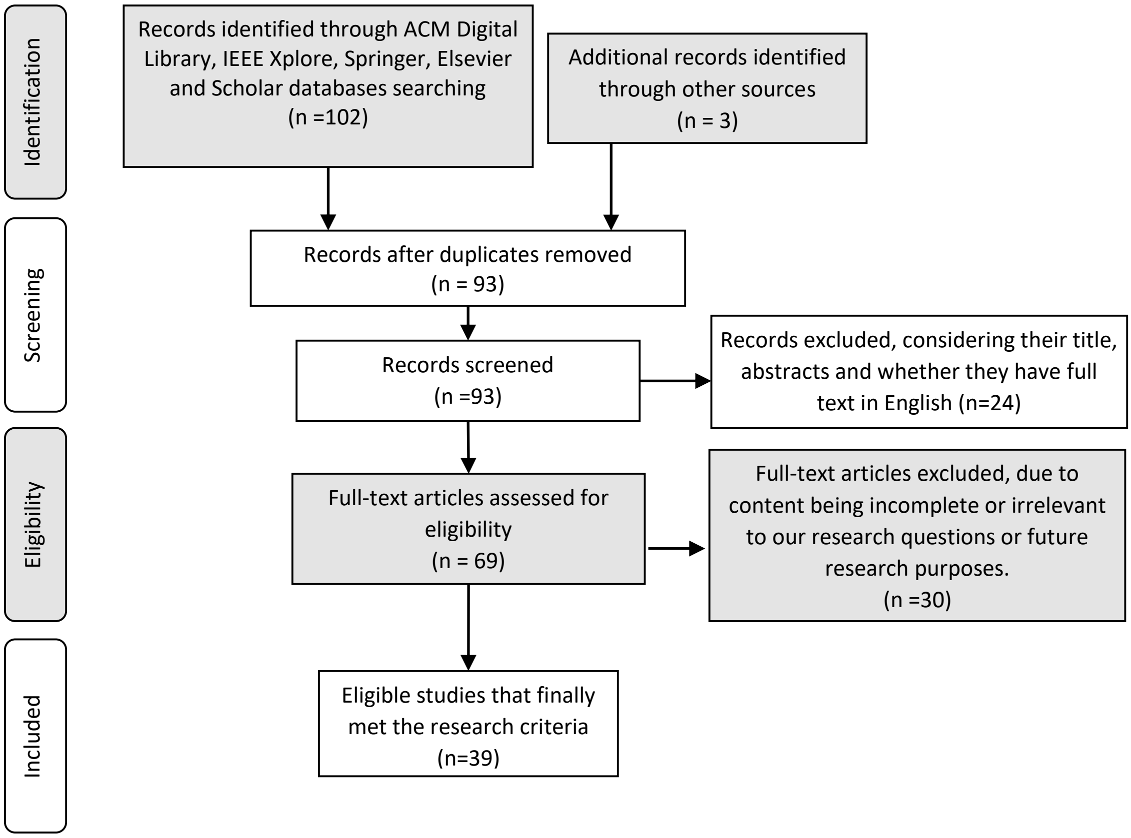


Figure 1. Relevant bibliography search process flowchart based on PRISMA statement (source: Authors' own work).

In our analysis we will not focus strictly on the definition of "competence" because we consider that definitions and concepts are tools scientists use to convey meaning and enable communication between people exchanging arguments; arguments and ideas matter. Nevertheless, it is worth noting at least at this point, the approach of ESCO (European Skills, Competences, Qualifications and Occupations) which, among other things, applies the same definition of "competence" as the European Qualification Framework (EQF). According to ESCO, "competence" means the proven ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development.

We started our analysis by taking a closer look at the competencies of university teachers by analyzing concepts, theories and HE teacher roles from the existing relevant literature. Figure 2 summarizes the steps that this section followed to arrive at tangible results. The result of the initial stage (stage 1) was the final collection of 39 articles based on the PRISMA methodology as described

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

in detail in the previous section. Then, in stage 2, these articles were studied focusing on concepts, theories, principles and teacher roles. Also, through this study, the competency dimension addressed in the articles was recorded as well as to what extent. For this point, two indicators have been created which will be mentioned in detail below. Then, in stage 3, following the theoretical framework and the studies from the existing bibliography that referred to teaching staff competencies, six dimensions of competencies were identified for our analysis. Also, at this stage, the discrete competencies were listed, which were then, in step 4, distributed into the above categories-dimensions. In the final stage (stage 5), the influence of the competencies of the teaching staff, as well as the other findings of our analysis are summarized and presented in the relevant tables.

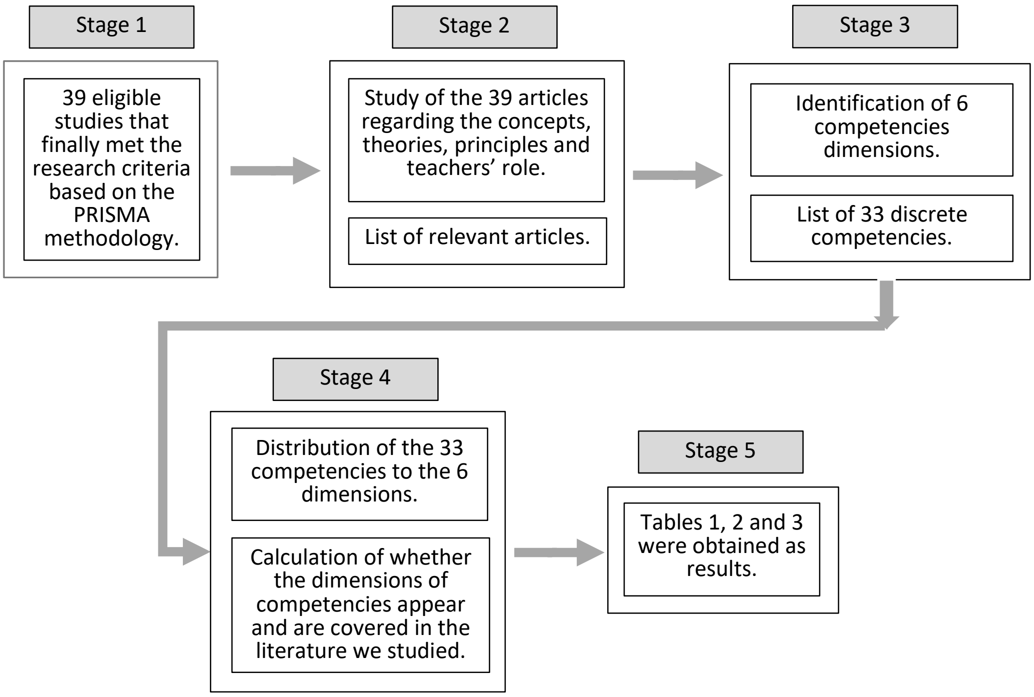


Figure 2. The study process of the 39 eligible articles (source: Authors' own work).

It is worth noting that, in general terms, all of the literature reviewed, asserts that university teacher roles are derived from traditional teacher functions. Consequently, it is necessary to clarify what these roles are, while at the same time, identify what competencies these roles require within the specificities of the duties and the mission of university professors (Alvarez et al., 2009).

2.1.3. A multi-dimensional approach of teacher competencies in higher education.

The professor's mission should not be limited to acting as a central point of knowledge transfer to students alone, but should, through interaction with students, aim at educating and shaping them in the sense of developing an integrated personality. A university professor has a dual role in

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

education, both as a teacher and as an educator. As a teacher, he/she is called upon to provide appropriate scientific knowledge to students. As an educator, he/she should have a role that contributes to the formation of the character and personality of a student who complies with the applicable rules (Mahulae et al., 2020), is kind to their fellow human beings, is active, creative, fair and cares about the environment (Cebrián et al., 2020). Jerome Delaney (2010), argues that a teacher should be distinguished by respect, honesty and a constant willingness to help students learn. Teachers must continually encourage their students, make themselves available and provide them with guidance. A teacher should be moral and fair to their students. Relevant research showed that the treatment of students by their teacher in a correct way and guided by fairness has a positive effect on their results (Blašková, 2014). The *Personality* competence dimension concerns the purely personal aspects of the teacher. It refers to mental and spiritual traits and behaviors that uniquely characterize the teacher.

In addition, several studies (López-Cámara et al., 2018; Bakhru, 2017; Cebrián et al., 2020), good teaching is that which is also based on the concept of expertise. However, this does not always mean that experience is the only necessary component for the development of expertise. Teachers need to know in detail how to teach their subject in terms of learning outcome (Park and Oliver, 2008) but at the same time in an ethical and rational way (G Fenstermacher, 2005). On the other hand, teaching in general is characterized by uncertainty regarding the diversity of learning environments and meeting the learning needs of students, requiring the teacher to adapt to each situation (Vogt and Rogalla, 2009). The *Professionalism* competence dimension clearly refers to the subject of one's professional activity in relation to how specialized a teacher is and how recognized they are in the field in which they teach.

According to EU Standards and Guidelines for Quality Assurance in European Higher Education (European Association for Quality Assurance in Higher Education, 2009), teachers are the most important source of knowledge and learning available to most students. Therefore, it is crucial that those who teach have a full knowledge and understanding of the subject they teach, the necessary skills and experience to impart them to students effectively in a variety of teaching contexts and also have access to feedback on their performance. Another study dealing with the conceptual approach to teaching staff development is that conducted by Ho (2000), which claims that conceptions that see teaching as a simple transmission of information are associated with didactic, uninspiring teaching approaches and also with surface learning in students, while conceptions that see teaching as facilitating student learning are more likely to be associated with teaching approaches commonly recommended for deep learning in students. Teachers should place the student at the centre when designing and structuring educational material. A teacher is, in a way like a coach who is as good as a parent (Singer and Moscovici, 2008; Iatrellis et al., 2017). The

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Educational competence dimension refers purely to aspects related to didactics. It is directly related to the way that teachers explain a subject to their students and how well they combine methods and techniques to make it understood by adapting it to the intellectual ability of their students.

Another important dimension is *Scientificity* which relates to the research that accompanies the teacher's work. The role of the university professor requires one's complete dedication to scientific research, the search for truth, the acquisition of knowledge as well as its promotion and dissemination through teaching and research, with the ultimate goal of impacting society. Teachers need to arrive at reliable knowledge by opening up new ways of thinking rather than one-dimensional views (Tigelaar et al., 2004). They must make the necessary connections with the knowledge they acquire in order to disseminate it, so that knowledge is constantly evolving and does not remain isolated and alienated. Research is an essential part of the knowledge building process. However, there is the perception that teachers who are devoted to research tend to teach worse because they are busy for long hours. However, results from very important research conducted by F. Hoffmann and P. Oreopoulos (2009) suggest that there is no strong correlation between research-focused teachers and teaching-focused teachers. Both categories of educators have effective and ineffective professors.

According to other views (Mahulae et al., 2020; Bakhru, 2017; Anderson, 2004; Caena, 2011; Tigelaar et al., 2004; Theall and Franklin, 2001; Long et al., 2014), a good teacher is one who makes an effort to be in constant contact and communicate effectively with their students by providing them with information and knowledge. Meaningful interactions enable university professors to objectively assess their students' progress and guide their future development appropriately (Blašková et al., 2014). The *Communication* competence dimension concerns the listening competencies, accuracy in verbal communication, persuasion and empathy that the teacher uses as a tool for the educational goal. The elements of this category enhance the effectiveness of the teacher in terms of the diffusion and understanding of knowledge by others. It contributes to the absorption of knowledge and avoiding misunderstandings with students and colleagues. With the right communication competencies, interaction and communication between people becomes effective and meaningful.

Teachers, in addition to the most important role they play, are also role models for their students. Our society demands digital competence from its citizens as never before. Teachers need to be able, not only to clearly demonstrate their digital skills to their students, but to transmit to them their creative and critical use of this technology. Digital technologies help to enhance and improve the learning process. Regardless of the teaching approach applied, the instructor's digital competence lies in coordinating the use of digital technology in the various stages of teaching, thus regulating the entire learning process. The digitally competent educator is properly trained to apply

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

digital technologies suitably, to design new teaching aids and to provide guidance and support to their students. Moreover, today more than ever, it is obvious that society needs flexible and resilient education systems as we face an unpredictable future. Wahab Ali (2020), with thorough research, focuses on the vulnerabilities of educational institutions that were revealed during the period of COVID-19. Through the findings of this research, it is confirmed that the digital competency of teachers is crucial since in times of lockdowns and social distancing as in the COVID-19 pandemic, online and distance learning is a necessity. The pandemic pushed universities toward e-learning. To better understand the nature of this competency dimension, the European Commission has set up the Digital Competence Framework for Citizens (Ferrari and Punie, 2013; Fitsilis, 2024) and the European Framework for the Digital Competence of Educators and in particular the DigCompEdu framework (Caena and Redecker, 2019), a widely accepted tool for measuring and certifying digital competency. This framework is the basis for teacher training and is addressed not only to the members of Europe but also beyond (Redecker, 2017). Unfortunately, for many teachers, computer technology and digital competence in general, are still considered as secondary to the educational process and not a central element of the educational mission. The *Digitality* competence dimension is essential for an effective teacher and concerns the set of skills, knowledge and attitudes that enable the creative and critical use of digital technology.

Following the theoretical framework and the relevant studies from the available literature, we initially aimed at identifying the main dimensions of the competencies of teachers in HE in a discrete way. Throughout the process of this exploration, in order to determine the competencies dimensions, it was found that investigating their characteristics was a powerful key. As a result, the characteristics of the identified dimensions were delineated, thus reinforcing the meaning of each dimension, giving it an explicit definition. The result of the described procedure was the determination of six dimensions of competencies with their respective characteristics, as presented in Table 1.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 1. A multi-dimensional approach of teacher competencies in HE (source: Authors' own work).

Competence dimensions	Characteristics	Appearance in
A: Personality	The *Personality* competence dimension concerns purely personal aspects of the teacher. It refers to mental and spiritual traits and behaviors that uniquely characterize the teacher. It also includes emotional mood, patience during discussions, character and temperament of a teacher.	(Caena, 2013), (Cochran-Smith et al., 2008), (Mahulae et al., 2020), (Crick, 2008), (Park and Oliver, 2008), (Bergsmann et al., 2015), (Blašková et al., 2014), (Delaney et al., 2010), (Blašková, 2014), (López-Cámara et al., 2018), (Pânişoară et al., 2014), (Bakhru, 2013), (Barth et al., 2007), (Tripathi, 2015), (Kugel, 1993), (Hoidn and Kärkkäinen, 2014), (Bakhru, 2017), (Anderson, 2004), (Caena, 2011), (G Fenstermacher, 2005), (Tigelaar et al., 2004), (Theall and Franklin, 2001), (Long et al., 2014), (Iqbal et al., 2019), (Cebrián et al., 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Krumsvik, 2014)
B: Professionalism	The *Professionalism* competence dimension clearly refers to the subject of one's professional activity in relation to how specialized a teacher is and how recognized he/she is in the field in which they teach. It also refers to how well he/she knows and applies methods and principles that govern their field, while	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Alvarez *et al.*, 2009), (Mahulae *et al.*, 2020), (Crick, 2008), (European Association for Quality Assurance in Higher Education, 2009), (Ho, 2000), (Park and Oliver, 2008), (Bergsmann *et al.*, 2015), (Vogt and Rogalla, 2009), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Blašková, 2014), (Berliner, 1994), (López-Cámara *et al.*, 2018), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi, 2015), (Kugel, 1993), (Hoidn and Kärkkäinen, 2014), (Bakhru, 2017), (Anderson, 2004), (Caena, 2011), (G Fenstermacher, 2005), (Tigelaar *et al.*, 2004), (Long *et al.*, 2014), (Metzler and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Characteristics	Appearance in
	ensuring the interconnection between theory and practice.	Woessmann, 2010), (Iqbal *et al.*, 2019), (Cebrián *et al.*, 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Krumsvik, 2014)
C: Educational	The *Educational* competence dimension refers purely to aspects related to didactics. It is directly related to the way that a teacher explains a subject to their students and how well he/she combines methods and techniques to understand it, by adapting them to the intellectual ability of their students. It also refers to a teacher's ability to be objective and non-discriminatory towards their students.	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Alvarez *et al.*, 2009), (Mahulae *et al.*, 2020), (Crick, 2008), (European Association for Quality Assurance in Higher Education, 2009), (Ho, 2000), (Singer and Moscovici, 2008), (Park and Oliver, 2008), (Bergsmann *et al.*, 2015), (Vogt and Rogalla, 2009), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Redecker, 2017), (Blašková, 2014), (López-Cámara *et al.*, 2018), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi, 2015), (Kugel, 1993), (Hoidn and Kärkkäinen, 2014), (Anderson, 2004), (Caena, 2011), (G Fenstermacher, 2005), (Tigelaar *et al.*, 2004), (Theall and Franklin, 2001), (Long *et al.*, 2014), (Iqbal *et al.*, 2019), (Cebrián *et al.*, 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Ali, 2019)
D: Scientificity	The *Scientificity* competence dimension concerns the expansion of new ways of thinking and not one-dimensional views. It includes the creation of the necessary	(Caena, 2013), (Cochran-Smith *et al.*, 2008), (Mahulae *et al.*, 2020), (Crick, 2008), (Blašková *et al.*, 2014), (López-Cámara *et al.*, 2018), (Caena, 2011), (Tigelaar *et al.*, 2004), (Cebrián *et al.*, 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Characteristics	Appearance in
	connections with the acquired knowledge, preventing its isolation. It supports that science and research are the guide to the well-being of peoples. It refers to the highly competent scientist and researcher who bring to light up-to-date, true and useful knowledge.	
E: Communication	The *Communication* competence dimension concerns the listening competencies, accuracy in verbal communication, persuasion and empathy that the teacher uses as a tool for the educational goal. The elements of this category enhance the effectiveness of the teacher in terms of the diffusion and understanding of knowledge by others.	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Mahulae *et al.*, 2020), (Crick, 2008), (Ho, 2000), (Singer and Moscovici, 2008), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Redecker, 2017), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi, 2015), (Kugel, 1993), (Bakhru, 2017), (Anderson, 2004), (Caena, 2011), (Tigelaar *et al.*, 2004), (Theall and Franklin, 2001), (Long *et al.*, 2014), (Iqbal *et al.*, 2019), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Krumsvik, 2014)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Characteristics	Appearance in
F: Digitality	The *Digitality* competence dimension concerns the set of skills, knowledge and attitudes that enable creative and critical use of digital technology for didactic purposes. It mainly includes key components of digital competence to locate and retrieve digital data, as well as to interact, communicate and collaborate through digital technologies. It also encompasses, the basic knowledge to protect digital devices and content, personal data and troubleshooting in digital environments.	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Alvarez *et al.*, 2009), (Crick, 2008), (Delaney *et al.*, 2010), (Redecker, 2017), (Caena, 2011), (Ali, 2020), (Selvi, 2010), (Krumsvik, 2014), (Ali, 2019)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

2.1.3.1. Competencies per dimension

Subsequently, we focused exclusively with distinguishing, assembling and then distributing into the appropriate dimensions all competencies of HE teachers mentioned in the literature at our disposal. Several of the competencies identified were inherently related to each other, for example morality, ethics, honesty. The inherent relationships among these competencies and their interrelated characteristics were carefully examined and then combined or merged to maintain a single term of almost the same meaning. For instance, as in our case the C5: Morality (see Table 2).

Next, all the identified skills were distributed into the six categories, *Personality*, *Professionalism*, *Educational*, *Scientificity*, *Communication*, *Digitality*. The distribution was guided by whether they were related to the respective dimension and / or whether there was a relevant reference in the literature. For example, in the *Personality* category, we distributed those competencies that have more to do with a person's spiritual characteristics and ways of behaving. Even if we identified an item such as "Morality" that could obviously be associated with a number of categories, the relationship with that category that was strongest prevailed.

The resulting classification is presented in Table 2. It is worth noting that Table 1, and Table 2 could have a different structure and focus if a different rationale, approach or criteria were adopted. For example, the focus could alternatively be on two broad categories such as innate and acquired competencies or focus on domains of expertise with specialized skills. However, there is no intention to undermine or challenge any other similar or alternative approach. Instead, the aim was to highlight the diversity of approaches and to encourage the exchange of views for the purpose of exploring the specific subject.

In addition, in Table 2, it was deemed necessary to add two indicators, $Apr_{(x)}$ and $Cvr_{(x)}$ as shown in their respective columns. The "appearance" indicator $Apr_{(x)}$ represents the percentage of each competency dimension that appears in the available literature we studied, while the "coverage" indicator $Cvr_{(x)}$ represents the percentage of coverage of each dimension in the eligible papers, i.e. to what extent the competencies of each dimension were covered by the selected documents. The definition of indicators $Apr_{(x)}$ and $Cvr_{(x)}$, as well as an example of calculating each indicator separately is as follows:

For indicator $Apr_{(x)}$,

$$Apr_{(x)} = \frac{n_{(x)}}{s_p} \text{ x } 100 \qquad (2.1)$$

Where,

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

$Apr_{(x)}$: *percentage appearance of the x dimension in the eligible papers.*

x: *the A, B, C, D, E, F dimension.*

$n_{(x)}$: *the number of documents indicating the corresponding dimension* x*.*

s_p: *the number of eligible documents. In our research we had 39.*

As an example, we want to find the $Apr_{(F)}$. That is, the percentage appearance of the $x = F$ dimension in the eligible papers.

Therefore, as shown in Table 2,

$$n_{(x)} = 12 \text{ and } s_p = 39$$

So,

$$Apr_{(F)} = \frac{n_{(F)}}{s_p} \text{x } 100 \Rightarrow$$

$$Apr_{(F)} = \frac{12}{39} \text{x } 100 \Rightarrow$$

$$Apr_{(F)} = 31\%$$

For indicator $Cvr_{(x)}$,

$$Cvr_{(x)} = \frac{\sum Cvr_{(x,i)}}{n_{(x)}} \text{x } 100 \qquad (2.2)$$

and,

$$Cvr_{(x,i)} = \frac{q_i}{q_{(x)}} \qquad (2.3)$$

Where,

$Cvr_{(x)}$: *total percentage coverage of the x dimension in the eligible papers.*

$Cvr_{(x,i)}$: *the coverage of dimension x from paper i (partial coverage).*

i: *3, 4, 5, 9, …. 48*

q_i: *the number of the discrete competencies of dimension x dealt with in paper i.* Note that the variable q_i is refers to a dimension in any case.

$q_{(x)}$: *the number of the discrete competencies of each dimension x.*

$n_{(x)}$: *the number of documents indicating the corresponding dimension* x*, as defined above.*

As an example, we want to find the $Cvr_{(F)}$. That is, the percentage coverage of the $x = F$ dimension in the eligible papers with:

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

$$q_3 = 1,\ q_4 = 6,\ q_5 = 3,\ q_8 = 3,\ q_{10} = 1,\ q_{18} = 1,\ q_{19} = 6,\ q_{32} = 4,$$

$$q_{41} = 4,\ q_{44} = 2,\ q_{47} = 5,\ q_{48} = 4$$

And

$$n_{(F)} = 12$$

Therefore,

$$Cvr_{(F)} = \frac{\sum Cvr_{(F,i)}}{n_{(F)}} \text{ x } 100 \Rightarrow$$

$$Cvr_{(F)} = \frac{Cvr_{(F,3)} + Cvr_{(F,4)} + Cvr_{(F,5)} + Cvr_{(F,8)} + Cvr_{(F,10)} + Cvr_{(F,18)} + Cvr_{(F,19)} + Cvr_{(F,32)} + Cvr_{(F,41)} + Cvr_{(F,44)} + Cvr_{(F,47)} + Cvr_{(F,48)}}{n_{(F)}} \text{ x } 100$$

$\Rightarrow$

$$Cvr_{(F)} = \frac{\frac{q_3}{q_{(F)}} + \frac{q_4}{q_{(F)}} + \frac{q_5}{q_{(F)}} + \frac{q_8}{q_{(F)}} + \frac{q_{10}}{q_{(F)}} + \frac{q_{18}}{q_{(F)}} + \frac{q_{19}}{q_{(F)}} + \frac{q_{32}}{q_{(F)}} + \frac{q_{41}}{q_{(F)}} + \frac{q_{44}}{q_{(F)}} + \frac{q_{47}}{q_{(F)}} + \frac{q_{48}}{q_{(F)}}}{n_{(F)}} \text{ x } 100 \Rightarrow$$

$$Cvr_{(F)} = \frac{\frac{1}{6} + \frac{6}{6} + \frac{3}{6} + \frac{3}{6} + \frac{1}{6} + \frac{1}{6} + \frac{6}{6} + \frac{4}{6} + \frac{4}{6} + \frac{2}{6} + \frac{5}{6} + \frac{4}{6}}{12} \text{ x } 100 \Rightarrow$$

$$Cvr_{(F)} = 56\%$$

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 2. The list of teaching staff competencies in HE and the papers in which they were addressed (source: Authors' own work).

Competence dimensions	**Competencies**	**Apr(x) (% of the x dimension's appearance in the eligible papers)**	**Cvr(x) (% of the x dimension's coverage in the eligible papers)**	**n(x) (The total number of papers referencing each dimension)**	**Papers**
A: Personality	C1: Energy C2: Helpfulness C3: Personal integrity and reliability C4: Multiculturalism C5: Morality	72%	50%	28	(Caena, 2013), (Cochran-Smith *et al.*, 2008), (Mahulae *et al.*, 2020), (Crick, 2008), (Park and Oliver, 2008), (Bergsmann *et al.*, 2015), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Blašková, 2014), (López-Cámara *et al.*, 2018), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi, 2015), (Kugel, 1993), (Hoidn and Kärkkäinen, 2014), (Bakhru, 2017), (Anderson, 2004), (Caena, 2011), (G Fenstermacher, 2005), (Tigelaar *et al.*, 2004), (Theall and Franklin, 2001), (Long *et al.*, 2014), (Iqbal *et al.*, 2019), (Cebrián *et*

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	**Competencies**	**Apr(x) (% of the x dimension's appearance in the eligible papers)**	**Cvr(x) (% of the x dimension's coverage in the eligible papers)**	**n(x) (The total number of papers referencing each dimension)**	**Papers**
					al., 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Krumsvik, 2014)
B: Professionalism	C6: Theoretical knowledge C7: Practical knowledge C8: Experience C9: Discipline C10: Creativity	87%	70%	34	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Alvarez *et al.*, 2009), (Mahulae *et al.*, 2020), (Crick, 2008), (European Association for Quality Assurance in Higher Education, 2009), (Ho, 2000), (Park and Oliver, 2008), (Bergsmann *et al.*, 2015), (Vogt and Rogalla, 2009), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Blašková, 2014), (Berliner, 1994), (López-Cámara *et al.*, 2018), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi,

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Competencies	Apr(x) (% of the x dimension's appearance in the eligible papers)	Cvr(x) (% of the x dimension's coverage in the eligible papers)	n(x) (The total number of papers referencing each dimension)	Papers
					2015), (Kugel, 1993), (Hoidn and Kärkkäinen, 2014), (Bakhru, 2017), (Anderson, 2004), (Caena, 2011), (G Fenstermacher, 2005), (Tigelaar *et al.*, 2004), (Long *et al.*, 2014), (Metzler and Woessmann, 2010), (Iqbal *et al.*, 2019), (Cebrián *et al.*, 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Krumsvik, 2014)
C: Educational	C11: Transmissibility C12: Studiousness C13: Planning C14: Adaptability	87%	41%	34	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Alvarez *et al.*, 2009), (Mahulae *et al.*, 2020), (Crick, 2008), (European Association for Quality Assurance in Higher Education, 2009), (Ho, 2000), (Singer and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	**Competencies**	**Apr(x) (% of the x dimension's appearance in the eligible papers)**	**Cvr(x) (% of the x dimension's coverage in the eligible papers)**	**n(x) (The total number of papers referencing each dimension)**	**Papers**
	C15: Coaching C16: Inventiveness				Moscovici, 2008), (Park and Oliver, 2008), (Bergsmann *et al.*, 2015), (Vogt and Rogalla, 2009), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Redecker, 2017), (Blašková, 2014), (López-Cámara *et al.*, 2018), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi, 2015), (Kugel, 1993), (Hoidn and Kärkkäinen, 2014), (Anderson, 2004), (Caena, 2011), (G Fenstermacher, 2005), (Tigelaar *et al.*, 2004), (Theall and Franklin, 2001), (Long *et al.*, 2014), (Iqbal *et al.*, 2019), (Cebrián *et al.*, 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Ali, 2019)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Competencies	Apr(x) (% of the x dimension's appearance in the eligible papers)	Cvr(x) (% of the x dimension's coverage in the eligible papers)	n(x) (The total number of papers referencing each dimension)	Papers
D: Scientificity	C17: Resourcefulness C18: Innovativeness C19: Imagination C20: Diligence C21: Persistence C22: Effectiveness	28%	25%	11	(Caena, 2013), (Cochran-Smith *et al.*, 2008), (Mahulae *et al.*, 2020), (Crick, 2008), (Blašková *et al.*, 2014), (López-Cámara *et al.*, 2018), (Caena, 2011), (Tigelaar *et al.*, 2004), (Cebrián *et al.*, 2020), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010)
E: Communication	C23: Listening C24: Persuasion C25: Empathy C26: Presentation C27: Collaboration	64%	70%	25	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Mahulae *et al.*, 2020), (Crick, 2008), (Ho, 2000), (Singer and Moscovici, 2008), (Blašková *et al.*, 2014), (Delaney *et al.*, 2010), (Redecker, 2017), (Pânişoară *et al.*, 2014), (Bakhru, 2013), (Barth *et al.*, 2007), (Tripathi, 2015), (Kugel, 1993), (Bakhru, 2017), (Anderson, 2004), (Caena, 2011), (Tigelaar *et al.*,

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Competencies	Apr(x) (% of the x dimension's appearance in the eligible papers)	Cvr(x) (% of the x dimension's coverage in the eligible papers)	n(x) (The total number of papers referencing each dimension)	Papers
					2004), (Theall and Franklin, 2001), (Long *et al.*, 2014), (Iqbal *et al.*, 2019), (Hoffmann and Oreopoulos, 2009), (Selvi, 2010), (Krumsvik, 2014)
F: Digitality	C28: Digital literacy C29: Digital pedagogy C30: Digital communication and cooperation C31: Creating digital content C32: Safety	31%	56%	12	(Caena, 2013), (Caena and Redecker, 2019), (Cochran-Smith *et al.*, 2008), (Alvarez *et al.*, 2009), (Crick, 2008), (Delaney *et al.*, 2010), (Redecker, 2017), (Caena, 2011), (Ali, 2020), (Selvi, 2010), (Krumsvik, 2014), (Ali, 2019)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Competencies	Apr(x) (% of the x dimension's appearance in the eligible papers)	Cvr(x) (% of the x dimension's coverage in the eligible papers)	n(x) (The total number of papers referencing each dimension)	Papers
	C33: Basic troubleshooting				

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

The results show that most of the existing literature is focused on the *Personality*, *Professionalism* and *Educational* competencies dimensions. s. In contrast, the *Digitality* and *Scientificity* dimensions are addressed in only a limited number of papers. Notably only 12 out of 39 papers addressed the *Digitality* dimension. We assume that this is due to the nature of this particular dimension, which is related to the rapid introduction of Information and Communication Technology (ICT) in education in recent years. In comparison, the other dimensions are rooted in traditional theoretical educational models, which have been extensively studied over decades. Moreover, the *Digitality* dimension has not been extensively covered in all its aspects by these few documents. This implies that these documents have not dealt with all six competencies distributed to this dimension (see Table 2). Although the European Commission (Redecker, 2017) recommends paying special attention to this dimension by proposing a specific framework, as mentioned, it is clear that there is a long way to go for further research on this dimension and its adoption by teachers. Our findings also reveal that almost all the reviewed papers—specifically, 34 out of 39, address the *Professionalism* dimension, underscoring its perceived necessity. However, these works focus on some of the five dimension's competencies and not with all. Specifically, the coverage rate of this dimension from these 34 papers was 70%. A complete example of how the Cvr(x) index is calculated is given directly above, in this section. It should be noted at this point that the term "cover" is relative and it simply relates to our study and specifically to Table 2.

Additionally, a considerable number of documents, dealt with the *Personality* and *Educational* competencies dimensions. This indicates that the specific competence dimensions are not only well-researched but are also widely recognized as essential components of the educational process.

Finally, Figure 3 illustrates the relationship between the percentage appearance and coverage level of the dimensions in the eligible papers.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

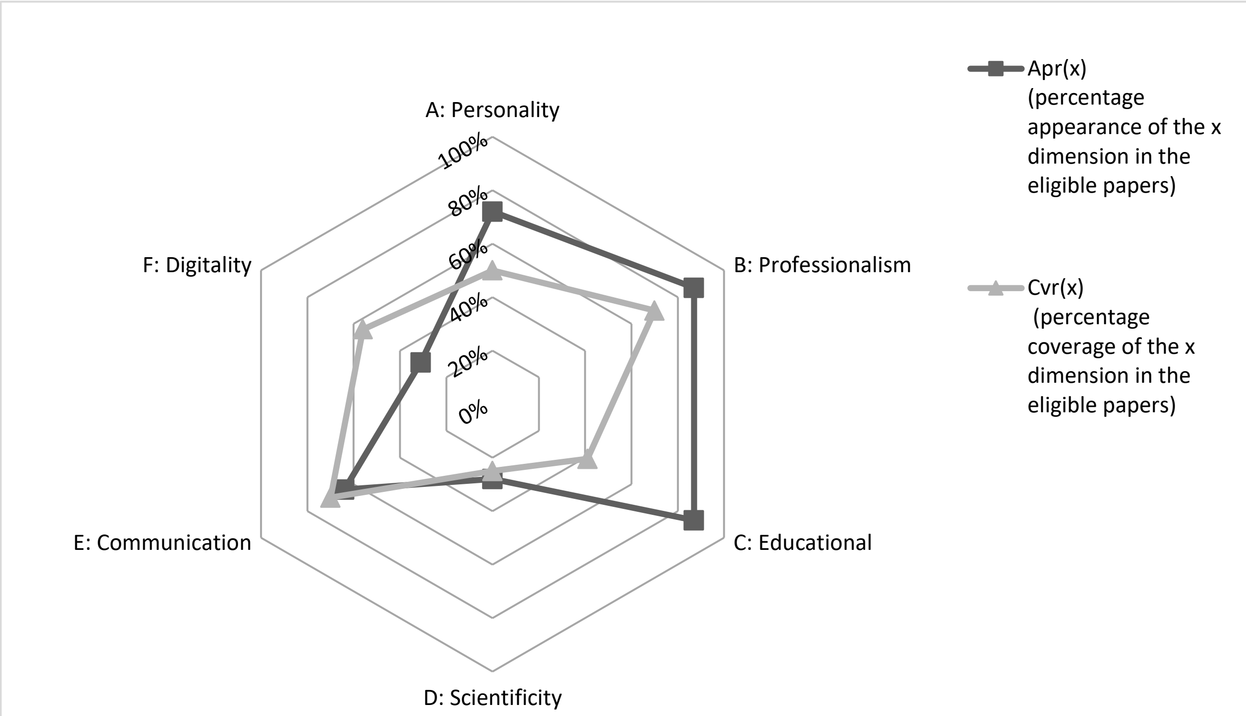


Figure 3. Percentage appearance Apr(x) and coverage level Cvr(x) of dimensions in the eligible papers (source: Authors' own work).

2.1.3.2. The influence of dimensions in the education process

Several of the papers investigated in our study also report the influence of the above dimensions on the educational process, on student learning outcomes and on society in general. A'en Vater Mahulae et al. (2020) reported that teacher professionalism has a positive and significant effect on teacher performance, which in turn positively impacts student learning outcomes. Students often conceptualise the professionalism of their teachers and this helps in the development and gradual cultivation of the professionalism that they reflect. Sanders & Rivers (1996) and Babu & Mendro (2003), through their research, reveal that of all the factors under the control of an educational institution, teachers are the strongest influence on student success. Metzler and Woessmann (2010) studied the effect of teacher competencies in HE on student achievements and found that the quality of teaching is directly related to student achievements, and it is very important for educators to develop strong teaching competencies to provide quality teaching.

Teacher competencies related to the *Educational* dimension ensure that the educator places the student at the center when designing educational material and guides the learning process effectively to enable self-directed learning (Tigelaar et al., 2004). Competencies falling under the *Communication* dimension ensure effective and meaningful interaction between teacher and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

student, thus contributing to the absorption of knowledge and avoiding misunderstandings. On the contrary, the teacher's inability to communicate with their students may lead to poor performance (Theall and Franklin, 2001). Additionally, the lack of communication and encouraging feedback from the teacher can lead to failure in the performance of students (Iqbal et al., 2019).

The competencies associated with the Scientificity dimension contribute to the systematic and creative work aiming at enhancing the level of knowledge, thereby fostering societal and cultural evolution. Furthermore, for teachers, these skills contribute to conveying the results of science to students in such a way as to be understood and inspiring for their proper future development. The data collected from the relevant papers and used to identify the “Digitality” dimension and the competencies included in it, at the same time revealed that they contribute to the educational process in various ways, providing substantial benefits to all involved. Redecker C. (2017) emphasizes that competencies that fall into the dimension of “Digitality” ensure the effective and responsible use of digital resources for learning, contribute to orchestrating the use of digital technologies in teaching and enhance assessment. The influence of the above six dimensions, including the capabilities contained in them, were identified and are presented briefly in Table 3.

Table 3. Influence of each competency dimension (source: Authors’ own work).

Competence dimensions	**Influence**	**Competencies**
A: Personality	Contributes to the development of students' personality and general behavior. It also contributes to the creation of positive role models. Teacher has a positive attitude towards students and exhibits respect towards them.	C1: Energy C2: Helpfulness C3: Personal integrity and reliability C4: Multiculturalism C5: Morality
B: Professionalism	Contributes to the diffusion of the latest knowledge and trends, ensuring the acquisition of knowledge of his/her students and colleagues. Professionalism has a positive and significant effect on teacher performance, which in turn reflects positively on students.	C6: Theoretical knowledge C7: Practical knowledge C8: Experience C9: Discipline C10: Creativity
C: Educational	Contributes to the cultivation of students' intellectual abilities. It also helps to consolidation knowledge. Teacher places	C11: Transmissibility C12: Studiousness C13: Planning

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Competence dimensions	Influence	Competencies
	the student at the center when designing and structuring educational material.	C14: Adaptability C15: Coaching C16: Inventiveness
D: Scientificity	Contributes to the development of knowledge and in general to the progress and growth of society. In addition, it contributes to the quality indicator of educational institutions.	C17: Resourcefulness C18: Innovativeness C19: Imagination C20: Diligence C21: Persistence C22: Effectiveness
E: Communication	Contributes to the absorption of knowledge and avoiding misunderstandings with students and colleagues. With the right communication competencies, interaction and communication between people becomes effective and meaningful.	C23: Listening C24: Persuasion C25: Empathy C26: Presentation C27: Collaboration
F: Digitality	In general, facilitates the use of digital technology in the classroom. Contributes to the achievement of educational goals and facilitates communication between teacher and students. It also contributes to the proper examination and assessment of digital tools as potential means for learning and teaching.	C28: Digital literacy C29: Digital pedagogy C30: Digital communication and cooperation C31: Creating digital content C32: Safety C33: Basic troubleshooting

2.1.4. Limitations

Our study relied on the existing literature without using data drawn from a dedicated questionnaire. While, the selection of the databases was not exhaustive, it included significant databases such as ACM Digital Library, IEEE Xplore, Springer, Elsevier, Google Scholar, largely meeting the search criteria we set. Additionally, to maintain a homogeneous base of teacher competencies, specific domains of expertise were intentionally excluded.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

2.1.5. Practical and social implications

A well-defined framework of teacher competencies in HE, can be a useful reference point not only for ensuring quality in the selection of teachers and their career-long professional development, but as well for forming national education policy strategies. The definition of teacher competencies framework contributes to facilitating effective dialogue for the evaluation and QA in education between agencies, authorities, researchers, teachers, policymakers, education managers and different communities at large.

These competencies are central not only of the teaching and learning process but also in the workplace and in society in general. These are increasingly recognized as essential for addressing modern educational and societal challenges. An adequately prepared community and management equipped with the required employee competencies is able to react immediately and in a positive way to any obstacle, yielding optimal results.

2.1.6. Concluding remarks

Throughout this literature review, thirty-nine (39) documents were studied in detail from a total of 102 results, which were eligible for inclusion based on PRISMA methodology. As a result: a multi-dimensional approach to teacher competencies in HE was identified, which consists of 6 main dimensions with their respective characteristics. Then, 33 discrete teaching staff competencies were identified according to the research studied and were distributed in the 6 dimensions: *Personality, Professionalism, Educational, Scientificity, Communication and Digitality*.

While recording the competencies of the teaching staff, it was found that the majority of the papers were targeted at and focused on certain competencies, while others received very little attention. We believe that this is because the educational community does not perceive certain competencies as imperative, and obviously for this reason the appropriate scope for their investigation has not been given. The value of such competencies is often confirmed by needs and circumstances. One finding of our research, that strengthens this conclusion, is related to the digital competence of teachers on which the selected papers studied have little focus. What is more, this vital issue - not previously given due attention in the education community - has proven urgent on a global level during the challenges faced in education worldwide during the COVID-19 pandemic. It is therefore necessary that the educational community focus on digital training and be adequately prepared, so that in such cases it is fully equipped and ready to react immediately and in a positive way to any obstacle in the educational process.

We believe that the purpose of this research is not only to conduct an additional literature review with all that it entails. More importantly, our mission is to raise awareness in the educational community about the fact that, many of the competencies thoroughly revealed by our analysis are

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

still considered as “secondary” to the educational process and not as necessary elements of the educational mission. Teacher competencies should constantly be examined through scientific studies in light of varying national educational policies and reforms. It is therefore imperative to redefine teacher competencies in order to respond to an uncertain future full of change and new challenges.

Finally, through the current part of literature review, we have identified critical questions that need further research, such as the fact that the relationship between teaching and students' academic achievement may seem obvious, but how can this be proven? How can competencies be measured? What is the relationship between teacher competencies in HE and students' outcome?

2.2. Part 2: Assessing teacher competencies through student feedback

This section presents an overview of the use of data mining techniques in universities today. It highlights how institutions leverage vast amounts of data, often sourced from student feedback, to enhance teaching quality. The section discusses the challenges posed by the sheer volume and complexity of this data and explores how it encompasses various aspects of student experiences, from course objectives to instructor competence. By analyzing such data, universities can refine their curriculum design and make informed decisions about resource allocation and administrative policies. Additionally, the methods presented in this section empower teachers to assess their teaching methods, identify strengths and weaknesses, and make necessary adjustments, ultimately improving the overall effectiveness of teaching and student learning.

2.2.1. Introduction

Nowadays, with the rapid spread of data and its huge availability arising from various educational processes, the majority of universities frequently use data mining methods to examine data collected and to extract hidden knowledge from student feedback with the aim of improving teaching quality in universities (Mayordomo et al., 2022; Misuraca et al., 2021; Nguyen and Habók, 2023; Rybinski, 2023). However, the datasets are too large or complex to be effectively processed using traditional software tools like spreadsheets or basic statistical software. Consequently, institutions face the challenge of leveraging this vast data to enhance student learning. These data generally capture student perspectives on various aspects such as their attainment of a course's learning objectives, the effectiveness of instructional methods, the competence of the instructor, course content, and even the overall student experience. When utilized, these data help academic institutions into their curriculum design and in the management decisions of their organizations. Furthermore, they enable teachers to better understand their teaching approaches, strengths and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

weaknesses. In this way, teachers are better equipped to make appropriate changes thus contributing to the overall effectiveness of their teaching and student learning.

The traditional techniques for student feedback analyses are based on close-ended questionnaire data collection and analysis (Richardson, 2005; Wong and Li, 2020). These questionnaires, are often based on Likert-scale items, and provide quantitative data that may be analysed using traditional statistical methods. However, the majority of academic institutions, including our university, additionally offer students the ability to provide their feedback using open-ended questions. Despite being a rich source of insights, such open-ended responses are often underutilized since they require more sophisticated analytical approaches (McDonald et al., 2019).

This subsection of the literature review examines works that assess student's feedback and the techniques used for such type of learning analytics.

2.2.2. The value of student feedback in teaching progress

Academic institutions have long aimed to collect and take their students' views into account, in an effort to improve their teaching and curricula (Su et al., 2023; M. Lane & D. Meth, 2021). Flodén (2017), with thorough research, focuses on student feedback and its positive acceptance from university teachers, emphasizing that both positive and negative student feedback are equally valuable to them. Mandouit (2018), concluded that when provided with professional learning to support the process, student feedback has the potential to have a positive impact on teaching practices and contributes to opening up a dialogue about teaching and learning.

Another pioneering study dealing with an alternative approach to student feedback on courses is that conducted by Marley (2021), which asserts that the positive and negative experiences of online learners are extremely useful for educators involved in the design and development of online teaching. According to Nasim (2017), in their feedback, students' views touch on multiple aspects of their teachers such as teaching style, lecture organization, knowledge, punctuality, and communication competence making them valuable in the learning process.

Other research (Marsh, 2007; Kumar & Jain, 2015) has found that, student feedback contributes to various aspects of academic institutions by providing information and knowledge to facilitate decision making in their organization. Marsh (2007) claims that student opinion mining, beyond serving as diagnostic feedback for academics about the effectiveness of their teaching, provides a measure of teaching effectiveness for decisions regarding personnel appointments and promotions. This constitutes a valuable component for use in the QA processes of academic institutions (Tsinidou et al., 2010; Fitsilis et al., 2024 under review).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Student feedback is collected in different forms, such as close-ended questions (usually in Likert scales), open-ended questions and free text comments. Different mediums are also used, such as, classroom feedback, social media, online platforms, etc., by students to express their perceptions of issues about their instructors and institutions (Sangeetha & Prabha, 2021). This whole array of educational data has led to more demanding data management systems and complex analytical approaches, resulting in the emergence of a specialized research field known as Educational Data Mining (EDM) (Slater et al., 2017).

2.2.3. Educational data analysis

The term 'educational data' refers to information that is collected and organized to capture various aspects of the educational environment. This includes data related to students, parents, schools, teachers, and the relationships among them. Analyzing this data often involves methodologies like Educational Data Mining (EDM) and Learning Analytics (LA) (Knight & Shum, 2017; Lang et al., 2017; Lester et al., 2017; Srinivasa & Kurni, 2021). These approaches form a continuous and iterative process—discovering insights and patterns from the data, which in turn enhance our understanding of educational settings and inform improvements over time.

A particularly significant area within this domain is learning analytics, which focuses on the measurement, collection, analysis, and reporting of data about learners and their contexts. The goal is to better understand and optimize learning processes and the environments in which they occur. Learning analytics is inherently multidisciplinary, drawing on fields such as education, data science, psychology, and computer science. Its practical applications are vast, ranging from improving individual student outcomes to guiding strategic decision-making at an institutional level.

The cyclical nature of these processes underscores their iterative character—new data and discoveries lead to further analyses, continuously refining our understanding and fostering more effective educational practices. By leveraging these insights, educators and researchers can create more personalized and impactful learning experiences for all stakeholders. The use of Educational Data Mining and Learning Analytics (EDM/LA) to analyze educational data is a continuous, cyclical process. It involves discovering insights and patterns from the data, which assist in gaining general knowledge or understanding about educational environments, students, teachers, schools, and related factors. The term 'cyclical' implies that this process is ongoing and iterative, allowing new data leads to further analysis and discovery, thereby refining our understanding over time.

The availability and type of educational data vary depending on the educational environment (traditional classroom education, computer-based education, or blended learning) and the information system used, such as Learning Management Systems (LMS) (Coates et al., 2005; Rhode, 2017), Intelligent Tutoring Systems (ITS) (Pham & Sampson, 2022; Roll & Wylie, 2016) or Massive

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Open Online Courses (MOOCs) (Al-Rahmi et al., 2019; Hew & Cheung, 2014; Voudoukis & Pagiatakis, 2022). Therefore, different types of data can be collected to address various educational problems (Fitsilis et al., 2024; Romero & Ventura, 2013). This data includes:

- Student data: Demographic information and previous academic performance.
- Educator data: Skills and professional experience.
- Teaching, learning, and assessment data: Lesson plans, educational activities, assessment methods, classroom management.
- Infrastructure, human, and financial resource data: Educational and non-educational personnel, materials, software, expenses.
- Social and emotional development data of students: Support, respect for diversity, and special needs.

2.2.3.1. Methods and approaches for analyzing educational data

Analyzing educational data involves various methods and approaches to derive insights and improve educational outcomes. These methods utilize data mining, statistical analysis, machine learning, and other techniques tailored to educational contexts. Some key methods and approaches commonly used are the following:

- **Descriptive Analytics**

  Descriptive analytics forms the foundation for all other types of analysis. It is the simplest and most common use of data in businesses and organizations today. Descriptive analytics answers the question "what happened?" by summarizing past data, typically in the form of tables and graphs. Monitoring key performance indicators of an organization is a typical case of data analysis falling under this category (Marr, 2012).

  Descriptive analytics focuses on summarizing historical data to understand past trends and patterns within educational settings. Techniques include aggregation, summarization, and visualization of data related to student performance, attendance, and demographic characteristics.

- **Predictive Analytics**

  Predictive analytics aims to forecast future trends and behaviors based on historical data. In education, predictive models can predict student outcomes, identify at-risk students, and optimize interventions. Techniques include regression analysis, decision trees, and predictive modeling.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

- **Prescriptive analytics**

  Prescriptive analytics goes beyond prediction to suggest actions that can optimize outcomes. In education, it recommends personalized learning paths, intervention strategies, and resource allocation based on predictive models and predefined objectives.

- **Social Network Analysis (SNA)**

  SNA examines relationships and interactions among students, teachers, and educational communities. It uses graph theory to visualize and analyze social structures, communication patterns, and collaborative learning networks.

- **Natural Language Processing (NLP)**

  NLP enables the analysis of unstructured data, such as student essays, forum posts, and teacher feedback. SA, a subset of NLP, is used to assess attitudes, emotions, and opinions expressed in text, providing insights into student engagement, satisfaction, and learning experiences.

- **Integrating SA**

  SA in educational data can enhance understanding of student sentiment towards courses, educational materials, and learning environments. By analyzing text data from surveys, social media, or learning platforms, educators can gauge student satisfaction, identify areas needing improvement, and tailor interventions to enhance the learning experience.

2.2.4. SA in education

SA, also known as opinion mining, is a field of study, directly related to NLP that analyzes views, sentiments, evaluations, attitudes and positions of people on issues through the computational processing of subjectivity in text (Liu, 2022). SA aims at taking the sentiment of a simple word, phrase, sentence or even an entire document into consideration and rating it appropriately. Essentially, using appropriate lexicons or ML approaches (Salazar et al., 2022), opinion polarity is determined, as to whether it is positive, negative or neutral (Hutto and Gilbert, 2014).

The initial challenges encountered in the practical application of SA and opinion mining stemmed from the rapid increase in the pace and volume of comment content observed on social media on the web. Since then, SA has penetrated and found application in a wide range of fields, focusing on customer opinions about products and services through their written comments. As a result, with the growing use of Internet and online reviews has established SA as a powerful method for understanding customer preferences in various business sectors, such as marketing, entertainment, policy-making, public transportation, travel, and academia. For the first time in

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

history, we now have a huge volume of opinionated data recorded in digital form for analysis (Liu, 2022). As a result, SA, considered as a big data task, is constantly arousing the interest of researchers.

Traditionally, in academic institutions, student feedback analysis systems were based on close-ended questionnaires that included a set of questions with predefined possible answers (Kumar and Jain, 2015). With SA techniques, students are not limited to grade only those features that are included in the questionnaire, but instead they can freely express their opinion. According to Nasim et al. (2017) textual feedback is the best way for students to express their views spontaneously, without being limited through predetermined answers. Despite its potentially huge size, this quality educational data collected from text, can now be easily analyzed with the help of computer systems using SA techniques (Shaik et al., 2022).

Kumar, A., & Jain, R. (2015) proposed an automatic evaluation system based on SA that collects student feedback in text form. Misuraca et al. (2021) used a dataset of text, built by Welch and Mihalcea (2016), with 1042 comments extracted from a Facebook group where students expressed their opinions on courses and instructors related to the Computer Science department at the University of Michigan (US). They observed that, concerning the polarity score calculation, it is worth adjusting lexicons to the specific language commonly used by students, particularly because the polarization of some terms may be different from the polarization used in common language. Nitin et al. (2015) observed that students provide feedback related to labs, projects, skills, etc. when it comes to components of a course which are non-theoretical (interactive, practical, collaborative) in nature. On the contrary, when it comes to theoretical components, student feedback tends to focus directly with the profile of their teachers.

Although there is research that has shown that the neutral comments, in SA, are often overlooked (Wang and Wu 2011), even student feedback categorized as neutral can contribute positively to understanding educational issues. Ortigosa et al. (2014), especially in the education domain, point out the necessity, even of the students' neutral feedback, in order to comprehend their attitudes. It is not our intention to review the entire body of literature concerning SA from a technical point of view. Nevertheless, we will attempt to provide a brief, and in our opinion necessary, overview in the following section.

Most SA approaches rely heavily on lists of lexical features (e.g., words, phrases), called sentiment lexicons, generally labeled according to their semantic orientation as either positive, negative or neutral (Liu, 2022). According to Feldman (2013), the sentiment lexicon is the most crucial resource for most SA algorithms. This lexicon is classified into two main categories, (1) polarity-based lexicons and (2) valence-based lexicons.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

In the first case, a polarity-based lexicon is used to determine the binary polarity of words i.e., either positive or negative. For example, "angry" has negative polarity while "family" positive. To calculate the total polarity of a text, the polarity of each word is searched in the lexicon and if found, is added to obtain an overall score. These lexicons, such as Linguistic Inquiry and Word Count (LIWC), Hu and Liu, etc., are widely known and validated. For example, the Hu and Liu lexicon initially created in 2004 (Hu and Liu, 2004), included 2006 words with a positive orientation and 4783 with negative.

However, the SA in practice has become increasingly demanding and complex, applying more demanding procedures to express, not only whether a linguistic term is positive or negative, but also the intensity of its polarity. These valence-based lexicons, such as VADER (Hutto and Gilbert, 2014), ANEW (Nielsen, 2011), SentiWordNet (Baccianella et al., 2010) etc., use word lists with a valence score. For instance, VADER lexicon uses a sentiment-intensity on a scale from -4 to +4. Table 4 shows the orientation valence for words and emoticons comprising the VADER sentiment lexicon.

Table 4. Example words from VADER lexicon (valence-based) (source: Authors' own work).

term	**"okay"**	**"good"**	**"great"**	**"horrible"**	**"sucks"**	**" :(" or ☹**
sentiment intensity	+0.9	+1.9	+3.1	-2.5	-1.5	-2.2

Beyond the sentiment word intensity provided, there are sentiment lexicons, such as VADER, which identify properties and specific characteristics of the text, especially in short comments, which affect the perceived sentiment intensity of the text, as shown in Table 5. For example, the exclamation mark, adds more orientation intensity. "Excellent!" is stronger than "Excellent". In addition, it recognizes many sentiment-laden emoticons such as ":)", ":D" and sentiment-laden initialisms and acronyms such as "lol".

It is worth noting that many lexicons of the two aforementioned categories, polarity-based lexicons and valence-based lexicons, were created based on or by combining earlier lexicons.

In other cases, machine-learning approaches are incorporated to achieve automated sentiment-related tracing processes in the text, considering the knowledge base obtained. Kang H. et al. (2012) proposed an improved Naïve Bayes algorithm to solve problems related to classification performance. Chen and Tseng (2011) used the Support Vector Machine (SVM), a ML algorithm to classify reviews.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 5. Contribution of special characteristics to the intensity of polarity (source: Authors' own work).

Case	Example	Result
Typical writing	"The lecture was nice."	Positive, with a certain grade
Punctuation	"The lecture was nice!"	increase intensity
Capitalization	"The lecture was NICE."	increase intensity
Adverbs	"The lecture was extremely nice."	increase intensity
	"The lecture was marginally nice."	reduce intensity
Conjunctions	"The lecture was extremely nice but the attendance conditions	change of polarity, with dominant polarity the second
Negation	"The lecture wasn't nice."	reversal of polarity

Al Amrani et al. (2018) proposed a hybrid approach by making a combination of ML techniques, using Random Forest (RF) and Support Vector Machine (SVM) algorithms, to perform SA on comments from product reviews offered by Amazon. In another noteworthy study, Kabir M. et al. (2021) used different ML algorithms and concluded that Boosting and Maximum Entropy algorithm outperform the other examined ML algorithms for detecting sentiments in online user reviews. However, it should be noted that most machine-learning approaches have drawbacks that are mainly related to the requirement of having an extensive and representative training dataset. In addition, significant training and classification (Hutto and Gilbert, 2014) time is required.

2.2.5. Concluding remarks

Nowadays, with the rapid spread of data and its huge availability arising from various educational processes, the majority of universities frequently use data mining methods to examine data collected and to extract hidden knowledge from student feedback with the aim of improving teaching quality (Zeng et al., 2023; Mayordomo et al., 2022; Misuraca et al., 2021; Nguyen & Habók, 2023; Rybinski, 2023). However, the datasets are too large or complex to deal with using traditional data application software. Therefore, institutions are confronted with the challenge of how to tap into this large amount of data with the ultimate goal of enhancing student learning. These data generally consist of student perspectives regarding various aspects such as their attainment of a course's learning objectives, the effectiveness of instructional methods, the competence of the instructor, the content of the course, and even the students' overall views on their academic institutions. When taken into account, this data helps academic institutions in their curriculum design and into the management-related decisions. Further, it provides the teacher with the opportunity to better understand their teaching approaches, strengths and weaknesses, and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

subsequently they are better equipped to make appropriate changes, thus contributing to the overall effectiveness of their teaching and student learning.

The traditional techniques for student feedback analyses are based on close-ended questionnaire data collection and analysis (Richardson, 2005; Wong & Li, 2020). These forms, based on Likert-scale items of various types, provide quantitative data. However, the majority of academic institutions, such as our university, additionally offer students the ability to express their feedback opinions through open-ended questionnaires (text fields). According to previous research, this institutional educational text data is commonly collected but not fully exploited and remains untapped, thereby hindering the acquisition of knowledge that would contribute to the improvement of teaching and learning (McDonald et al., 2019). This is because unlike Likert-scale based data, which is quickly and easily analyzed using simple statistical methods, open-ended comments require special treatment and analysis. This creates a pressing need for more innovative approaches to effectively harness and utilize these untapped educational data resources.

This is where SA, a subfield of opinion mining, comes into play. SA involves the use of computational methods to identify and categorize subjective information within textual data. By applying SA to student feedback, institutions can gain deeper insights into various aspects of the educational experience, such as the attainment of course's objectives, the effectiveness of teaching methods, and overall student satisfaction with their academic environment. This technique allows for a more nuanced understanding of student perspectives, enabling targeted improvements in curriculum design and teaching strategies. In our research, we utilized SA not to challenge existing methodologies and approaches but to demonstrate the potential of SA in uncovering valuable insights from educational text data. By doing so, we hope to encourage the adoption of similar approaches across academic institutions.

In conclusion, the integration of SA into the evaluation of student feedback represents a promising frontier in educational research. By tapping into the rich, yet underutilized, source of textual data, academic institutions can make more informed decisions, ultimately leading to better educational outcomes. This research underscores the importance of embracing advanced analytical techniques to fully exploit the potential of educational data, fostering an environment of continuous improvement and innovation in teaching and learning.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

3. Research Design and Methodology

3.1. Introduction

This chapter thoroughly outlines the approach, design, and methods used to address the research problem. Before delving into our methodology, it is crucial to note the previously stated aims and objectives. In brief, our aim is to define a competency-based model for the quality of teaching from the perspective of processes in HE, with the goal of evaluating and enhancing academic excellence. The overall scope of this research is to establish a framework that encompasses both the competencies of teachers and the methods through which educational data can be utilized to measure these competencies. The key elements towards this aim are as follows:

- To endeavor to collect and synthesize previous research thoroughly in order to identify teaching staff competencies derived from the roles and tasks attributed to university professors. This entails identifying the competencies of teachers in HE across the board, ensuring a homogeneous frame and avoiding the inclusion of any domains of expertise that may require additional specialized skills.
- To explore and highlight the way in which large and complex educational textual data can be harnessed for our purposes, offering valuable insights to academic institutions.
- To implement techniques and approaches for extracting valuable information from student feedback.
- To evaluate specific dimensions of competencies of teachers in HE. This includes a comprehensive analysis of various datasets, aiming to provide a well-rounded assessment of the educators' strengths and areas for improvement.
- To create a coherent framework for validating our methods. This validation process is crucial to ensuring that the methods we develop are robust, scalable, and applicable across various educational settings.

The following section begins with an exploration of the research philosophy, providing a foundational understanding of the theoretical framework underpinning our study. Following this, we will describe the steps we followed, explaining each element of the process in detail. The chosen research strategy and methodology will then be outlined to illustrate how our approach is operationalized. Finally, the chapter will focus on the methods, tools, and techniques necessary for the research procedures and data collection in this thesis.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

3.2. Research philosophy

3.2.1. The nature of reality

Our research is based on the assumption that these competencies can be analyzed and measured through objective data, such as student comments.

From a realist perspective, teachers' competencies are considered real, independent of students' perceptions, and can be quantified using tools like SA and topic modeling. This approach supports the idea that these competencies have a stable essence that can be recorded and evaluated through systematic data analysis methods.

However, constructivism suggests that reality is shaped by social and personal perceptions. According to this view, teachers' competencies are not objectively given but constructed through the experiences and expectations of students. The comments we analyzed in our study may be influenced by personal and cultural factors, making teachers' competencies a dynamic and subjective concept. This highlights that, while our research focuses on measurement, we must also account for the subjectivity that students bring to their comments. Figure 4 illustrates the dual nature of reality in our study, highlighting the interplay between the objective measurement of teachers' competencies and the subjective perceptions shaped by students' experiences.

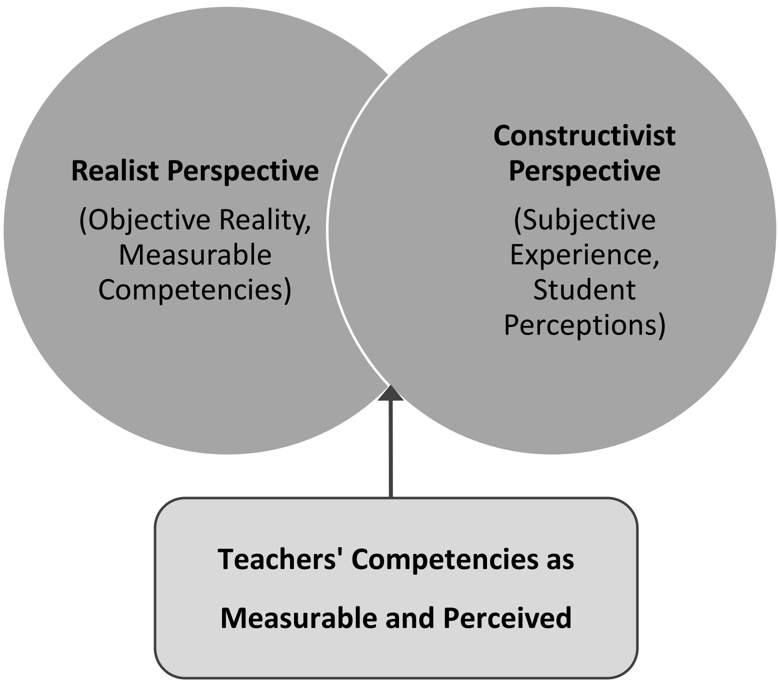


Figure 4. Pragmatic research framework (source: Authors' own work).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

It also emphasizes that, while teachers' competencies can be objectively measured (as suggested by the realist perspective), they are also shaped by students' subjective experiences and perceptions (as suggested by the constructivist perspective). This intersection represents the complexity of understanding teachers' competencies in our study, where both objective data and subjective insights hold significant value.

In summary, the ontological approach in our study recognizes both the need for objective measurement and the inevitable influence of subjective perceptions. This dual perspective is essential in our study, as it acknowledges the objective evaluation of competencies while also recognizing that students' feedback is influenced by personal and contextual factors (Lynham & Guba, 2018; Scotland, 2012).

3.2.2. Epistemology

In our study, the epistemological approach is based on collecting and analyzing student comments, aiming to highlight key teacher competencies. The knowledge we seek to gain relies on transforming subjective comments into objective data through tools like SA and TM, supporting a scientific approach to understanding teaching effectiveness (Bryman, 2016).

We follow a positivist epistemological approach, which assumes that knowledge can be achieved through scientific and measurable methods. Through this process, students' comments are converted into quantitative data, allowing us to draw conclusions about teachers' competencies. The choice of a positivist approach helps us identify and measure trends within the data, providing a more objective understanding of these competencies. The following diagram in Figure 5 outlines the epistemological framework of our research, demonstrating how subjective student comments are transformed into objective knowledge about teachers' competencies through systematic data analysis.

However, in our study, we also recognize that knowledge may not be entirely objective. The process of analyzing comments may obscure subjective information, such as students' emotional reactions or their social backgrounds. In this context, interpretivist epistemology would argue that understanding competencies requires interpreting students' personal and cultural contexts (Guba & Lincoln, 1994). Balancing these epistemological approaches ensures that our analysis, while grounded in data, remains sensitive to the subjectivity inherent in student feedback.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

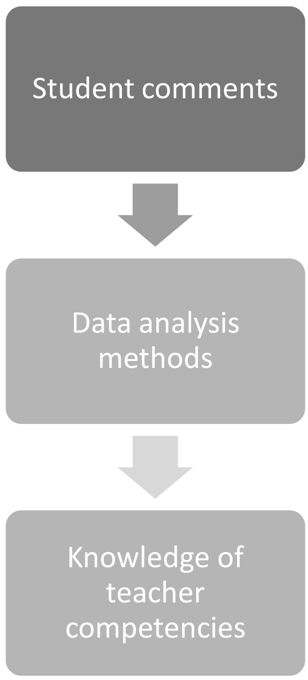


Figure 5. Knowledge acquisition in research (source: Authors' own work).

3.2.3. Axiology

Student values are central to this study, as the data examined originates from their comments on teachers' competencies. Therefore, it is essential to explore the role that values play in research and knowledge development. These values reflect the students' perceptions and expectations of the educational process and guide their comments.

From the researcher's perspective, axiology also concerns the decisions made in selecting methodological tools (Heron & Reason, 1997). In our study, we chose tools like SA and topic modeling because they reflect the value we place on objectivity and measurability. However, the very choice of these tools is based on values that recognize the importance of quantifying subjective comments to achieve a more objective understanding of teachers' competencies.

The values of both the students and the researcher coexist in this research and influence the outcomes. The analysis of student comments reflects the values they assign to the educational process, such as their attention to communication, understanding, and organization by the teacher. Therefore, the results of our research are based not only on objective data but also on value judgments that shape how we understand the educational reality. This is consistent with pragmatic research approaches that aim to produce practical knowledge that can be applied to real-world teaching contexts (Saunders, 2014). The following diagram in Figure 6 highlights the various values that shape our research methodology. It emphasizes the balance between the pursuit of objective

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

measurement and the recognition of students' subjective experiences in evaluating teacher competencies.

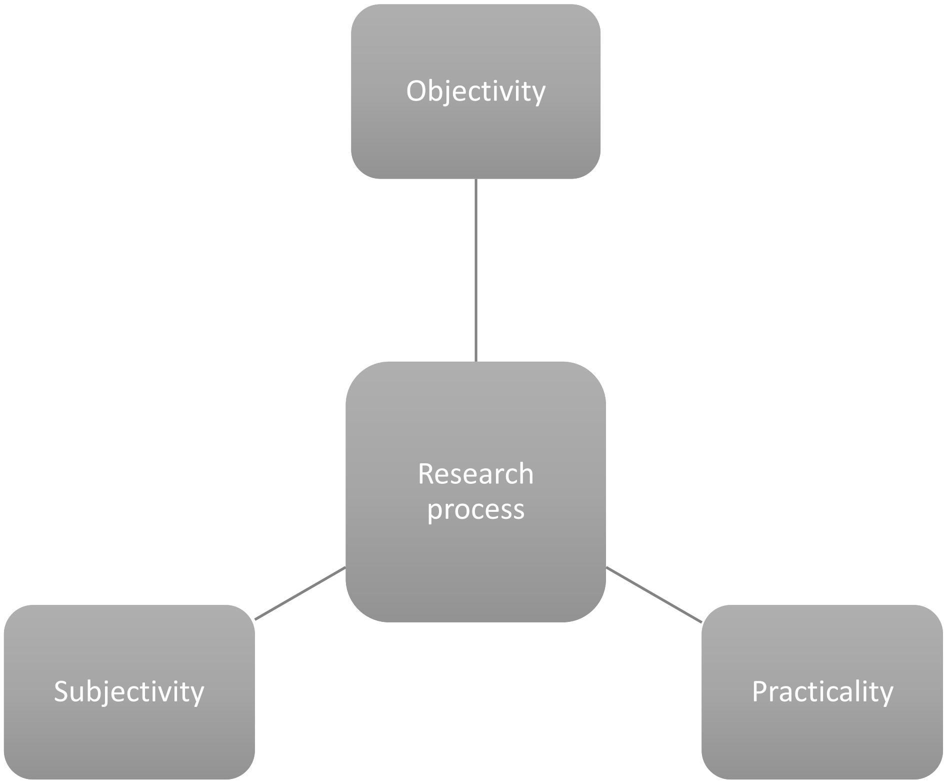


Figure 6. Values in educational research (source: Authors' own work).

3.2.4. Positivism vs. Interpretivism

The approaches of positivism and interpretivism are fundamental theoretical directions that influence the choice of methodological tools in research (Heshusius & Ballard, 1996). In our study, we have adopted a positivist approach, as we focus on quantifying and analyzing student comments through objective methods, such as SA and topic modeling. This approach allows us to draw clear, measurable conclusions about teacher competencies, giving us the ability to understand the key trends emerging from the data.

Interpretivism, on the other hand, argues that social reality cannot be fully analyzed using quantitative methods. According to this approach, student comments must be interpreted based on their social and personal context (Potrac et. al., 2014), as they include important subjective elements that cannot be captured through simple data analysis.

Although our research is primarily grounded in positivism, we recognize that the data studied—student comments—contain subjective elements that require careful interpretation. Our findings,

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

therefore, emerge from a combination of quantitative analysis and interpretive methods, which help us understand the complexity of the data.

The following diagram in Figure 7 illustrates the spectrum between positivist and interpretivist approaches in our research. It emphasizes how our study incorporates elements of both perspectives, aiming for a balanced understanding of teacher competencies through quantitative and qualitative methods.

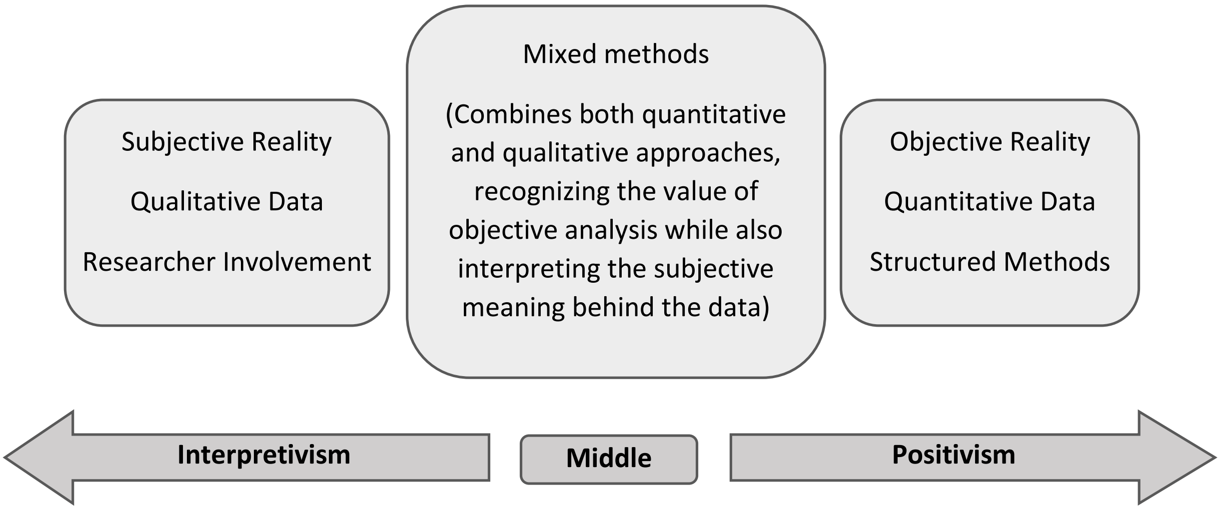


Figure 7. Positivism vs. Interpretivism (source: Authors' own work).

3.2.5. Pragmatism

Focusing on practicality and the use of research to achieve specific goals, pragmatism is central in our study, as our research aims to identify the competencies teachers should have and find effective ways to measure them. Pragmatism does not emphasize whether a method is purely quantitative or qualitative but rather the suitability of the tools used to achieve research objectives.

In our case, the methods chosen, including SA and topic modeling, were evaluated based on their ability to reveal meaningful insights about teacher competencies through student comments. This means that a pragmatic approach combines the best elements of both quantitative and qualitative methods, aiming to produce useful and actionable results. This pragmatic stance allows for flexibility and ensures that the research produces actionable insights that can be applied in educational settings (Tashakkori & Teddlie, 2010).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Pragmatism, therefore, underscores that our research is not limited to strict philosophical positions but aims at the practical utility of its results. The techniques we used are those that offer the best potential to achieve a reliable and applicable measurement of teacher competencies. For example, using BERTopic to validate the competency model based on student feedback was chosen for its ability to provide meaningful conclusions about the educational process. This approach makes pragmatism suitable for our study, as the ultimate goal is to produce knowledge that can be used for evaluating and improving teaching practices. The following diagram in Figure 8, encapsulates the essence of pragmatism in our research. It underscores the focus on practical utility and flexibility in methodology, illustrating how our chosen techniques serve to produce meaningful and applicable results in evaluating teacher competencies.

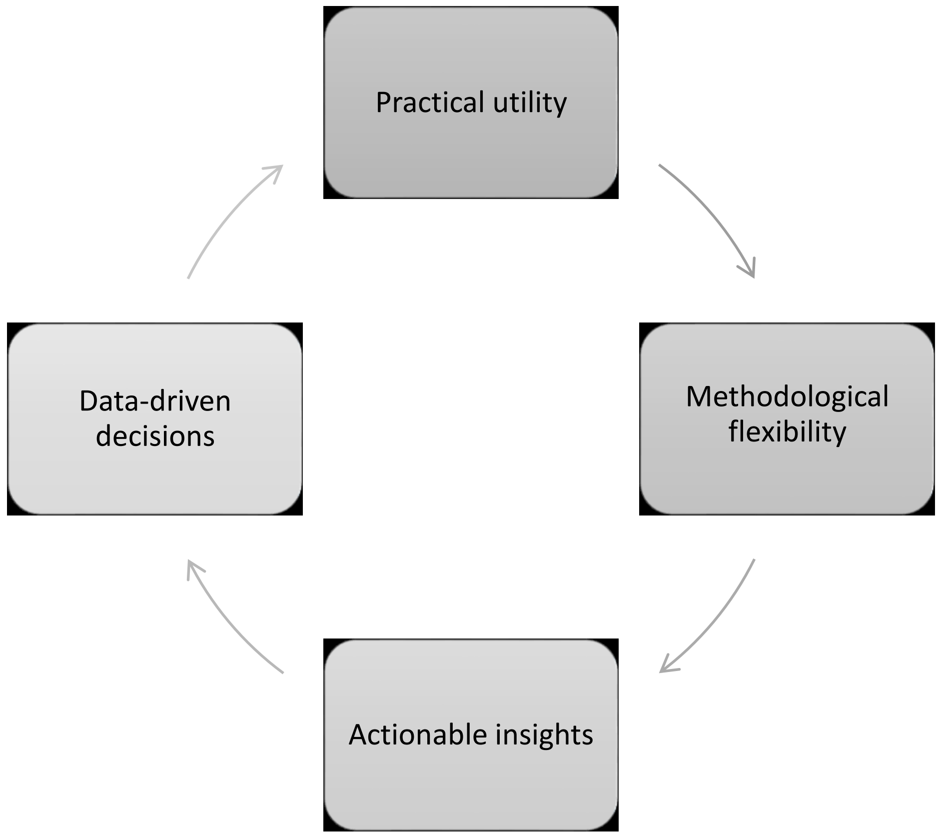


Figure 8. Pragmatic research framework (source: Authors' own work).

Finally, pragmatism allows flexibility in selecting and using tools to meet research goals in the most effective way. In our research, this philosophical approach guides our methodological choices, as they focus on the practical value of the results, which can be applied in real-world settings for evaluating and developing teachers. Essentially, pragmatism in this dissertation highlights the

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

importance of using the right tools to achieve useful and actionable results, regardless of whether these tools are traditionally quantitative or qualitative.

3.3. Research design and methodology

In this section, we provide a comprehensive overview of the stages involved in the research process we followed. A visual depiction of these stages is presented below. Specifically, we divided the overall process into three phases, as illustrated in Figures 9, 10, and 11, respectively. For better clarity and coherence, each stage of each phase is described in detail in the following subsections. We chose to depict the overall methodology using a flowchart for each phase, rather than a linear representation to emphasize the parallel nature of the procedure, particularly regarding the detailed understanding of our research design.

3.3.1. Detailed framework of the initial research process: From objectives to identifying gaps (phase 1)

The first phase of this research aimed to establish a clear and structured approach to investigating teacher competencies in HE and the role of student feedback in assessing these competencies. This phase was critical in laying the groundwork for the subsequent research stages, as it provided a systematic pathway from defining objectives to identifying specific gaps in the existing literature. By structuring the initial research process into distinct steps, a logical flow was established, as depicted in Figure 9, ensuring that the study's objectives were comprehensively addressed and that all necessary groundwork was completed before advancing to more complex methodologies.

The focus during this phase was placed on reviewing and analyzing relevant literature to both identify the core competencies required of teachers in HE and to explore the potential of student feedback as a data source. The detailed steps followed during this phase were designed to ensure that the study would be grounded in existing research, while also identifying opportunities for novel contributions to the field.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

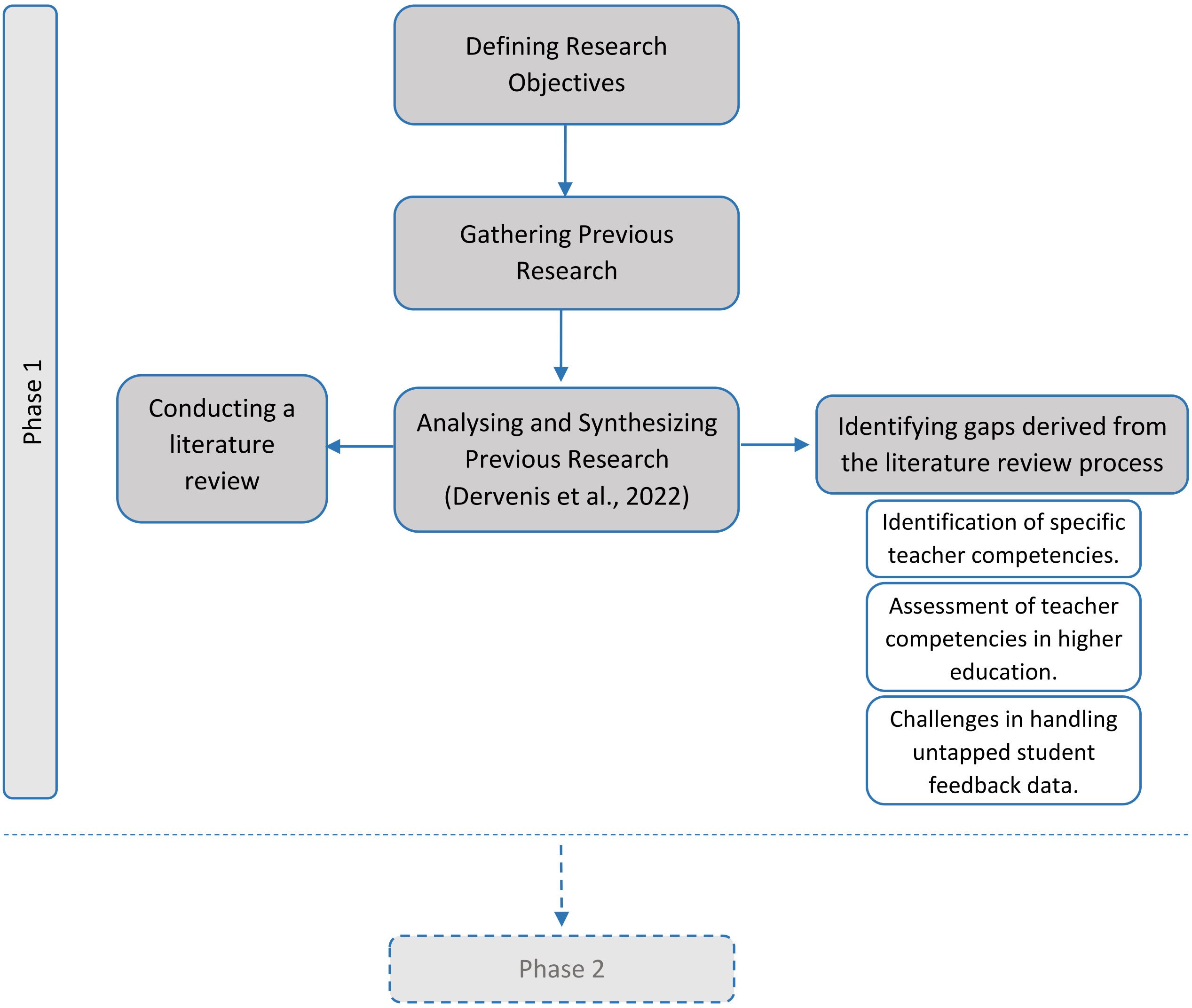


Figure 9. Framework for initial research process: from objectives to identifying gaps (phase 1) (source: Authors' own work).

First, the research objectives were defined to guide the overall direction of the study. Then, an extensive literature review was conducted to gather a broad base of previous research related to teacher competencies and feedback analysis. This review allowed for the synthesis of existing knowledge and the identification of key gaps, which informed the focus areas for the research. Specifically, these gaps highlighted the need for more targeted research into competencies specific to higher education, as well as the challenges of effectively utilizing untapped student feedback data.

The analysis conducted in this phase also set the stage for the development of the methodology employed in later stages. By identifying the competencies most pertinent to HE and the potential challenges associated with utilizing student feedback as a data source, a foundation was established for the implementation of advanced analytical techniques, including SA and machine learning. This

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

systematic approach ensured that the study would be both theoretically sound and practically relevant, addressing key issues in the field of educational research.

The following subsections provide a detailed explanation of each step within this initial phase, outlining how the objectives were defined, how the literature was reviewed and synthesized, and how the specific gaps in the research were identified. Each of these steps played a crucial role in shaping the overall direction and scope of the research.

3.3.1.1. Defining research objectives

The initial step in our research process involved defining the research objectives. These objectives were shaped by the overall aim of understanding and identifying key teacher competencies in higher education, and determining how student feedback could offer insights into these competencies. The focus was placed on investigating the specific competencies required for effective teaching in HE and exploring how student feedback could serve as a measurable indicator of these competencies. This step established the foundation of the research, ensuring that the competencies being studied were aligned with current educational needs and the available data.

3.3.1.2. Gathering previous research

Following the definition of research objectives, a range of relevant literature was gathered, including empirical studies, theoretical frameworks, and reviews related to teacher competencies and student feedback in higher education. Emphasis was placed on sourcing research from academic journals, databases, and books that addressed competency models, educational effectiveness, and feedback analysis. The goal at this stage was to create a comprehensive literature base that would inform the understanding of the subject and provide context for identifying gaps and areas requiring further research.

3.3.1.3. Analysing and synthesizing previous research

A thorough literature review was conducted to analyze and synthesize previous studies. The review focused on extracting key insights and identifying existing models, frameworks, and theories regarding teacher competencies and student feedback. Attention was given to how competencies were defined and assessed in different contexts, as well as the methodologies used for analyzing feedback data. This review provided a broad understanding of the current knowledge in the field and shaped the direction of the ongoing research. The literature review revealed several important gaps, which informed the subsequent phases of research. It was noted that while multiple competency frameworks were present in the literature, there was limited consensus on a core set of competencies specific to higher education. Additionally, many studies focused on generic teaching competencies, with fewer addressing competencies tailored to the HE context. This

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

highlighted the need for further research to define and investigate the specific competencies that are essential in this setting. The analysis also revealed a lack of consistent methods for assessing teacher competencies, particularly through student feedback. This gap suggested the potential to explore student comments and SA as methods for evaluating teacher competencies.

Finally, the literature pointed to challenges in utilizing student feedback data effectively. Although vast amounts of feedback were available, there was limited research that systematically analyzed this data, particularly through techniques such as ML or natural language processing. This presented an opportunity for the current research to apply SA and TM to unlock the valuable insights embedded in untapped student feedback, offering new perspectives on teacher performance and effectiveness.

3.3.2. Methodological framework for analysing educational text data: Approaches and techniques (phase 2)

The second phase of our methodology focuses on the investigation and analysis of educational data, particularly textual data derived from student feedback (see Figure 10). This phase is crucial for understanding how insights can be extracted from complex educational texts and how various analytical techniques can be applied to interpret these data effectively. By employing a methodological framework tailored to educational text data, this phase seeks to uncover patterns, sentiments, and themes that may inform our understanding of teacher competencies and their impact on student experiences.

This phase is structured into three main steps: investigating and analyzing educational data, applying text data analysis techniques, and comparing different analytical approaches. Each of these steps plays a vital role in ensuring that the research not only draws meaningful insights from the data but also utilizes a variety of methods to validate findings. This comprehensive approach allows for a nuanced understanding of the data and enhances the robustness of our research outcomes.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

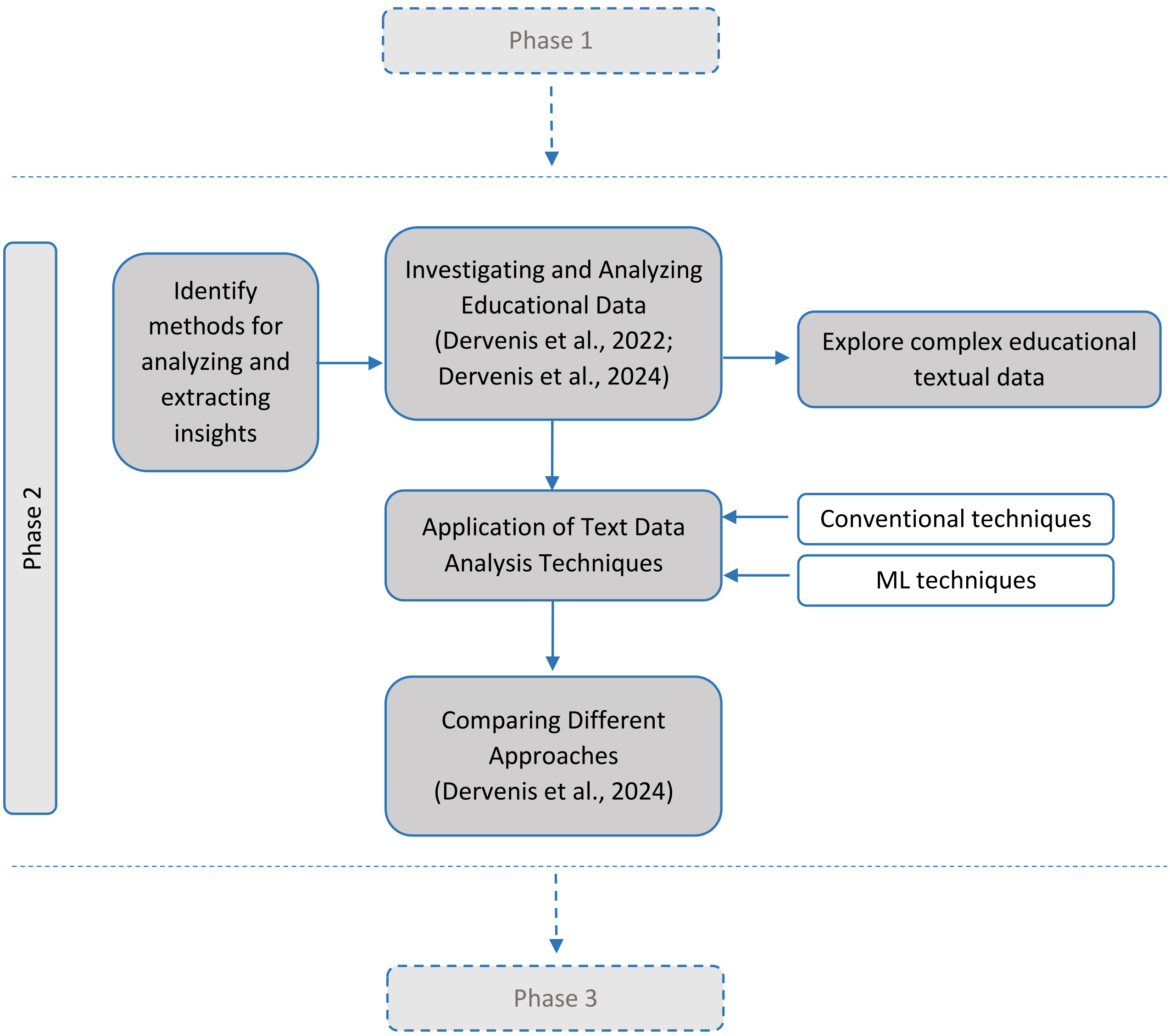


Figure 10. Methodological framework for analyzing educational text data (phase 2) (source: Authors' own work).

3.3.2.1. Investigating and analysing educational data

In this stage, the focus is on identifying appropriate methods for analyzing educational data, particularly in relation to student feedback. Various analytical approaches are considered, including qualitative and quantitative methods, to ensure a holistic understanding of the data. Techniques such as content analysis, topic modeling, and SA are evaluated for their potential to extract valuable insights from the data. By assessing these methods, a clear strategy is developed to systematically analyze the feedback and derive meaningful conclusions about teacher competencies.

Following the identification of analytical methods, the next step involves exploring the complexities inherent in educational textual data. This includes understanding the diverse formats and contexts of the data collected, as well as the various factors that can influence student feedback. The challenge lies in addressing the nuances of language, tone, and sentiment expressed

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

in student comments, which may vary significantly. This exploration not only informs the analytical techniques that will be applied but also highlights the importance of considering the context in which feedback is provided. By delving into these complexities, the research aims to uncover deeper insights that may not immediately be apparent.

3.3.2.2. Apply text data analysis techniques

In the second step of this phase, both conventional and ML techniques are applied to the educational feedback data. The conventional approach includes SA based on lexicons, where predefined lists of words and their associated sentiment scores are utilized to gauge the overall sentiment of student comments. This method allows for the systematic classification of feedback into positive, negative, or neutral categories, providing a foundational understanding of the students' perceptions. Additionally, manual coding is employed to identify themes and patterns through a thorough review of the comments, along with traditional statistical methods that quantify specific aspects of the feedback, such as frequency analysis to pinpoint commonly mentioned sentiments or themes.

At the same time, machine learning techniques are leveraged to enhance the analysis further. Advanced computational methods, such as NLP and machine learning algorithms, are employed to categorize feedback, assess sentiment, and extract topics from the textual data. Unlike conventional techniques, machine learning approaches utilize algorithms that can learn from data, improving their accuracy and ability to detect complex sentiment patterns within larger datasets. By incorporating both conventional lexicon-based SA and machine learning techniques, the research enhances the analytical depth, ensuring that key trends and insights are highlighted.

3.3.2.3. Compare different approaches

The final step of this phase involves comparing the effectiveness of the various analytical approaches employed throughout the research. This comparative analysis evaluates the strengths and limitations of conventional and machine learning techniques, assessing how each method contributes to the overall understanding of the educational data. By identifying which approaches yield the most significant insights, this step informs the research methodology for subsequent phases and highlights the best practices for analyzing educational text data. Ultimately, this comparative analysis aims to refine the research approach and ensure that the findings are robust and reliable.

3.3.3. Model development and validation: Methodological framework (phase 3)

The third phase of the research methodology focuses on model development and validation, which is essential for translating the insights gained from the previous phases into actionable

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

frameworks. This phase aims to create a robust model that can effectively assess teacher competencies based on student feedback and provide a foundation for future research and practical applications in educational settings. By utilizing various analytical techniques, including SA and topic modeling, this phase seeks to ensure that the developed model is both accurate and reliable. This phase comprises four critical steps, as illustrated in Figure 11, model creation and implementation, development of a validation framework, extraction and analysis of the final results and final presentation and conclusions.

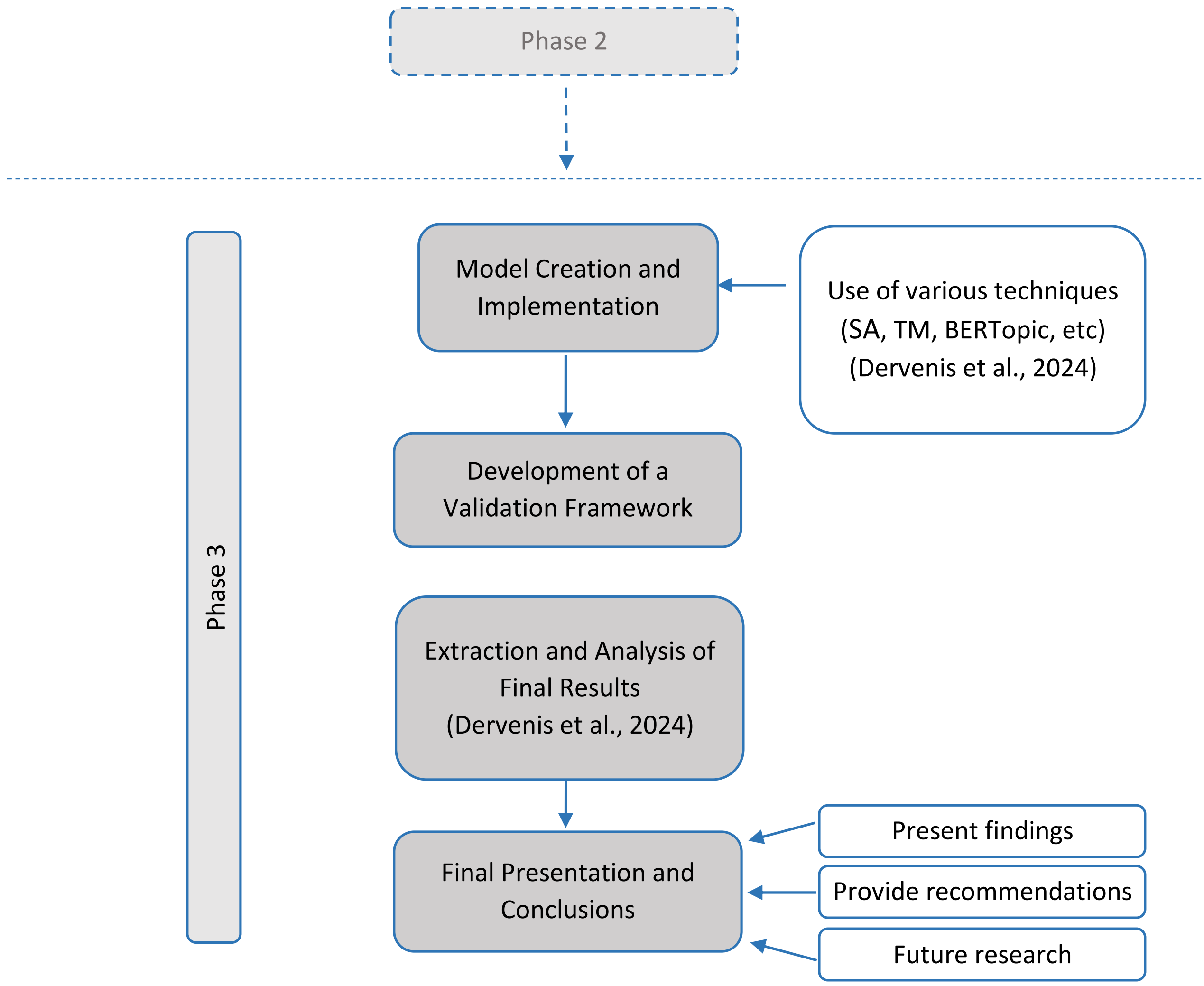


Figure 11. Model development and validation (phase 3) (source: Authors' own work).

Each of these steps plays a vital role in the overall research process, ensuring that the model not only meets the defined objectives but also addresses the complexities of educational data. This comprehensive approach to model development enhances the validity and practical implications of the research, providing valuable insights for stakeholders in the education sector.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

3.3.3.1. Model creation and implementation

In this initial step of this phase, a variety of techniques are employed to create and implement the model that assesses teacher competencies. SA is applied to understand the emotional tone of student feedback, enabling the identification of areas where teachers excel or require improvement. Additionally, TM techniques, such as Latent Dirichlet Allocation (LDA) and BERTopic, (BERT stands for Bidirectional Encoder Representations from Transformers) are utilized to uncover underlying themes in the data. These methodologies allow for a more nuanced interpretation of student comments, highlighting the competencies that resonate most with learners.

The model's implementation involves the careful integration of these techniques, ensuring that they work in concert to produce a comprehensive assessment of teacher performance. By combining qualitative insights from SA with quantitative data from topic modeling, the model aims to provide a holistic view of teacher competencies in higher education. This phase also emphasizes the importance of iterative testing and refinement, ensuring that the model evolves based on feedback and findings from the analysis.

3.3.3.2. Development of a validation framework

The validation framework was a key element in ensuring both the relevance of the competencies identified through the literature review and the effectiveness of the methods used to measure these competencies. The purpose of this validation was twofold:

1) to assess whether the competencies identified from the academic literature reflect the competencies that students value in their professors, and
2) to confirm that the approaches used to measure these competencies through SA, topic modeling, and other text analysis techniques are accurate and meaningful.

The key question was whether the competencies recognized as significant by the academic community—such as pedagogical skills, communication abilities, and subject knowledge—are similarly valued by students in their feedback. Through this process, we aimed to determine, if the academic model aligns with students’ perceptions. If students consistently highlight these same competencies in their evaluations, it provides strong evidence that the model is sound and representative of their actual concerns. The chart provided in Figure 12 illustrates the process followed for model validation.

However, the validation process extended beyond competent matching. It also critically examined the measurement framework used to analyze student feedback. By employing methods such as lexicon-based SA and ML models, we sought to measure and quantify these competencies from the students' comments. A key part of the validation was to confirm whether these tools and

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

techniques were effective in capturing the nuances of student perceptions and if the derived measurements were reflective of the competencies in question. For example, if the analysis techniques consistently identify sentiment related to a professor's communication or teaching style, and these align with the competencies highlighted in the literature, it demonstrates that the measurement methods are working effectively. Additionally, the process sought to reveal whether some competencies that may be undervalued in the literature—such as interpersonal skills or approachability—are nonetheless captured by the student feedback and recognized through the analysis framework. This would indicate that the chosen methods are robust and capable of identifying a broader range of competencies that may not have initially been emphasized in academic discussions.

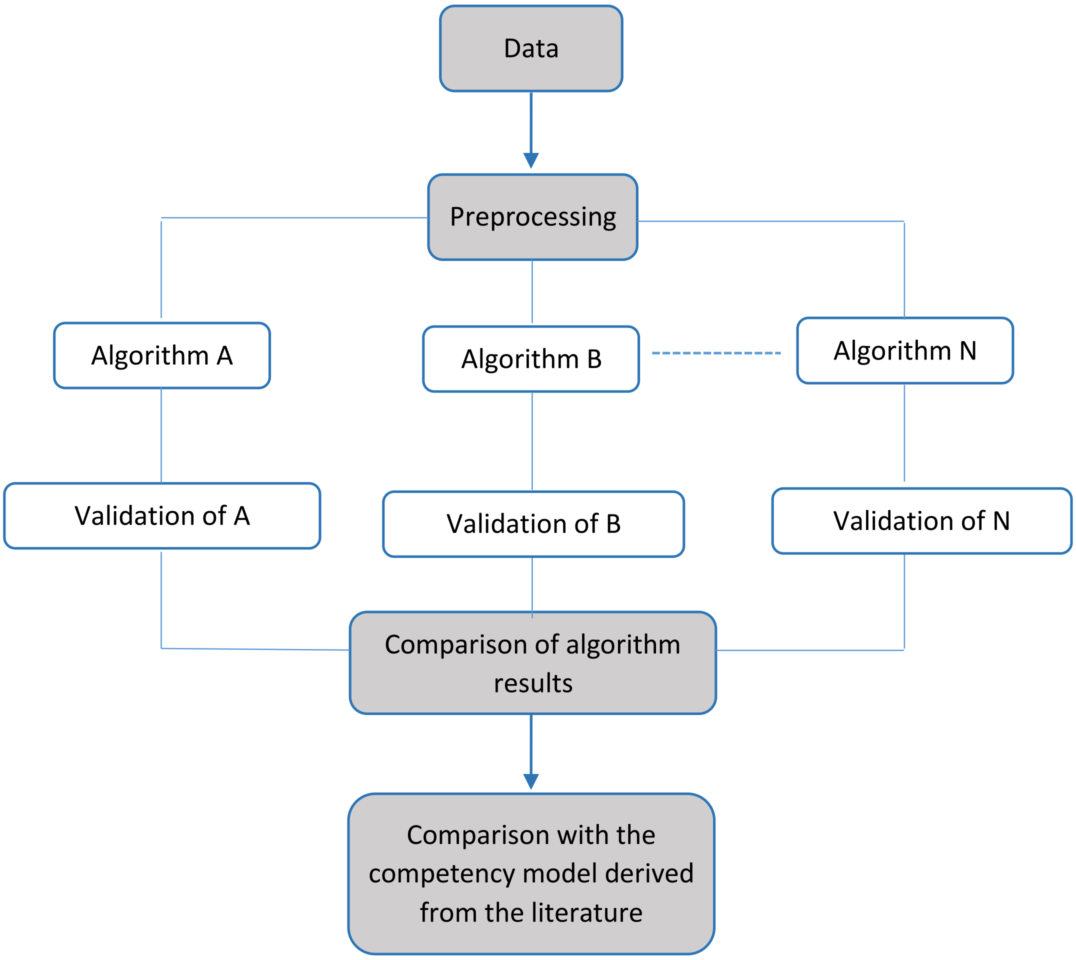


Figure 12. Model validation (source: Authors' own work).

Ultimately, the goal of this validation framework was not only to verify that the competencies identified through the literature review are important to students, but also to ensure that the

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

analytical tools and techniques used to measure these competencies from textual data are valid. If the results show that students' feedback consistently highlights competencies aligned with the literature, and the methods used are able to accurately measure these competencies, it confirms the value of the entire framework. This would demonstrate that the system is both theoretically sound and practically useful in evaluating teaching performance through student-generated data. In conclusion, the validation framework ensured the credibility of both the competency model and the measurement methods used to analyze and assess these competencies from student feedback. By validating both what is being measured and how it is being measured, the framework supports the effectiveness of using student comments as a reliable tool for evaluating teacher competencies in higher education.

3.3.3.3. Extraction and analysis of results

The extraction and analysis of the results was a crucial step in understanding both the competencies of professors as identified in the literature and how these competencies are reflected and measured through student feedback. This phase not only involved identifying which competencies emerged from the data but also examined how the students' comments were effectively utilized to measure and assess these competencies. The key aspect of this stage was to extract meaningful insights from the student feedback using the analytical techniques applied earlier—such as SA, topic modeling, and other text data analysis methods. These techniques allowed for the quantification of competencies and provided a structured way to interpret student feedback. For example, SA identified whether students expressed positive or negative sentiment regarding specific competencies such as communication skills, subject knowledge, or teaching style. Additionally, TM uncovered key themes and topics within the comments that correlated with the competencies derived from the literature.

Additionally, this phase also explored whether the students’ feedback emphasized competencies that might not have been the primary focus in the academic literature. In some cases, students might prioritize interpersonal skills, fairness, or accessibility, which were underrepresented in academic models of teaching competencies. The ability to capture these competencies through text analysis methods demonstrated the flexibility and depth of the approach, showing that the student feedback data could reveal both anticipated and unanticipated insights. Moreover, the analysis examined how effectively the student comments could be transformed into actionable metrics for assessing teacher performance. The results of the analysis indicated not only whether students aligned with the competencies recognized by the academic community, but also how reliable and useful student feedback was for measuring these competencies. For instance, if certain competencies were frequently mentioned in the feedback with consistent sentiment, it confirmed that these were aspects of teaching that mattered most to

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

students. This consistency validated the use of student comments as a legitimate and valuable source of data for competency measurement.

The extraction and analysis process also served to identify any gaps or areas where the current methods might fall short in fully capturing student perceptions. This step ensured that the final results provided a comprehensive understanding of how well the analytical framework measured the competencies in question. By leveraging student feedback data, this phase demonstrated the practical value of using such data to evaluate teaching performance and informed recommendations for future improvement in both the analysis techniques and the competency model itself.

In summary, the final results provided twofold insight: a deeper understanding of the teacher competencies that matter to students, and confirmation that the methods used to analyze and measure these competencies through student feedback were effective. This outcome reinforced the practical utility of using student-generated data to assess teacher performance in higher education, bridging the gap between theoretical competency models and real-world student expectations.

3.3.3.4. Final presentation and conclusions

The final step of this phase focuses on the presentation of findings, providing recommendations for practice and suggesting areas for future research. The results were communicated through clear and concise reports, visualizations, and presentations, ensuring that the insights remained accessible to a wide audience, including educators, administrators, and policymakers.

Recommendations derived from the findings emphasize actionable strategies for enhancing teacher competencies and improving student feedback mechanisms. Furthermore, this step identifies gaps in the current research and proposes avenues for future exploration, encouraging continued inquiry into the intersection of teacher competencies and student experiences. By concluding this phase with a comprehensive presentation of findings, the research aims to foster informed discussions and decisions within the educational community.

3.4. Methods and tools

In this section, we present the methods and tools that were considered and ultimately chosen to carry out the steps of the methodology described earlier in this chapter. The purpose of this selection process was to find the most suitable approaches for analyzing the data and drawing meaningful insights. The choices made were informed by both the literature review and the specific

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

needs of this research, ensuring that the selected tools effectively supported the analysis and helped achieve the research goals.

3.4.1. Methods and tools used for the literature review

In this subsection, we delve into the methods and tools that were utilized during the literature review phase of the research. This crucial phase adopted a systematic approach to gathering pertinent studies and articles, ensuring a comprehensive exploration of the subject matter. Initially, to obtain a holistic understanding of the relevant field, a systematic review was conducted following the PRISMA methodology (Moher et al., 2009).

To facilitate this process, an extensive search was performed across several key databases, including ACM Digital Library, IEEE Xplore, Springer, Elsevier, and Google Scholar. This multi-database approach was essential for capturing a broad spectrum of literature related to the topic. Boolean operators were used in refining our search queries, allowing us to combine various search terms effectively. Additionally, proximity operators (Hammerstrøm et al., 2009) were used to locate terms within a specific range, ensuring that the searches remained comprehensive and relevant. The search terms included a variety of concepts such as academic staff, higher education, faculty, tertiary education, professor quality, profiles, competencies, skills, and quality teaching.

Using these terms, we crafted targeted queries that yielded results closely aligned with the objectives of our review. Throughout this process, results that were pertinent to the research purpose were meticulously examined, while irrelevant outcomes were systematically filtered out in subsequent stages. The search queries were executed between February and March 2021, ultimately generating a total of 102 results that served as the foundation for our analysis. Figure 13 illustrates the process that was performed based on the template of PRISMA 2020 flow diagram for new systematic reviews (Page et al., 2021) showing the final phase where our bibliographic research included the papers that met the search criteria.

Following this, we focused on identifying and categorizing all competencies cited in the literature available to us. We carefully examined the intrinsic relationships between the competencies we identified, grouping those with related characteristics or shared elements. Our objective was to differentiate, gather, and then assign these competencies to specific dimensions relevant to HE teachers as discussed in the literature.

Additionally, we found it necessary to introduce two indicators. The “appearance” indicator, $Apr_{(x)}$, represents the proportion of each competency dimension that appears in the literature we reviewed, while the “coverage” indicator, $Cvr_{(x)}$, reflects the percentage of each dimension covered in eligible studies—specifically, the extent to which the competencies in each dimension were

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

represented in the selected sources. With this method, we gained a clear view of not only what was referenced in the literature we examined but also the degree to which each aspect was covered.

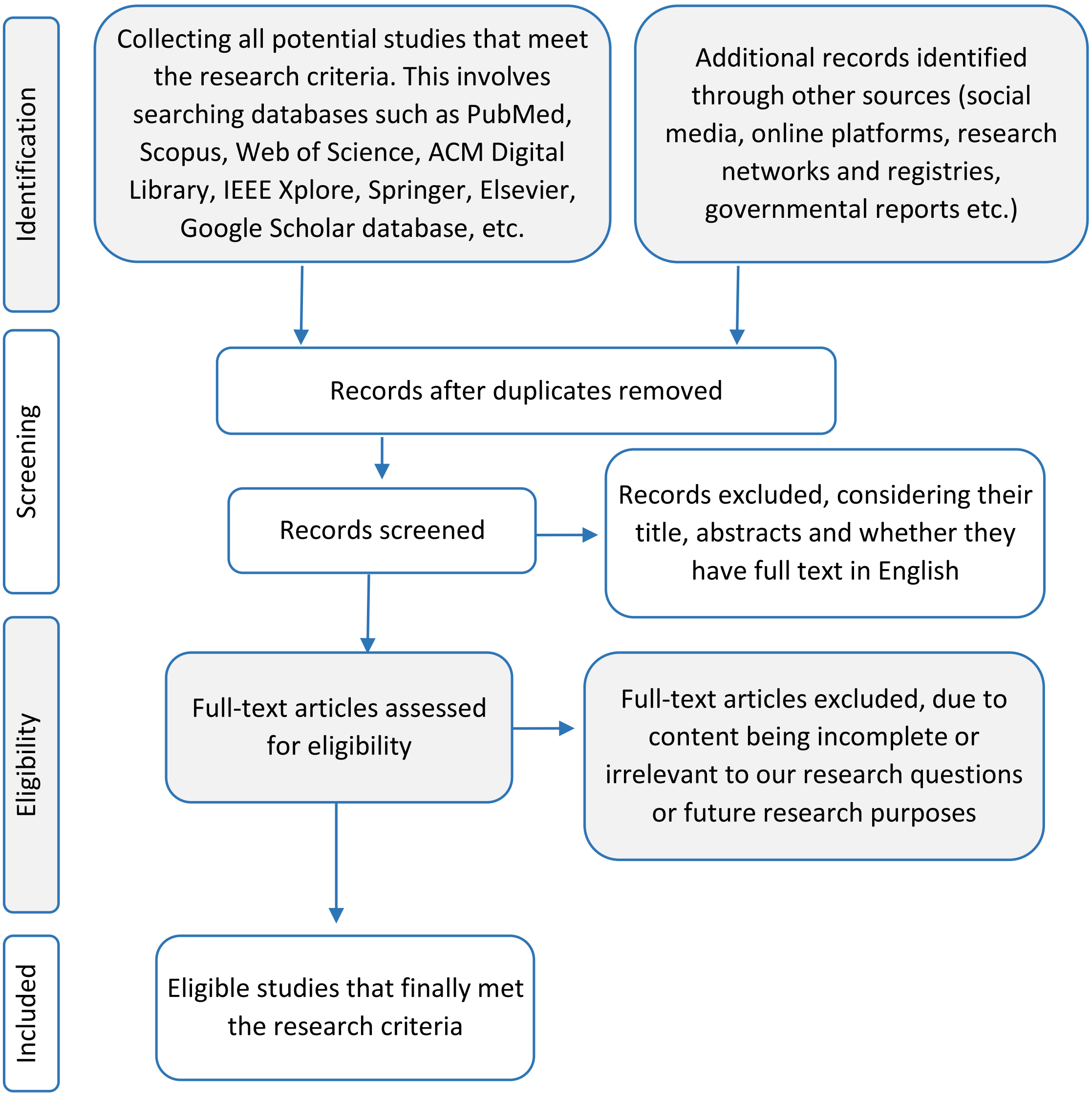


Figure 13. PRISMA statement flowchart (Page et al., 2021).

3.4.2. Data collection methods

In our research, we utilized textual data derived from student feedback at our university, collected through both a structured, standardized feedback process and a customized survey designed to gather specific insights relevant to our research. This additional survey enabled us to obtain focused responses directly connected to our objectives. In addition to structured sources, we also incorporated opportunistic data—feedback gathered informally through natural student

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

interactions (oral discussions) and open-ended survey responses. The categorization of our data as "opportunistic" stems from the way it was collected: it includes feedback gathered informally through natural student interactions, rather than within a strictly structured framework (Ferragina & Manzini, 2000). This allowed us to capture spontaneous, unfiltered comments that arose from open-ended survey questions specifically designed to gather insights relevant to our research objectives.

By incorporating opportunistic data, our study benefited from the addition of more authentic viewpoints, often revealing perspectives that structured data collection methods might overlook. This approach enriched our analysis by providing a broader, real-world view of the competencies we aimed to study. This multi-source approach not only enriched the depth of our analysis but also allowed for the inclusion of natural data points broadening our understanding of the competencies under study (Willig, 2019).

Meaningful comparative evaluations of algorithms yield robust results when tested on consistent datasets using identical performance metrics (Witten et al., 2016). Generally, data used for testing can originate from three primary sources: actual domain-specific data, pre-generated datasets, and randomly generated data. Actual (or real-world) data is drawn directly from cases within the study’s application field, such as the structured and opportunistic student feedback used in our research. Pre-generated data, by contrast, is collected from standardized datasets or repositories, while randomly generated data can be used when there is insufficient real data available for comprehensive training and testing.

Data preparation should be carefully aligned with the evaluation goals and minimize issues such as:

- Insufficient sample size: Larger, diverse samples improve experimental reliability and provide clearer insights into the performance of tested algorithms. As noted by Hattie and Timperley (2007), sufficient sample sizes are essential in educational research to obtain valid insights into the effectiveness of interventions and algorithms, ultimately leading to improved educational outcomes.
- Homogeneity in test set features: A lack of variety may obscure an algorithm’s weaknesses. As Cawley and Talbot (2010) explain, homogeneity in the features of the test set can lead to an incomplete evaluation, emphasizing the need for diversity to identify potential weaknesses in algorithm performance.
- Incorrect or unknown labels: Inaccurate labeling leads to poor model training and suboptimal deployment performance. Zhang and Wallace (2015) highlight that incorrect

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

or unknown labels can severely impact the training process, underscoring the necessity for precise labeling to achieve effective machine learning outcomes.

- Non-representative training data: Training on data that doesn't represent the actual problem domain results in a model that may underperform in real-world scenarios. According to Kohavi and Provost (1998), utilizing non-representative training data can hinder a model's ability to generalize, emphasizing the necessity for data that accurately reflects the problem domain for optimal performance.

Our study's data, which includes both structured student feedback and opportunistic data sources, provides a robust real-world dataset. By integrating structured responses and spontaneously gathered feedback, we enhanced the dataset's relevance and reliability, allowing for a more comprehensive and nuanced analysis of HE teachers' competencies.

3.4.3. SA method

Sentiment Analysis (SA) can be referred to by various names in the literature. It may be mentioned as subjectivity analysis, opinion mining, or appraisal extraction (Misuraca et al., 2021). According to Liu, (2012), "Sentiment analysis is the field of study that analyzes people's opinions, emotions, evaluations, appraisals, and attitudes towards entities such as products, services, organizations, individuals, topics, events, and their attributes.". Essentially, SA refers to the process of taking a text or text snippet as input and extracting valuable information related to the expressed sentiments within that text. This process is typically performed on a collection of texts referred to as a corpus.

SA plays a significant role in many applications of business intelligence, such as credit rating assessment or organizational reputation evaluation. It is essential for the development of analytics that aim to assess the limitations and shortcoming of offered products and services in order to improve their quality. Furthermore, SA is one of the most prominent examples of descriptive analysis. As a research field, SA can be considered part of computational linguistics and NLP, which involves the interaction between computers and human (natural) language (Saraswat et al., 2020; Mejova, 2009). It employs techniques and methods from machine learning, NLP and statistics, with the goal of classifying text into sentiment classes. The sentiment polarity is one of the desired properties of the text being processed. It constitutes the main methodological characteristic of SA and is commonly categorized into three classes: Positive, Neutral and Negative (Garg, 2020; Liu, 2012; Chakraverty & Saraswat, 2017; Liu, 2012; Wilson et al., 2009).

However, this does not imply that binary classification (positive, negative) or multiclass classification with more categories, (i.e., Strongly Positive, Positive, Neutral, Negative, Strongly Negative) cannot be applied. The number of classes used for text categorization is dictated by the

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

specific problem at hand. Furthermore, besides polarity identification, there are more sophisticated techniques that recognize emotional states in the text, such as "joy," "sadness," or "anger." Finally, it is important to distinguish between polarity and intensity (strength) with which the sentiment is expressed in the text. For example, one might have a positive opinion about a professor but not hold a strong opinion since they have not attended their class for a significant period of time.

SA can play a pivotal role in various facets of the learning process. Not only does it allow educators to receive feedback on their instructional methods, but it can also facilitate the identification of students requiring extra support (Dervenis et al., 2024; Altrabsheh et al., 2013; Poulos & Mahony, 2008; Wen et al., 2014). The state of the art in SA distinguishes three dominant approaches aimed at addressing this concern. These approaches are briefly yet concisely presented in the following section.

3.4.3.1. Algorithms and techniques used in SA

In our research, we focus on Random Forest (RF) as well as on the Artificial Neural Networks (ANNs) (Shaik & Srinivasan, 2019; Babu & Kanaga, 2022). RF is a powerful supervised learning algorithm that excels in classification tasks. The first algorithm, as implied by its name, consists of an ensemble of decision trees. Each individual decision tree in the RF makes a class prediction, and the class with the most votes becomes the final prediction of the model. The fundamental idea behind the method is that a large number of uncorrelated models functioning as a committee will outperform any of the individual constituent models. The low correlation among the models is key, as the uncorrelated models can produce predictions that are more accurate than any of the individual predictions (Dervenis et al., 2022). This is because the trees compensate for each other's errors through their overlapping nature. Furthermore, unlike individual decision trees, the decision trees belonging to a RF are constructed by randomly selecting predictor variables based on which the splitting of nodes in the tree occurs. These characteristics, combined with the absence of pruning, result in lower correlation among the trees and, therefore, greater diversity among them. RF, and subsequently decision trees, belong to the category of nonparametric ML algorithms.

Artificial neural networks are a form of ML algorithm that seeks to emulate, to the extent feasible, the operation of the human brain to solve complex problems. The mathematical model underpinning this particular method strives to replicate the structure of the biological neural network, thereby acquiring characteristics that enable parallel data processing and the execution of tasks that are challenging to describe, such as pattern recognition. ANNs represent a relatively new area of research that began its development on an international level a few decades ago, around 1980. Initially, McCulloch & Pitts (1943), presented a model of a neural network based on neurons and their connections. This model forms the basis of AI neural networks, which have the

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

perceptron as their basic structural unit and "learn" through processes that modify the weights of their connections (synaptic weights). After the learning process, the network can undergo the recall process and compute its output for a given input vector. In other words, it applies the learned model. The use of these techniques requires a large volume of data, and most of them rely on correctly labeled data during the training process. In contrast, lexicon-based techniques can avoid this limitation.

The lexicon-based approach we employed utilizes a lexicon that associates words with corresponding sentiment weights, capturing the meaning attributed to each word. Depending on the implementation, weights can take two values (positive/negative sentiment) or belong to a range of values. The polarity values of each word in the text are transmitted to an algorithm to generate the final sentiment score, which represents the sentiment conveyed by the text. We can distinguish two main approaches in the application of Lexicon-Based methods based on the way sentiment bearing words are identified during the creation of the lexicon (Kausar et al., 2019). The first approach is the Lexicon-based approach, in which lexicons can be created either manually or automatically using seed words, typically adjectives, as indicators of the semantic orientation conveyed by the text. These seed words are expanded with their synonyms to enrich the initial lexicon as much as possible. The implementation of this technique can be further expedited by utilizing a preexisting lexicon from a variety of resources available for sharing, primarily in the English language. Examples of such lexicons include NRC Emotion Lexicon, MPQA, WordNet, SenticNet, SentiWordNet, WordNet-Affect, and others (Pradhan et al., 2023; Mohamad S. & Mohamed A., 2022; Beckwith et al., 2021; Hardeniya & Borikar, 2016). Furthermore, we have the corpus based approach that aids in identifying words that carry sentiment by utilizing heuristic rules to comprehend the meaning of a word based on its context, ultimately assigning it to the corresponding sentiment category, which involves calculating the polarity of a word based on its co-occurrence frequency with another word of known polarity.

3.4.4. Topic modeling method

In our study, we employed TM techniques to analyze open-ended feedback collected from students at UTH. This approach allowed us to extract meaningful themes related to teacher competencies that resonate with students. We utilized two prominent methods: LDA and BERTopic, each offering distinct advantages in uncovering latent topics within textual data.

LDA is a generative probabilistic model that assumes each document is a mixture of topics, with each topic characterized by a distribution of words. The model works by first defining a fixed number of topics. It then iteratively assigns words to topics based on their co-occurrence in documents, refining these assignments until convergence is achieved (Blei et al., 2003). In our

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

implementation, we preprocessed the student feedback to ensure that the data was clean and relevant, including tasks such as removing stop words, stemming, and lemmatization. By applying LDA, we aimed to identify the underlying themes in student comments, providing insights into the competencies they value in their instructors. This method proved effective in capturing the diversity of opinions expressed by students, allowing us to formulate a preliminary understanding of their expectations, as highlighted by Xing et al. (2020).

Following the initial analysis with LDA, we turned to BERTopic, which builds upon transformer-based embeddings and enhances TM through advanced techniques. One of the key features of BERTopic is its ability to utilize a guided approach, leveraging seed words to direct the topic extraction process (Ravenda et al., 2024; Raman et al., 2024). In our case, we experimented with various iterations of guided seed words, systematically tweaking parameters to identify the optimal model. This iterative process allowed us to refine our topic representations and capture more nuanced insights into the competencies of interest to students.

By combining these methods, we aimed to validate the competency model derived from the literature. The goal was to align our findings from the student feedback with established competencies, ensuring that our model not only reflected theoretical frameworks but also resonated with the practical experiences and expectations of students. The validation of this competency model will be elaborated on in a subsequent section, where we will discuss how the insights gained from TM contributed to our understanding of effective teaching practices in higher education.

In summary, our approach to TM through LDA and BERTopic enabled us to dissect and interpret student feedback meaningfully. By identifying relevant teacher competencies, we hope to contribute to ongoing discussions about effective teaching practices and the factors that enhance student engagement and satisfaction.

3.5. Summary

In this chapter, we explored the research framework and the methodological approaches we followed in our study. Beginning with an introduction to the main objectives, we described the theoretical foundations by conducting an in-depth examination of the research philosophy. We analyzed key philosophical dimensions, such as the nature of reality, epistemology, and axiology, and evaluated the contrasts between positivism, interpretivism, and pragmatism. These perspectives guided our approach to the study's design and implementation.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

The chapter was then organized into three main methodological phases. Phase 1 covered the initial research process, where we defined our objectives, conducted a literature review, and identified gaps in the field, thereby establishing a strong foundation for our study. Phase 2 addressed the methodological framework for text data analysis, highlighting various techniques and comparing approaches for effective knowledge extraction. In Phase 3, we developed and implemented models, detailing the steps taken to generate, validate, and interpret the final study results.

Following the step-by-step methodology, we presented the specific methods and tools used for the literature review and data collection, emphasizing the practical techniques chosen to ensure reliable analysis. The chapter then focused on applications of SA and topic modeling, providing an overview of the algorithms and techniques applied.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

4. Implementation and Results

4.1. Introduction

The present chapter aims to provide a comprehensive overview of the implementation process and results derived from the analysis of student feedback, aligning with the overarching goal of this dissertation. The methods and tools discussed herein are the result of a systematic application of advanced data analysis techniques across various phases of the research. Specifically, this chapter integrates insights from five of our published studies, each contributing distinctively to the analysis of student feedback and the evaluation of teaching quality. While some studies focus directly on identifying and validating teacher competencies through SA and TM, others demonstrate the applicability of ML and advanced tools for extracting meaningful insights from educational data.

To structure the presentation, the chapter is divided into four (4) distinct implementation studies:

Study 1: Predicting student performance using ML algorithms – showcasing the potential of machine learning methods for analyzing educational outcomes.

Study 2: SA of student feedback – employing both lexicon-based and machine learning techniques to assess teacher competencies.

Study 3: Comparative case study: ChatGPT vs. a typical data mining process – exploring innovative and conventional approaches to analyzing textual data.

Study 4: Analysis of students' perspectives through topic modeling – utilizing TM to extract and validate key teaching competencies based on student feedback.

These studies, while varying in focus and methods, collectively illustrate the versatility and robustness of data analysis techniques applied in this research. The findings presented here not only provide empirical support for the competency-based model derived earlier but also highlight the ways in which analytical tools can contribute to a deeper understanding of academic quality and teaching excellence.

4.2. Implementation

This section presents the implementation of methods and techniques used in this research. The five cases discussed include both predictive modeling of student performance and analysis of student feedback to evaluate different methods, algorithms, and approaches, with the ultimate goal of assessing and validating teacher competencies.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

4.2.1. Study 1: Predicting student performance using ML algorithms

In this case we use ML algorithms in order to predict the performance of students, taking into account both past semester grades and socioeconomic factors. We run two models:

- a 2-class one predicting a "pass" or "fail" result and
- an expanded 5-class model, where we predict in which grading group the student will fall in the next semester.

The process we followed to reach our results, as illustrated in the following block diagram (see Figure 14).

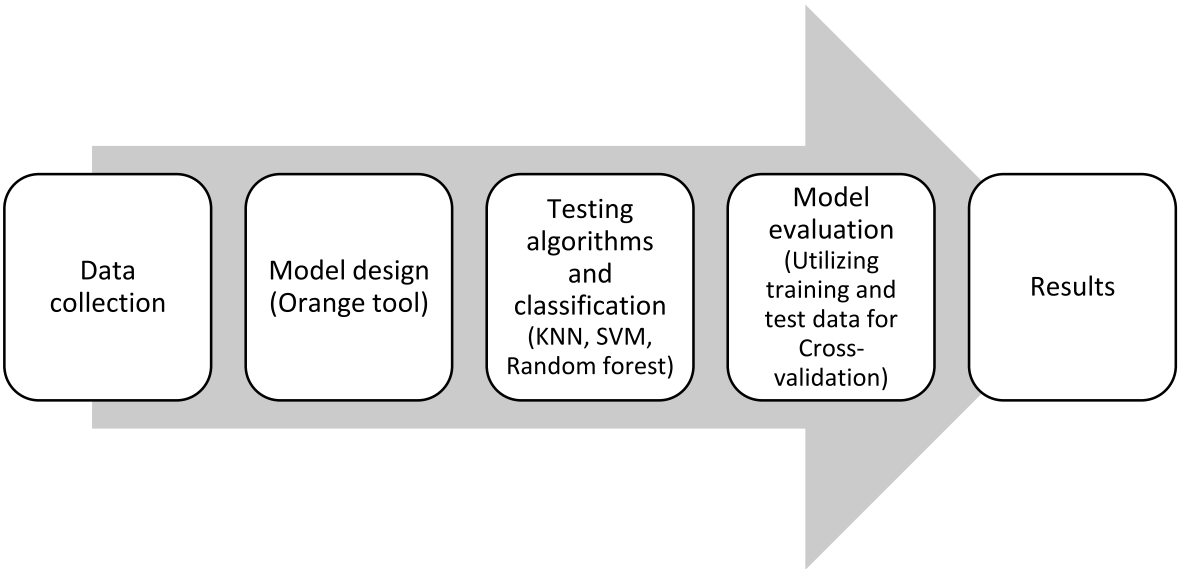


Figure 14. Predicting student performance using machine learning algorithms: The procedure followed (source: Authors' own work).

4.2.1.1. Data collection

We used student performance data to classify and predict student *grading*. More specifically, we used the "Student Performance" dataset, which is publicly available in the UCI repository[1]. This is a popular repository of datasets for testing ML algorithms. These data relate to the performance of secondary school students in two Portuguese schools (Ramaswami et al., 2019). Data characteristics include student grades, demographic, social, and school characteristics and were collected using school reports and questionnaires. The subject we focused on was mathematics

[1] https://archive.ics.uci.edu/ml/datasets/student+performance

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

(mat) and the variable to predict was the third quarter grade. The dataset has 649 samples (students) and 33 variables, presented in Table 6.

Table 6. Model variables and description.

Variable	**Description**
school	School student (binary: 'GP' - Gabriel Pereira ή 'MS' - Mousinho da Silveira)
sex	student gender (binary: 'F' - female ή 'M' - male)
age	student age (numeric: from 15 to 22)
address	type of residence (binary: 'U' - urban or 'R' - rural)
famsize	family size (binary: 'LE3' – less than or equal to 3, or 'GT3' – greater than 3)
Pstatus	parents' marital status (binary: 'T' - living together, or 'A' - apart)
Medu	mother's education (arithmetic: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
Fedu	father's education (arithmetic: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education, or 4 - higher education)
Mjob	mother's work (categorical: 'teacher', 'health' care related, civil 'services' (e.g. administrative ή police), 'at home', 'other')
Fjob	father's work (categorical: ‘teacher', 'health' care related, civil 'services' (e.g. administrative ή police), 'at home', 'other')
reason	reason for school choice (categorical: close to 'home', school 'reputation', 'course' preference, 'other')
guardian	student guardian (categorical: 'mother', 'father', or 'other')
travel time	school-home travel time (arithmetic:1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, 4 - >1 hour)
study time	weekly study time (arithmetic:1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, 4 - >10 hours)
failures	number of previous failures in the course (arithmetic: n if 1<=n<3, else 4)
schools up	additional school support (binary: yes, or no)
famsup	educational support from family (binary: yes, or no)
paid	additional paid courses (Math or Portuguese) (binary: yes, or no)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Variable	Description
activities	extracurricular activities (binary: yes, or no)
nursery	went to kindergarten (binary: yes, or no)
higher	wants higher education (binary: yes, or no)
internet	Internet at home (binary: yes, or no)
romantic	with relation (binary: yes, or no)
famrel	level of family relationships (arithmetic: from 1 - very bad to 5 - excellent)
free time	free time after school (arithmetic: from 1 - very low to 5 - very high)
Go out	going out with friends (arithmetic: from 1 - very low to 5 - very high)
Dalc	alcohol consumption on weekdays (arithmetic: from 1 - very low to 5 - very high)
Walc	alcohol consumption on weekends SCs (arithmetic: from 1 - very low to 5- very high)
health	health status (arithmetic: from 1 - very bad to 5 - very good)
absences	number of absences (arithmetic: from 0 to 93)
G1	first term grade (arithmetic: from 0 to 20)
G2	second term grade (arithmetic: from 0 to 20)
G3	third term grade (arithmetic: from 0 to 20)

4.2.1.2. Model design

We created our model in Orange, as illustrated in Figure 15. The Orange platform (Demšar et al., 2013) is an open-source data visualization, ML and data mining toolkit with a visual programming front-end that allows users to expedite data analysis and easily produce interactive data visualizations.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

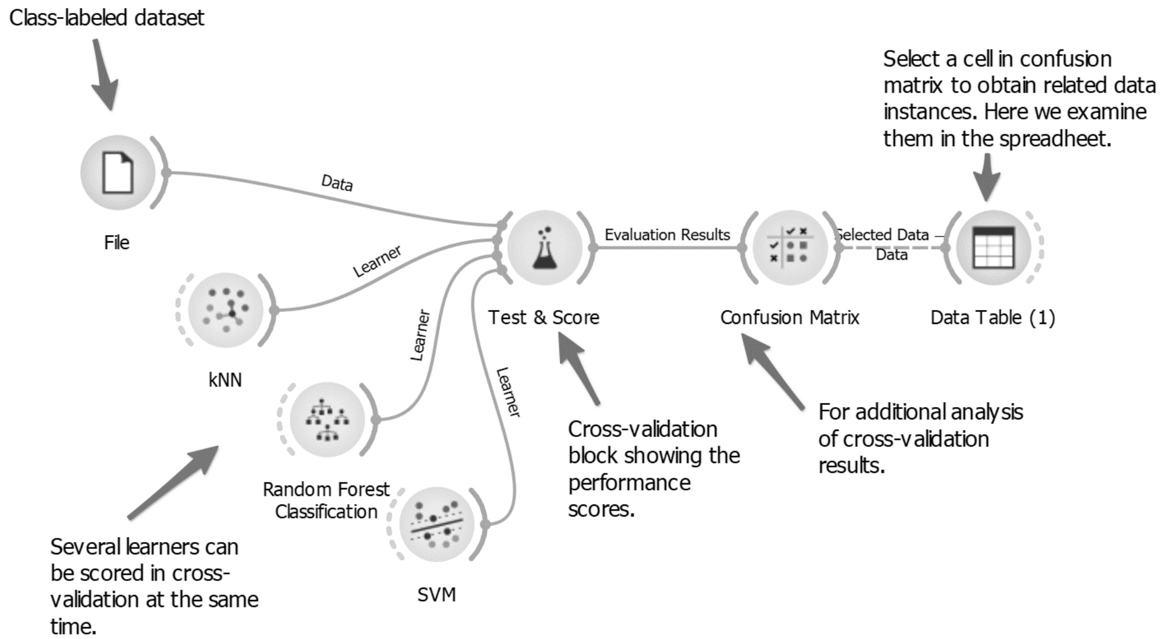


Figure 15. Model designed (source: Authors' own work).

As shown in the Figure 15, we test three classification models, KNN, Random Forest and SVM, by calculating their confusion matrices to evaluate their performance using specific metrics, which are detailed further below.

4.2.1.3. Testing algorithms and classification

We tested and compared different classification algorithms in order to evaluate which one achieves the highest accuracy. In particular, we tested the K-nearest neighbors (KNN), Random Forest (RF), and Support Vector Machines (SVM) algorithms.

KNN is a simple but efficient classification algorithm, hence it is quite widespread (Zhang et al., 2017). For each sample of the test data, it finds the K nearest neighbors from the training data. Finding the nearest neighbors is done with some distance metric, such as the Euclidean distance (Cunningham & Delany, 2021). It then determines which class of the K neighbors has the majority and returns it as output.

In the SVM algorithm the categorization of the data is based on finding an optimal line (for two-dimensional data) or an optimal hyperplane (for higher dimensions) that separates the data creating the maximum margin (Cervantes et al., 2020). The ability to generalize the use of SVMs to non-linear data relies on the kernel trick. In the event that linear separation is not possible, appropriate visualizations are used that transfer the set of data to a larger dimension in order to finally achieve their separation (Nalepa & Kawulok, 2019). SVM is primarily a binary classifier, i.e. it has the ability to categorize into two classes. RF is a classification method that uses a large number

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

of Classification and Regression Trees (CART) in order to provide higher accuracy than a single decision tree (Shaik & Srinivasan, 2019). RF generates a large number of unpruned trees, which are quite different from each other due to their random construction. Thus, the trees are not correlated with each other, so RF can avoid overfitting to the training data and can achieve higher accuracy than a single tree. Moreover, they can handle large datasets efficiently and can be used for both classification and regression.

4.2.1.4. Model evaluation

For evaluation purposes, training and test data are needed. An algorithm builds a model based on training data, but its performance is evaluated against test data that has not been used to build the model. Since there is usually only one data set, the cross-validation technique (Bergmeir & Benítez, 2012) is widely used to evaluate an algorithm. Under this technique, we repeatedly divide the dataset 10 times (see Figure 16) into training data (90%) and test data (10%) (10-fold cross-validation), ensuring that every data point is used in the test subset for evaluation.

Figure 16. K-fold cross-validation with k = 10 (source: Authors' own work).

To measure performance, we calculate the confusion matrix, as shown below in Table 7, for two categories, which we call Positive and Negative:

Table 7. Confusion matrix.

		Category prediction	
		Negative	Positive
Real Category	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

The rows represent the actual class, while the columns represent the predicted class. Correct predictions are located along the diagonal of the matrix starting from the top left cell (1,1). For the evaluation, the following metrics were used, as shown in Table 8:

Table 8. Classification evaluation metrics.

Metric	Calculation	Description
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$	Accuracy indicates the percentage of correct predictions made by the classifier. It provides a general assessment of the model's performance across all classes. However, accuracy alone may not be sufficient if the dataset is imbalanced or if different classes have varying degrees of importance.
Precision	$\frac{TP}{TP + FP}$	A model with low precision is capable of identifying a substantial number of positive examples, but it also exhibits a high rate of false positives, incorrectly classifying many non-positive instances as positive. Conversely, a model distinguished by high accuracy may not succeed in capturing all positive examples, but the instances it categorizes as positive are highly likely to be true positives.
Recall	$\frac{TP}{TP + FN}$	Recall i measures the model’s ability to correctly detect positive instances. A model with high recall performs well in detecting positive instances in the data, although it may also incorrectly classify certain negative instances as positive. Conversely, a model with low recall is unable to identify all (or at least a significant portion of) the positive instances.
F1-Score	$\frac{2 * Precision * Recall}{Precision + Recall}$	It is used to balance the results of Accuracy and Recall into a single measurement of algorithm performance. Therefore, it can effectively capture

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Metric	Calculation	Description
		the model's performance even in cases where the data exhibits significant class imbalance. A middle-range F1 score indicates that either precision or recall is relatively low, although the metric does not specify which one.

4.2.1.5. Results and discussion

In 2-class implementation, we used two classification categories: pass or fail (i.e. predicting if the third semester grade is going to be less than or greater than 10), and tested the three classification models described earlier. The values of the parameters used were; for KNN number of neighbors K=5, for RF number of trees N=50, and for SVM the RBF kernel was selected. These values are widely used in the literature and in the algorithms' documentation, leading to good results. The evaluation results of the models are shown in Table 9.

The metrics AUC (Area Under Curve), CA (Classification Accuracy), F1, precision and recall are indicative of the performance of the algorithms (the higher the values, the higher the performance). Between all three cases we have the best result with the random forest algorithm, followed by KNN and finally SVM. An important observation stemming from the confusion matrices of Figure 17 is that all algorithms' performance is commensurate in predicting a "pass"/"fail" classification.

Table 9. Evaluation results (2-class implementation) (source: Authors' own work).

Algorithm	AUC	CA	F1	Precision	Recall
KNN	0.938	0.881	0.910	0.906	0.916
SVM	0.904	0.833	0.829	0.879	0.913
Random Forest	0.968	0.906	0.930	0.935	0.924

For example, RF predicts 113 failures (out of 130 true failures) and 245 successes (out of 265 true successes), while KNN predicts 105 failures and 243 successes (88.1% accuracy) and SVM predicts 87 failures and 242 successes (83.3% accuracy). Therefore, it is safe to claim that indeed, we are in a position to safely predict, taking into account both past semester grades and socioeconomic factors, whether a student will pass of fail the math course next semester.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

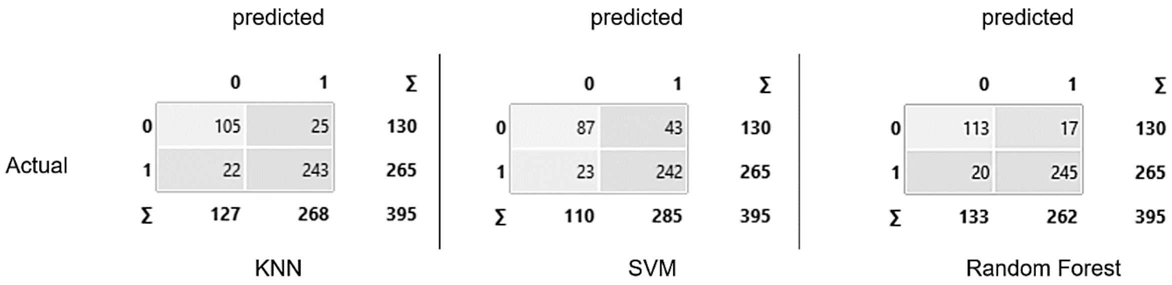


Figure 17. KNN, SVM and Random forest confusion matrix (2-class implementation) (source: Authors' own work).

In order to further investigate the algorithms' capability in providing more fine-grain predictions of the students' performance, we increased the number of classes from 2 ("pass"/ "fail") to 5 (Table 10), and keeping the rest of the parameters unaltered, we rerun the models. The class distribution used is adopted from the original work of Cortez and Silva and shown in Table 5, along with the distribution of the dataset's grades per class.

Table 10. Categories of students' grades (for Math) (source: Authors' own work).

Country	**0 (fail)**	**I (sufficient)**	**II (satisfactory)**	**III (good)**	**IV (excellent/very good)**
Portugal/France	0-9	10-11	12-13	14-15	16-20
Student distribution per grade category	130	103	62	60	40

As we can see in Table 11, AUC, the measure of the ability of the classifier to distinguish between classes, is 90.5% for RF, 87.1% and 82.2% for KNN and SVM respectively, indicating a good performance for our model. However, the CA percentages are lower than those obtained with the 2-class implementation (67.6% vs 90.6% for RF). Nevertheless, since we are dealing with imbalanced data and non-binary classification, AUC is deemed as a more appropriate metric for assessing the model's performance.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 11. Evaluation results (5-class implementation) (source: Authors' own work).

Algorithm	**AUC**	**CA**	**F1**	**Precision**	**Recall**
KNN	0.871	0.600	0.597	0.602	0.600
SVM	0.822	0.491	0.466	0.488	0.491
RF Learner	0.905	0.676	0.666	0.672	0.676

This can be also deducted from the confusion matrices in Figure 18, where correct classification requires larger values along the diagonal. Indeed, in the case of the RF implementation, we can see that the errors are mostly evident in class I and class II, while for the rest of the classes the error is quite negligible, and in all cases, it mostly concerns classifying a sample in the immediately neighboring classes, with approximately 81.3% of data points classified in the correct class. Overall, in this 5-class implementation, the performance of the KNN and especially of SVM deteriorates compared to the 2-class implementation, suggesting that may not be suitable for this purpose (especially SVM) or that their parameters require adjustment for better performance.

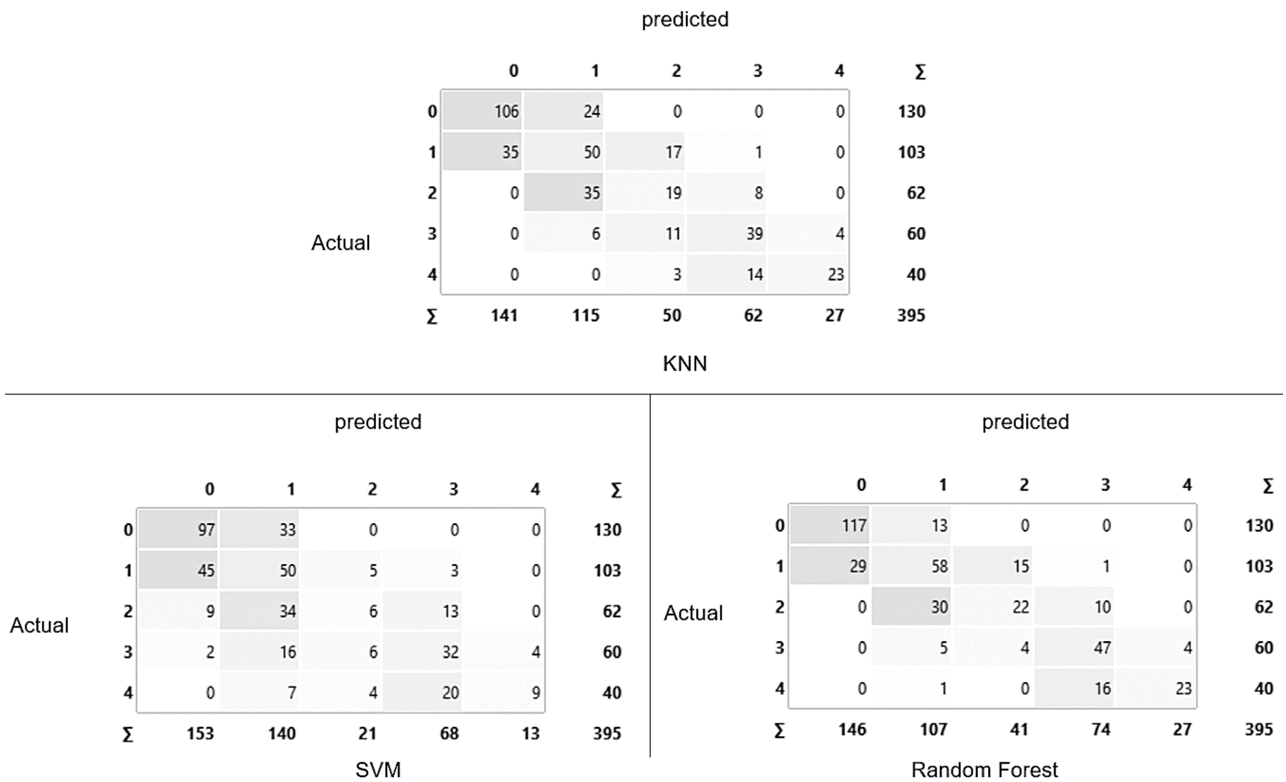


Figure 18. KNN, SVM and Random Forest confusion matrix (5-class implementation) (source: Authors' own work).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

4.2.1.6. Conclusions

This case study, we utilized a dataset of student scores for a single module. Initially, we investigated a two-category classification, i.e., whether the student passed the course or not. The results were very satisfactory, with our model predicting the third semester “pass” or “fail” with a very high level of precision. In order to investigate the algorithms’ performance in giving more fine grain result, we then classified student grades into 5 categories. The results in this case were also satisfactory, with the best algorithm based on the AUC metric being RF (as compared to SVM and kNN). Analysis of the results, using confusion matrices, revealed that although some of the performance indicators were reduced when compared to the two class implementation, the results are commensurately high taking in consideration that with the five class implementation we opt for more fine grain classification results. It is well established that a huge amount of educational data is generated every day and remains untapped. Educational institutions must, by all means exploit this data in order to get insight and support accurate and timely interventions towards improving various aspects of educational services provided. Our approach revealed that techniques and methods using ML algorithms can contribute in harnessing this vast amount of data with multifaceted benefits for the entire educational community.

4.2.2. Study 2: Assessing teacher competencies in higher education: A SA of student feedback

In this case, we used SA methods to measure teachers’ communication competencies. Utilizing the data of over 700 feedback comments from students of our university, we assessed specific competencies through the sentiment intensity that these comments contained. The model designed for the SA, as well as the entire experimental phase, was implemented using the Orange platform, an open source data mining and visualization platform. We highlight how text educational data from student feedback could provide valuable insight to academic institutions. This data will be analyzed and studied as to how it could contribute to the assessment of teacher communication competencies through opinion mining methods and SA. We will also investigate whether students perceive their professors’ competencies differently depending on the nature of the course.

4.2.2.1. SA approach

SA, also known as opinion mining, is a field of study, directly related to NLP that analyzes views, sentiments, evaluations, attitudes and positions of people on issues through the computational processing of subjectivity in text (Liu, 2012). SA aims to evaluate the sentiment of a simple word,

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

phrase, sentence or even an entire document and rate it appropriately. Essentially, using appropriate lexicons or ML approaches (Salazar et al., 2022), opinion polarity is determined, as to whether it is positive, negative or neutral (Hutto and Gilbert, 2014).

The initial challenges encountered in the practical application of SA and opinion mining stemmed from the huge pace and volume of comment content observed on social media on the web. Since then, SA has penetrated and found application in a wide range of fields, focusing on customer opinions about products and services through their written comments. As a result, with the growing familiarity of users with the Internet and online review writing, SA has established itself as a very powerful method of deciphering customer desires in various marketing companies, in the movie industry, in policy- making, in public transportation, in the travel industry, as well as in academic institutions. For the first time in history, we now have a huge volume of opinionated data recorded in digital form for analysis (Liu, 2012). As a result, SA, considered as a big data task, is constantly arousing the interest of researchers.

Traditionally, student feedback analysis in academic institutions, systems, as a rule, were based on close-ended questionnaires that included a set of questions with predefined possible answers (Kumar and Jain, 2015). With SA techniques, students have no limitation to grade only those features that are included in the questionnaire, but instead opinions are freely expressed. According to Nasim et al. (2017) textual feedback is the best way for students to express their views spontaneously, without being limited through predetermined answers Shaik et al., 2022).

Kumar, A., & Jain, R. (2015) proposed an automatic evaluation system based on SA collecting student feedback in the form of text. Misuraca et al. (2021) used a dataset of text, built by Welch and Mihalcea (2016), with 1042 comments extracted from a Facebook group where students expressed their opinions on courses and instructors pertaining to the Computer Science department at the University of Michigan (US). They observed that, concerning the polarity score calculation, it is worth adjusting lexicons to the specific language commonly used by students, particularly because the polarization of some terms may be different from the polarization used in common language. Nitin et al. (2015) observed that students provide feedback that is usually related to labs, projects, skills, etc. when it comes to components of a course which are non-theoretical (interactive, practical, collaborative) in nature. On the contrary, when it comes to theoretical components, student feedback is concerned directly with the profile of their teachers.

Although there is research that has shown that the neutral comments, in SA, are not of particular research interest (Wang and Wu, 2011), even student feedback categorized as neutral can contribute positively in terms of understanding educational issues. Ortigosa et al. (2014), especially in the education domain, point out the necessity, even of the students' neutral feedback, in order to comprehend their attitudes.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

It is not our intention to review the entire body of literature concerning SA from a technical point of view. Nevertheless, we will attempt to provide a brief, and in our opinion necessary, overview in the following section.

Most SA approaches rely heavily on lists of lexical features (e.g., words, phrases), called sentiment lexicons, generally labeled according to their semantic orientation as either positive, negative or neutral (Liu, 2012). According to Feldman (2013), the sentiment lexicon is a key resource for most SA algorithms. Lexicons are classified into two main categories, (1) polarity-based lexicons and (2) valence-based lexicons.

In the first case, a polarity-based lexicon is used to determine the binary polarity of words i.e., either positive or negative. For example, "angry" has negative polarity while "family" positive. To calculate the total polarity of a text, the polarity of each word is searched in the lexicon and if found, is added to obtain an overall score. These lexicons, such as LIWC, Hu and Liu, etc., are widely known and validated. For example, the Hu and Liu lexicon initially created in 2004 (Hu and Liu, 2004), included 2006 words with a positive orientation and 4783 with negative.

However, the SA in practice has become more demanding, applying more complex procedures in order to express, not only whether a linguistic term is positive or negative, but also its polarity's intensity. These valence-based lexicons, such as VADER, ANEW, SentiWordNet etc., use word lists with a valence score. For example, using the VADER lexicon, which measures sentiment-intensity on a scale from -4 to +4. More specifically the term “okay” scores +0.9, “great” scores 3.1, “horrible” -2.5, “ :(“ or “☹” scores -2.2 and so on. Beyond the sentiment word intensity provided, there are sentiment lexicons, such as VADER, which identify properties and specific characteristics of the text, especially in short comments, which affect the perceived sentiment intensity of the text. For example, the exclamation mark, adds more orientation intensity. "Excellent!" is stronger than "Excellent". In addition, it understands many sentiment-laden emoticons such as ":)", "D" and sentiment-laden initialisms and acronyms such as "lol". It is worth noting that many lexicons of the two aforementioned categories, polarity-based lexicons and valence-based lexicons, were developed by building upon or combining earlier lexicons.

In other cases, machine-learning approaches are incorporated to achieve automated sentiment-related tracing processes in the text, considering the knowledge base obtained. Kang H. et al. (2012) proposed an improved Naïve Bayes algorithm to solve problems related to classification performance. Chen and Tseng (2011) used the SVM, a ML algorithm to classify reviews. Al Amrani et al. (2018) proposed a hybrid approach by making a combination of ML techniques, using RF and SVM algorithms, when applying SA to comments from product reviews offered by Amazon. In another powerful study, Kabir M. et al. (2021) used different ML algorithms and concluded that Boosting and Maximum Entropy outperform the other examined ML algorithms for detecting

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

sentiments in online user reviews. However, it should be noted that, most machine-learning approaches have drawbacks that are mainly related to the requirement of training data set. This dataset must be extensive and representative. In addition, significant training and classification (Hutto and Gilbert, 2014) time is required and therefore computing power (CPU processing) and high memory.

In each case, SA can be conducted using different techniques with some alterations. In this case, aiming at presenting a well-defined approach, we describe it step-by-step and next apply it on student feedback comments of our university, with no intention to challenge any other similar or alternative approach. Rather, we aim to highlight the diversity of approaches and to encourage the exchange of views for the purpose of exploring the specific subject.

4.2.2.2. Method

In this subsection, we provide the method that we applied for the SA procedure based on the lexicon approach. The initial stage (stage 1) concerns the collection of student comments and their import as the dataset in the input channel. The data can be imported in various file formats such as Excel (.xlsx), comma-separated (.csv) and native tab-delimited (.tab) files, etc., depending on the system used. Then, in stage 2, the text preprocessing process takes place, during which some natural language issues such as Transformation, Tokenization, Stopwords filtering, Normalization (stemming and lemmatization) etc. should be considered. The process of text preprocessing consists of a sequence of steps and its objective is to prepare the raw text data for the next phase, thus acquiring a clean and consistent form that can then be fed into a model for further analysis. Figure 19 illustrates the order of distinct steps followed in our approach during text preprocessing. Additionally, Table 12 summarizes thoroughly, all of the text data challenges handled in these steps.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

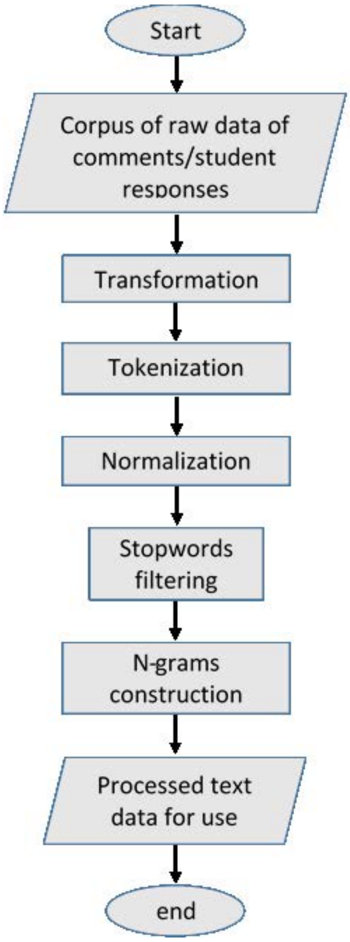


Figure 19. Text preprocessing phase (Stage 2) (source: Authors' own work).

In stage 3, the sentiment intensity for each segment of the data processed in the previous stage is detected. To do this, a prepared sentiment lexicon was used containing words and corresponding sentiment scores for each word. This assignment of polarity is quite important, as the purpose is to calculate the overall score of a student comment on a particular topic, although in the individual words, aspects, etc. different grades may be assigned. For instance, a comment consisting of seven tokens may have a positive average grade, even though five of the tokens have negative scores and only the other two have a positive one. In the final stage (stage 4), all the SA insights are converted into relevant reports. Reports enable interactivity through different forms such as charts, graphs, etc. At this stage of the SA steps, visualization is a very important segment for the overall monitoring of the results and the receiving of actionable insights. To enable the decision of the right course of action, it is crucial to know which aspects got high a score, which ones got low scores and generally, to have an idea as to which areas need our attention more than others.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 12. Text data challenges (source: Authors' own work).

Process	Description
Transformation	Removes URLs from text. Detects and retains only the text between html tags. For example, from valuable text retains the "*valuable text*" If necessary, applies lowercase transformation.
Tokenization	Tokenization is the method of splitting text into smaller components, called tokens or units. These units may be words, bigrams, sentences and even paragraphs. In addition to breaking up a sequence of strings into pieces, a characteristic task in the process of tokenization is the discarding of certain characters such as punctuation.
Normalization (stemming and lemmatization)	This method is used to find out the stem or lemma of a word. Words are stemmed (based on root form) using different stemming algorithms such as, Porter stemmer (Balakrishnan and Lloyd-Yemoh, 2014), Snowball stemmer (Khyani *et al.,* 2021) etc. while lemmatization reduces a word to its base or dictionary form (called a "lemma") by considering its meaning. For instance, the word "flying" is stemmed as "fly" by removing the "ing" suffix while the word "better" using lemmatization is lemmatized to "good". It is very important to remove the endings of the words, to return the root word so that it is given the intensity of polarity based on the dictionary used.
Stopwords filtering	At this stage, stopwords from texts are removed, i.e. pronouns, articles, that carry very little useful information such as "he", "it", "the", "and", etc., without changing the semantics of a text. These words occur commonly across all the texts in the corpus and need to be dissociated in order to retain only the meaningful words and improve the performance of our model. Furthermore, at this stage we can remove words that make no contribution when analyzed, such as verbs or prepositions with high frequency of occurrence and of negligible interest (be, with, etc.).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Process	Description
N-grams construction	The N-gram consists of successive tokens, with determined window (N tokens), that occur in texts. N-grams of texts are extensively useful in text mining and NLP tasks (Xun *et al.,* 2015). For instance, "Difficult lesson" is a 2-gram and "Difficult lesson in lab" is a 4-gram with different meaning. It is quite interesting to determine one or the other case and to find out the frequency of their occurrence and polarity in a plethora of student feedback.

4.2.2.3. Implementation

At the end of each semester at our university, students are asked to answer a questionnaire about each course and its instructor, anonymously online. For this purpose, the Quality Assurance Unit (QAU) of the University of Thessaly (UTH), which is the central body for the coordination and support of the QA procedures of our university, sends a relevant link to the students' email. The questionnaire is harmonized with the requirements of the Hellenic Authority for Higher Education (HAHE) and contains closed and open-ended questions. However, as a continuation of our previous research (Dervenis et al., 2022) and independently of the aforementioned process, we aimed to measure the teacher competencies referring to the dimension of *Communication* (Table II). For this reason, we additionally developed a specific online questionnaire, oriented to our purpose, with which we asked students to provide responses to open-ended questions about Listening, Persuasion, Empathy, Presentation and Collaboration provided by their teacher. For the needs of our research, we initially focused on a course taught by two different professors in order to determine any differences in the nature of the teachers and not of the course. Subsequently, we repeated the same procedure with the same teacher in two different courses, to examine this time whether possible differences were due to the nature of the course and not of the teacher.

Table 13. Teacher competencies that refer to the Communication dimension with the corresponding open-ended questions asked to students (source: Authors' own work).

Teacher competence to be measured	Open-ended question asked to students.	Question coding
C22: Listening	Note general comments about your professor regarding their competence in actively listening (i.e., making a conscious	Q1

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

	effort to hear not only the words said but, more importantly, the complete message being communicated).	
C23: Persuasion	Note general comments about your professor regarding their powers of persuasion used to inspire students and enhance / boost their interest in the subject.	Q2
C24: Empathy	Note general comments about your professor: regarding the empathy they feel for their students. (Empathy as a basic dimension in emotional intelligence, is the ability to put oneself in another person's shoes and understand what they are feeling).	Q3
C25: Presentation	Note general comments about your professor regarding their presentation abilities (i.e., do they deliver their presentation in a clear and concise manner?).	Q4
C26: Collaboration	Note general comments about your professor regarding their collaboration with students (i.e., are they cooperative, do they detect an unwillingness to cooperate and prevent it, do they take into consideration the smooth functioning of the course?).	Q5

It is worth noting that, from a technical point of view, the platform allows us to choose the preferred language for data entry and processing. For our purposes, we judged that we should rely on the VADER lexicon which is fully open-sourced under the MIT License, in the English language, fully validated and widely accepted across multiple distinct domains including the field of education (Shafana and Safnas, 2022; Wook et al., 2020). Furthermore, the VADER sentiment lexicon has been compared to many other well-established SA lexicons such as Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), Affective Norms for English Words (ANEW), SentiWordNet (SWN), SenticNet (SCN), WordNet and the Hu-Liu, and the results were indeed quite remarkable. The comparison results revealed that the VADER lexicon performs at a similar level to that of individual human raters. The classification accuracy metric "Precision" varied between 0.69 and 0.99 across various domain contexts (Hutto and Gilbert, 2014).

At the same time, it was essential that the students express themselves in their mother tongue, Greek, in order for their responses to convey meaning and sentiment as accurately as possible

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

(Sharma et al., 2014). Therefore, to achieve maximum accuracy and efficiency, all the comments were initially collected in the Greek language and subsequently translated into English by accredited translators. In this way, we ensured the highest degree of accuracy both from the students' feedback in their mother tongue, as well as by utilizing the VADER English dictionary which provided us with the validity needed for the purposes of our research. We then entered the final data into our system.

This phase was implemented again using the Orange platform (Demšar et al., 2013). Our model takes the sequences of a sentence as input and gives the intensity of the sentiment polarity of the sentence as output. In Figure 20, the workflow depicts the model designed according to our approach.

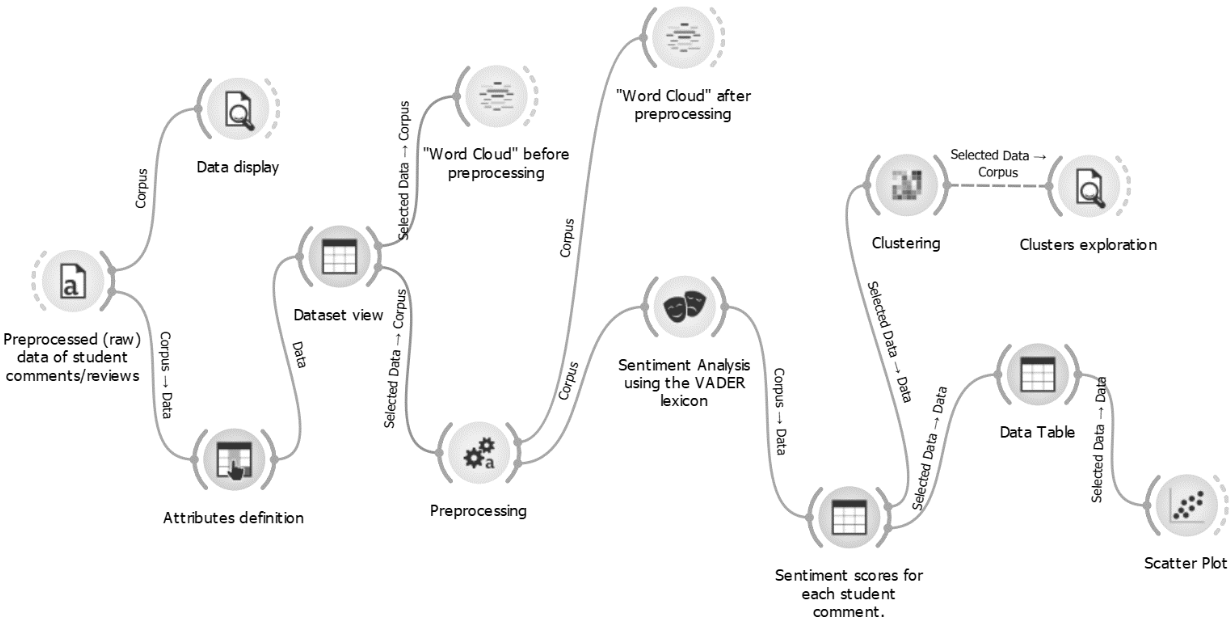


Figure 20. Workflow of model designed for SA (source: Authors' own work).

After importing the raw data from the student feedback into the system, the definition of the attributes is done. That is, we assign, which data will be the target for analysis and which will not. In our case, the questions and their answers' content is considered for analysis. Then, we processed the "raw" data in order to clean it up by removing noise, such as stopwords ('a,' 'and,' 'is,' 'the,' etc.), which do not contribute meaningful insights to the analysis.

Then, we utilized the VADER lexicon, which covers multiple domains and special cases for SA that include the proper handling of sentences with typical negations (e.g., "not good"), the use of contractions as negations (e.g., "wasn't very good"), conventional use of punctuation to signal

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

increased sentiment intensity (e.g., "Good!!!"), conventional use of capital letters to signal emphasis (e.g. "NICE") and other special cases such as those mentioned in the previous section (section 2). Subsequently, we computed and reported the sentiment score as positive, negative or neutral to reach the compound score of each student comment received. The compound score of a specific sentence is determined by summing up the sentiment scores of each individual word within the sentence (Amin et al., 2019). This calculation is achieved by applying the following formula:

$$compound\ score = \frac{x}{\sqrt{x^2 + a}} \qquad (4.1)$$

Where,

compound score: the sentiment score received of a specific sentence

x: the sum of individual sentiment scores for each word in the specific sentence (as described in detail previously)

a: the normalization factor that typically has the value of 15

The sentiment score resulting, indicates the overall intensity of sentiment in the sentence and falls within the range of -1 to +1. Figure 21 shows a representative sample of comments with the corresponding compound (sentiment) scores for each.

Question	Students Comment	pos	neg	neu	compound
Q4	He is to the point and ...	0.176	0.059	0.765	0.5964
Q1	I believe he presents th...	0.175	0.066	0.758	0.3182
Q1	I think that students pa...	0.108	0.072	0.821	0.1779
Q1	Your active listening ca...	0.387	0.075	0.538	0.6997
Q2	His persuasion skills ar...	0.148	0.076	0.776	0.2115
Q2	he has told us very poli...	0	0.076	0.924	-0.1027
Q2	The professor uses ever...	0.176	0.088	0.736	0.3678
Q2	His calm way of teachi...	0.333	0.09	0.577	0.5859
Q2	He tried to help us wit...	0.272	0.091	0.638	0.6597

Figure 21. Sample comments from students with their corresponding scores (source: Authors' own work).

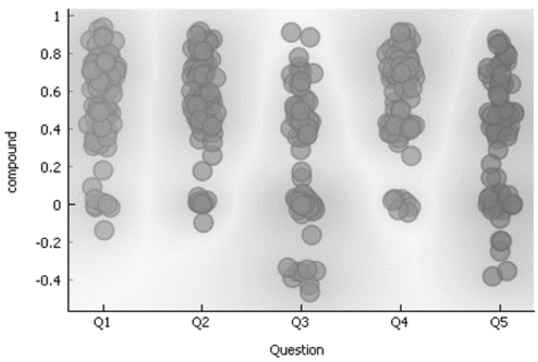
4.2.2.4. Results and discussion

Our research was developed in 2 dimensions (cases). In the first, we aimed to investigate and compare the competencies that students perceive in two different professors. In order to ensure a common base, we focused on one course, taught by both professors. Our approach detected a

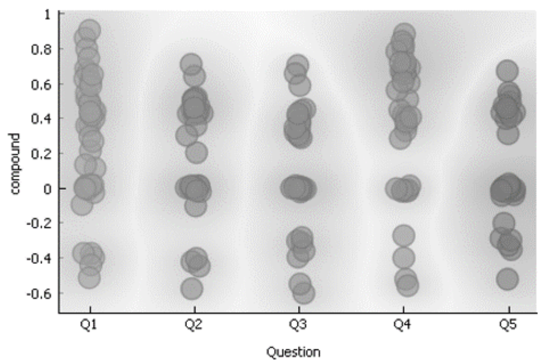
Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

different degree of competencies between teachers, which seems perfectly normal. Figure 22 illustrates the results extracted, referring to the grade of sentiment received from student comments regarding prof_1 in course_1 and prof_2 in course_1. The data is displayed as a collection of points, with the x-axis depicting the question number and the y-axis depicting comment intensity (ranging from -1 to +1) as defined by our system.

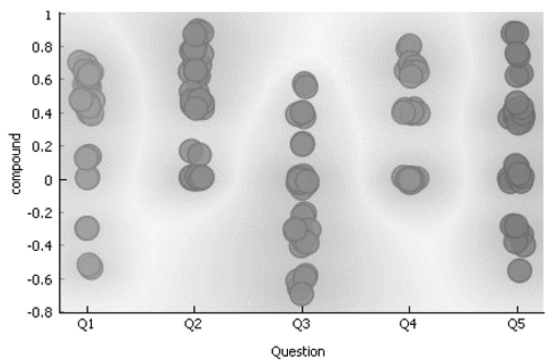
Figure 22 shows that Q1, Q2 and Q4, for prof_1, received the most positive responses with the greatest intensity, thus indicating that the corresponding competencies of “Listening”, “Persuasion” and “Presentation” were evident in prof_1. On the contrary, the feedback comments referring to “Empathy” and “Collaboration”, had the lowest sentiment score. In addition, for each question, it should be noted that only a few student comments received a negative compound intensity. Also, in Figure 22, it is observed that the scores received by the students' comments concerning the competencies of prof_2 in the same course (course_1) are distributed over a wider range of values of intensity, without clearly indicating the existence of the specific competencies concerning prof_2.



Prof_1 in course_1



Prof_2 in course_1



Prof_2 in course_2

Figure 22. Sentiment intensity of student comments for prof_1 in course_1, prof_2 in course_1 and prof_2 in course_2 (source: Authors’ own work).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Summarizing and comparing the results of the first case, depicted in Figure 22, concerning the competencies of 2 different professors in the same course (course_1), the students' point of view is clearly established, as depicted in Figure 23. Despite the obvious differences between the two professors that the students discerned, as for example in the "Presentation" skill, it is found that in both professors the students "see" the lack of teachers’ empathy as evident.

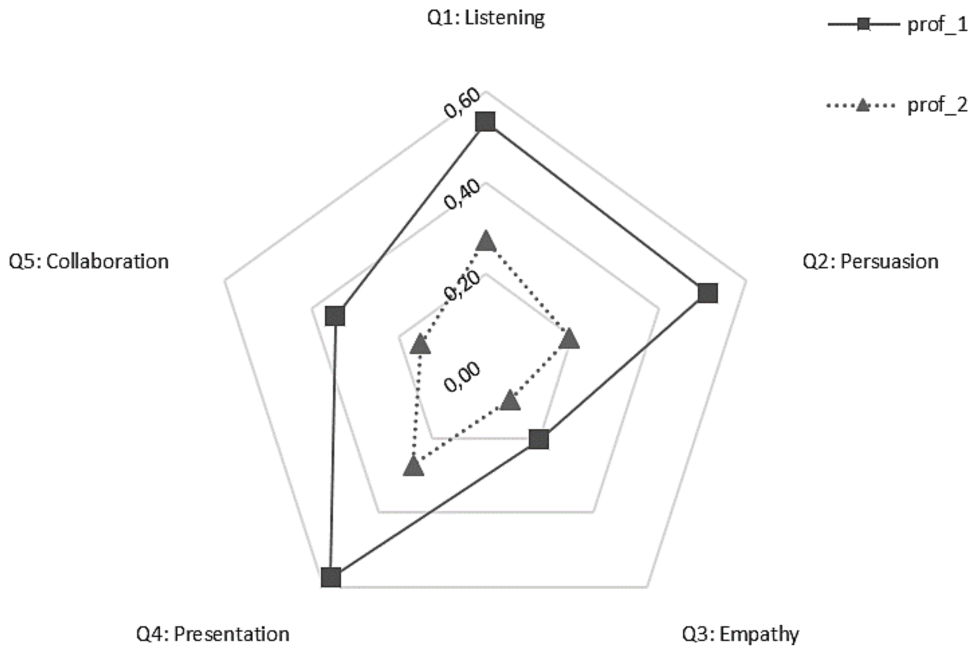


Figure 23. Mean sentiment score of student comments received, regarding the competencies of 2 different instructors of the same course (course_1) (source: Authors’ own work).

Case 2, which concerns the same professor in 2 different courses, is also quite interesting. The aim of this case was to investigate whether students perceive one professor’s competencies differently depending on the nature of the course. For this reason, we attempted to measure the specific competencies of prof_2 in course_2, in addition to course_1. Figure 22 depicts the results referring to prof_2 in course_2. As observed, there are questions (Q2 and Q4) that correspond to the “Persuasion” and “Presentation” competencies which received comments with negligible negative sentiment scores, in contrast to the case of prof_2 in course_1. However, a moderate amount of feedback comments referring to “Listening”, “Empathy” and “Collaboration” received negative compound scores, as in case 1.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Taking into account the results regarding course_1 and course_2 for prof_2, the competence, or lack thereof, when it comes to empathy expressed, listening and collaboration, is equally evident in students regardless of the nature of the course. Simultaneously, teacher competencies such as powers of persuasion used to inspire students and enhance / boost their interest, as well as presentation competencies, seem to depend on the nature of the course. In order to reinforce this observation, we posed an additional question to the students of courses 1 and 2. We asked them to express their opinion about which of the 2 courses they considered to be more difficult. The results showed an overwhelming difference between the 2 courses, since more than 90% of the students considered course_1 to be more difficult than course_2. We concluded that there are indeed some competencies such as persuasion, presentation, etc. which depend on the nature of the course. For example, according to students, a course that is easy to understand contributes to the emergence of the presentation skill possessed by the professor who teaches it.

Next, we visualized our data in a cluster matrix using the k-means algorithm. The following heat map diagram, in Figure 24, concerns the comment clustering of the case of prof_1, course_1. The far-right column indicates the prevailing question in the specific comment cluster. An interesting finding emerging through the cluster exploration was the nature of the content of each cluster. We found that comments referring to the same question, and to a specific teacher competence, were clustered together, based solely on sentiment and not on question type. For example, it seems clear that the comments related to Q1 were "grouped" together because of the common characteristics detected in the specific comments and not because they were stated to be related to Q1. In other words, the student opinions on specific teacher competencies showed homogeneity in their sentiment and beliefs.

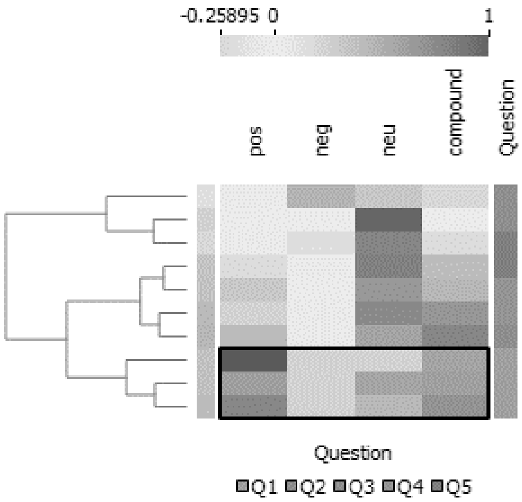


Figure 24. Data clustering using K-means algorithm (source: Authors' own work).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

4.2.2.5. Limitations

In this particular case, during the data collection process, our research was limited to gathering feedback data from students of UTH on two courses and for two professors. Although an expanded collection of student feedback including multiple courses with different characteristics such as semester, cognitive area, degree of course's difficulty, etc., would require additional time and effort, we believe it could contribute to revealing important findings and conclusions.

In addition, the language in which the students expressed themselves was their mother tongue, Greek, in order for their responses to convey meaning and sentiment as accurately as possible. Although from a technical point of view in our model, we had the ability to rely on a multilingual sentiment lexicon, we preferred the validity of the VADER lexicon. Because this choice limited us to the use of the English language, in order to achieve maximum accuracy and efficiency, all the student comments initially collected in the Greek language were subsequently translated into English by accredited translators.

4.2.2.6. Conclusions

In this case we highlighted the necessity of extracting valuable knowledge from student feedback and exploiting this data with the aim of improving teaching and learning in universities. We also emphasized that student feedback comments regarding open-ended questions are unique but unfortunately remain untapped due to the fact that current analyses are primarily limited to traditional techniques focused on closed-ended questions. As a result, not only do teachers miss out on the opportunity to gain insight into teaching procedures and the unique elements that characterize them in the educational process, but the educational system is also deprived of the potential benefits for all involved.

In this study, we proposed a SA approach for mining and quantifying student opinion in order to bridge the aforementioned gap. Following our previous research (Dervenis et al., 2024) and using our approach with a dataset from the UTH, teacher communication competencies were assessed measuring the intensity of sentiment of each student comment in two cases; for two teachers in the same course and for one teacher in two different courses. The analysis of the results, indicated that students predominantly expressed similar views regarding competencies of the same instructor, with minimal variations noted. From our research, it was also revealed that teacher competencies such as "Presentation" and "Persuasion" are more evident in courses with a lower degree of difficulty and vice versa in the opposite case. More specifically, according to our research and based on student opinion, a course that is easier for students to understand may highlight that the presentation and the persuasion competencies of the specific professor. Simultaneously, according to the data collected from the students in our study, competencies such as "Listening",

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

"Empathy" and "Collaboration" possessed by a teacher appeared to be independent of the nature of the course.

4.2.3. Study 3: Chat GPT vs. Typical data mining process: A comparative case study on teaching competencies via SA

In this case an attempt is made to analyze the data from student feedback in two different ways:

a) Using traditional tools for SA and

b) using a generative AI tool such as ChatGPT, which implements a generative pre-trained transformer that can be used for such application.

The benefit will be significant, since, if AI chatbots can reliably perform such tasks, it will be quite easy for anybody to perform such analyses and to benefit from insights gained.

4.2.3.1. ChatGPT and traditional lexicon-based approach

As previously mentioned in subsection 4.2.2 of this dissertation, most SA methods rely heavily on sentiment lexicons, which are lists of lexical features such as words and phrases, categorized according to their semantic orientation as positive, negative, or neutral (Liu, 2012). These lexicons serve as a crucial resource for SA algorithms, as highlighted by Feldman (2013). Sentiment lexicons can be broadly classified into two main categories: polarity-based lexicons and valence-based lexicons. It is important to acknowledge that many lexicons in both the Polarity-based Lexicons and Valence-based Lexicons categories have been developed by building upon or combining earlier lexicons.

In certain cases, machine-learning approaches are incorporated to automate the process of SA in text, leveraging the knowledge base acquired. For instance, Kang H. et al. (2012) proposed an enhanced Naïve Bayes algorithm to address classification performance issues. Chen and Tseng (2011) employed the SVM, for review classification. Al Amrani et al. (2018) presented a hybrid approach that combined ML techniques, utilizing the RF and SVM algorithms, to perform SA on product reviews from Amazon. In another significant study, Kabir M. et al. (2021) employed various ML algorithms and determined that Boosting and Maximum Entropy outperformed other examined methods for sentiment detection in online user reviews. However, it is worth noting that most machine-learning approaches have certain limitations, primarily related to the need for extensive and representative training datasets. Furthermore, substantial training and classification time, as well as considerable computational resources (CPU processing power and high memory), are required.

Recently, Large language models (LLMs) have received research and commercial attention for their amazing abilities across a number of domains, problems and generally on various NLP tasks.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

In this context, all LLMs have been evaluated in performing SA to varying extents. In their survey, Liu et al., presented the trends of LLMs and an extensive number of uses, including SA in a number of different domains (Liu et al. 2023). For example, fine-tuning pre-trained BERT models were used for aspect-based SA. Aspect-based SA aims to identify the sentiment expressed towards specific aspects mentioned in a text. The process involves fine-tuning pre-trained BERT models, which have shown exceptional performance and achieved impressive results in this task. BERT-based studies have only utilized either, the last output layer, disregarding the valuable semantic knowledge embedded in the intermediate layers or BERT's intermediate layers for fine-tuning the process and enhancing the performance of SA (Song et al. 2020).

SA using ChatGPT can be done by framing the analysis as a language generation task. Although ChatGPT is not specifically designed for SA, it can be used to generate responses that reflect the sentiment intensity based on the input provided. It's important to note that the accuracy of SA using ChatGPT depends on several factors, including the quality and diversity of the training data, prompt formulation, and the model's understanding of context. Comparing SA using ChatGPT with traditional lexicon-based methods poses several challenges, as displayed in Table 14.

Table 14. Challenges faced; ChatGPT and traditional lexicon-based method comparison (source: Authors' own work).

Generative languages	**Traditional lexicon-based methods**
The model is based on a generative language model that relies on a large amount of training data to generate responses	They use predefined lists of words with pre-assigned sentiment values.
Generative languages have a contextual understanding of language and can capture complex relationships between words and sentences. This contextual understanding allows them to capture nuances and subtle sentiment cues that may be missed by lexicon-based methods.	Lexicon-based methods focus on specific lexical features, such as individual words or phrases, to determine sentiment polarity.
Generative languages generate textual responses, which makes it more difficult to evaluate SA in a straightforward manner.	Lexicon-based methods provide explicit sentiment labels
The performance of generative languages varies based on the specific training data and fine-tuning process used. Conducting SA across different domains or datasets may	This is not required from lexicon-based methods in most of their applications

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

require additional adaptations and fine-tuning to ensure fair
and accurate comparisons.

4.2.3.2. Method

At the end of each semester at our university, students are asked to answer a questionnaire about each course and its instructor. The questionnaire is harmonized with the requirements of the HAHE and contains both close and open-ended questions. However, as a continuation of our previous research (Dervenis et al., 2022), we aimed to measure the teacher competencies that refer to the dimension of *Communication* (Table 15). For this reason, we additionally developed a specific online questionnaire, oriented to our purpose, with which we asked students to provide responses to open-ended questions about Listening, Persuasion, Empathy, Presentation and Collaboration provided by their teacher.

Table 15. Teacher competencies related to the "Communication" dimension, along with the corresponding open-ended questions posed to students (source: Authors' own work).

Teacher competence to be measured	Open-ended question asked to students.
C22: Listening	Note general comments about your professor regarding their competence in actively listening (i.e., making a conscious effort to hear not only the words said but more importantly, the complete message being communicated).
C23: Persuasion	Note general comments about your professor regarding their powers of persuasion used to inspire students and enhance / boost their interest in the subject.
C24: Empathy	Note general comments about your professor regarding the empathy felt for their students. (Empathy as a basic dimension in emotional intelligence is the ability to put oneself in another person's shoes and understand what they are feeling).
C25: Presentation	Note general comments about your professor regarding their presentation abilities (i.e., do they deliver their presentation in a clear and concise manner?).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Teacher competence to be measured	Open-ended question asked to students.
C26: Collaboration	Note general comments about your professor regarding their collaboration with students (i.e., are they cooperative, do they detect an unwillingness to cooperate and prevent it, do they take into consideration the smooth functioning of the course?).

The phase was implemented using the Orange platform (Demšar et al. 2013). In Figure 25, the workflow depicts the model designed according to our approach. Specifically, the initial stage (stage 1) concerns the collection of student comments and their import as the dataset in the input channel.

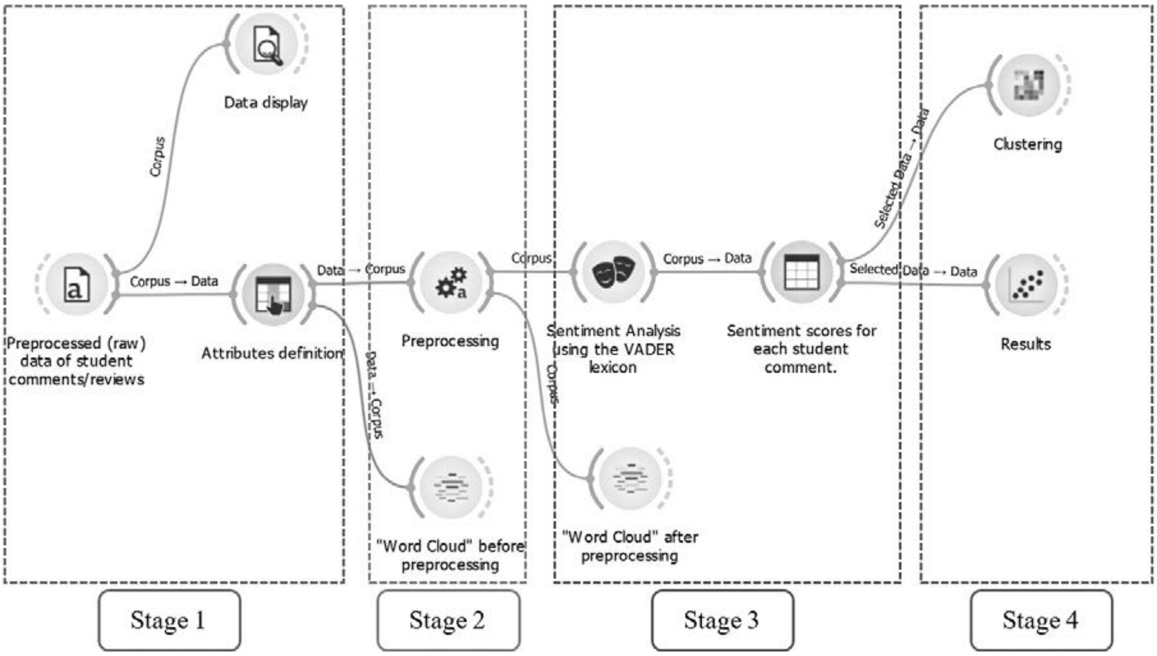


Figure 25. Workflow of model designed for SA (source: Authors' own work).

Then, in stage 2, text preprocessing takes place, during which some natural language issues such as Transformation, Tokenization, Stopwords filtering, Normalization (stemming and lemmatization) etc. should be considered. Text preprocessing consists of a sequence of steps and its objective is to prepare the raw text data for the next phase, acquiring a clean and consistent form that can then be fed into a model for further analysis. In general, text preprocessing is an integral part of NLP applications. Figure 26 illustrates the order of distinct steps followed in our approach during text preprocessing.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

In stage 3 of our model (see Figure 25), the sentiment intensity for each segment of the data processed in the previous stage is detected. To do this, a prepared sentiment lexicon containing words and corresponding sentiment scores for each word was used. This assignment of polarity is quite important, as the purpose is to calculate the overall score of a student comment on a particular topic, although in the individual words, aspects, etc. different grades may be assigned. For instance, a comment consisting of seven tokens may have a positive average grade, even though five of the tokens have negative scores and only the other two have a positive one. It is worth noting that, from a technical point of view, the platform allows us to choose the preferred language for data entry and processing. We judged that, for our purposes, we should rely on the VADER dictionary, which is in the English language, widely accepted and fully validated (Hutto and Gilbert, 2014). At the same time, it was essential that the students express themselves in their mother tongue, Greek, in order for their responses to convey meaning and sentiment as accurately as possible (Sharma et al. 2014).

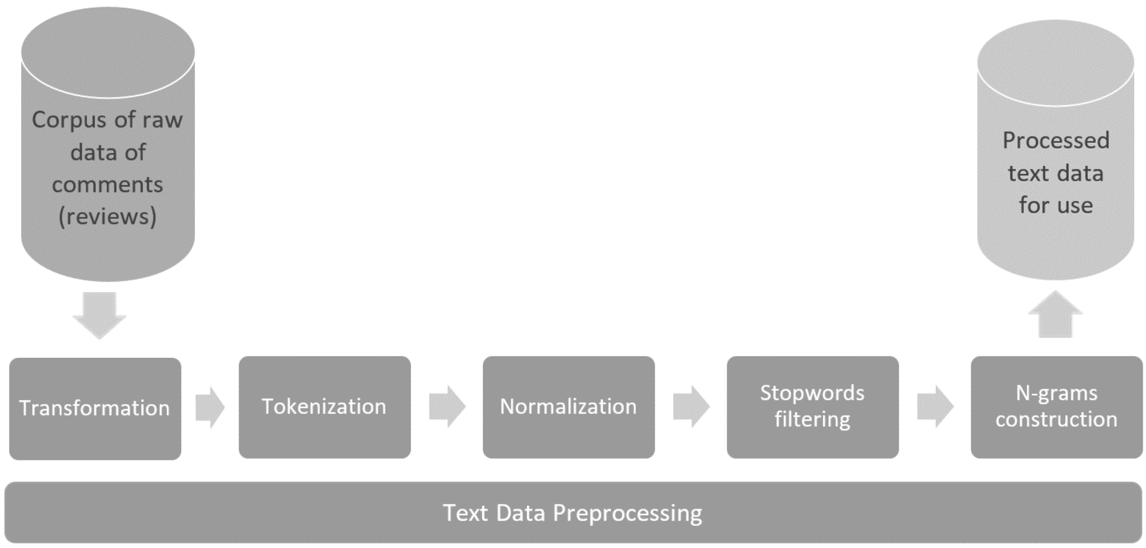


Figure 26. Text preprocessing phase (stage 2) (source: Authors' own work).

For achieving maximum accuracy and efficiency, all comments, initially expressed in the Greek language, were subsequently translated into English by accredited translators. In this way, we ensured the highest degree of accuracy both from the students' feedback in their mother tongue, as well as by utilizing the VADER English dictionary which provided us with the validity needed for the purposes of our research. Subsequently, we computed and reported the sentiment score as positive, negative or neutral to finally reach the compound score of each student comment

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

received. Figure 27 shows a representative sample of comments with the corresponding sentiment scores for each.

title	Question	Students Comment True	pos	neg	neu	compound
239	Q3	The professor has the skill to ...	0	0.144	0.856	-0.4019
362	Q5	He is willing to cooperate wi...	0.092	0.138	0.769	-0.2263
345	Q5	He understands the students...	0.171	0.134	0.695	0.1779
141	Q2	He is generally very convinci...	0.421	0.134	0.446	0.6997
77	Q1	The professor's ability to und...	0.344	0.129	0.527	0.5378
136	Q2	He uses good arguments wh...	0.259	0.127	0.613	0.4215
253	Q4	He tries hard to make everyt...	0.257	0.124	0.619	0.3612
379	Q5	He tries and wants to cooper...	0.109	0.122	0.769	-0.0772
372	Q5	He is willing to cooperate an...	0.121	0.121	0.759	0

Figure 27. Sample comments from students with their corresponding scores (source: Authors' own work).

In the final stage (stage 4), all the SA insights are converted into relevant reports. Reports enable interactivity through different forms such as charts, graphs, etc. At this stage of the SA steps, visualization is a very important segment for the overall monitoring of the results and the receiving of actionable insights. To enable the decision of the right course of action, it is crucial to know which aspects got high a score, which ones got low scores and generally, to have an idea as to which areas need our attention more than others.

After the aforementioned lexicon-based method, we applied a different novelty approach, in order to compare and identify any discrepancies and better understand the strengths and limitations of each with the overall aim of improving analysis. For this purpose, we used a processing tool based on AI technology, the ChatGPT, an AI chatbot created by Open AI (Roumeliotis & Tselikas, 2023). First, we initiated contact with ChatGPT by asking if it could perform SA based on student feedback. Then, the chatbot responded by utilizing its capabilities in NLP suggesting specific instructions to follow. We followed its instructions and provided new input so that it could comprehend and clarify our original query.

The guidelines followed, were comprised of general steps regarding student feedback data collection, data format, dataset size, etc. We were then asked to provide all the data we had collected. Responding to the Chatbot's request, we entered the dataset in sections, so that each section of data concerning student feedback to a specific question, i.e. one dataset of student responses for each of the five questions we asked the students. In this way, we wanted to make the answers relating to a specific question distinct, in order to facilitate our interaction with the AI chatbot and to be able to refer to each set of answers separately. We also aimed to "help" ChatGPT

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

to generate a more accurate, relevant and targeted response, as seen in the results received and presented in the next section.

The figure below (see Figure 28) shows the respective stages followed in our interaction with the ChatGPT in order to reach our purpose. It is worth noting that in general, the stages of action-reaction process between human and AI chatbot may differ depending on the query and context.

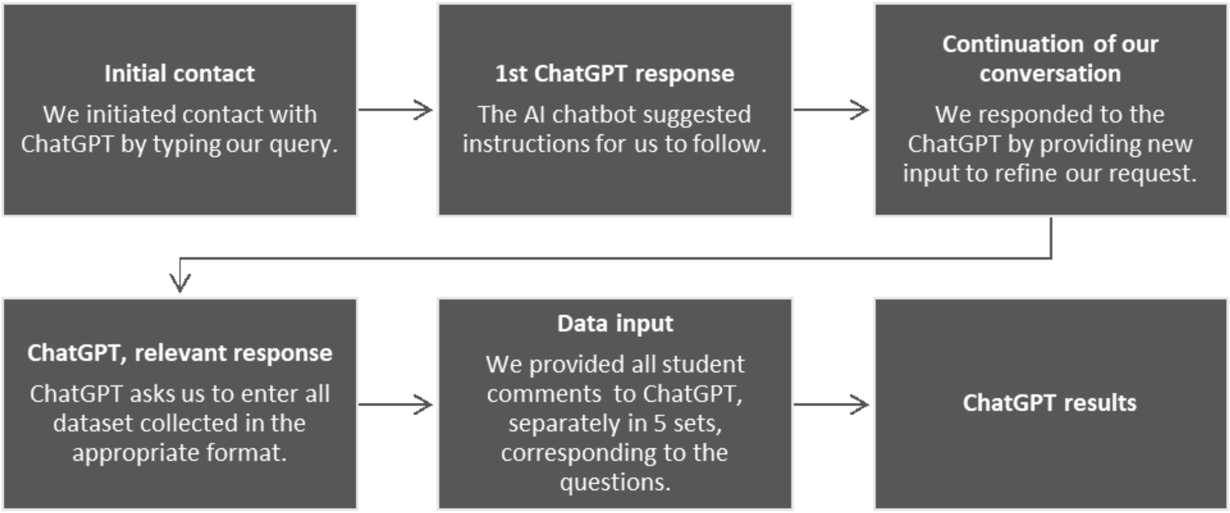


Figure 28. The steps in our engagement with ChatGPT (source: Authors' own work).

4.2.3.3. Results and discussion

In the initial approach of our analysis (typical data mining approach), we focused on identifying and measuring SA of student feedback based on lexicon and utilizing Orange, a data mining and visualization platform. The results, referring to the grade of sentiment received from student comments regarding the teacher of a particular course are depicted in Figure 29. The information is presented in the form of a group of dots, where the horizontal axis represents the number of the question (displayed in different colors), and the vertical axis represents the degree of comment intensity on a scale from -1 to +1, based on our system's assessment. It is worth noting that the overall sentiment intensity of a particular comment is determined by combining its positive, negative, and neutral scores according to our model.

Based on Figure 29, it can be observed that Q1, Q2, and Q4 received the most positive feedback with the highest level of intensity. This suggests that teacher demonstrated the competencies of

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

"Listening," "Persuasion," and "Presentation". Conversely, comments related to "Empathy" and "Collaboration" received the lowest sentiment score.

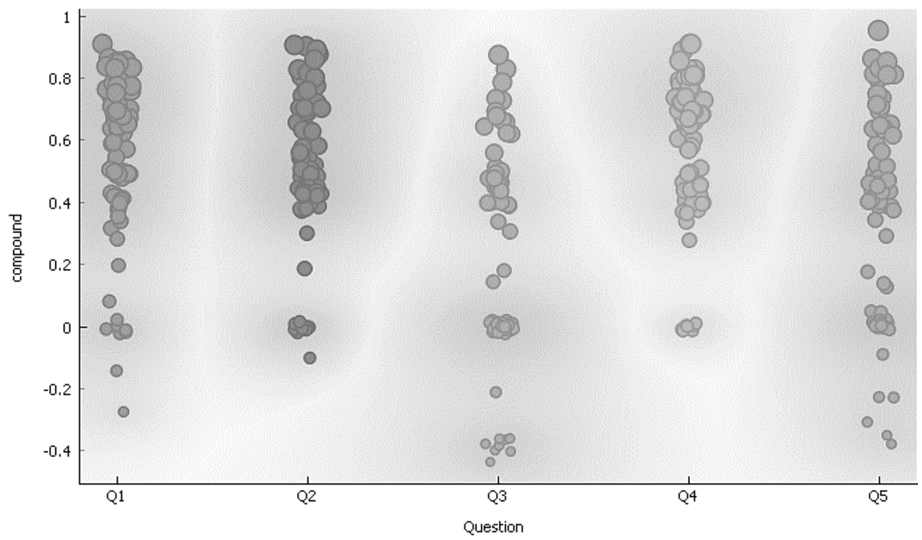


Figure 29. Sentiment score of student comments about a teacher of a particular course (source: Authors' own work).

It is important to note that only a small number of student comments received negative compound intensity for each question. Diagram in Figure 30 depicts the average sentiment score of comments provided by students regarding their teacher's competencies in a particular course.

The student feedback we had available to us was provided to ChatGPT on a question-by-question basis. Table 16 describes the results received from ChatGPT regarding each set of student comments. We observed that the ChatGPT identified with clarity the competencies that the students' comments were referring to, even though the chatbot had never been fed the questions we asked the students. For example, for the first part of the student feedback regarding the "Listening" competence, the chatbot responded to it distinctly as evidenced by the phrase: "It appears that you have provided a large number of statements about a professor's performance, specifically regarding their active listening skills and ability to engage and teach students effectively". Regarding evaluating teacher skills through student feedback, ChatGPT took a specific and clear position identifying our initial question (without having been provided with it).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

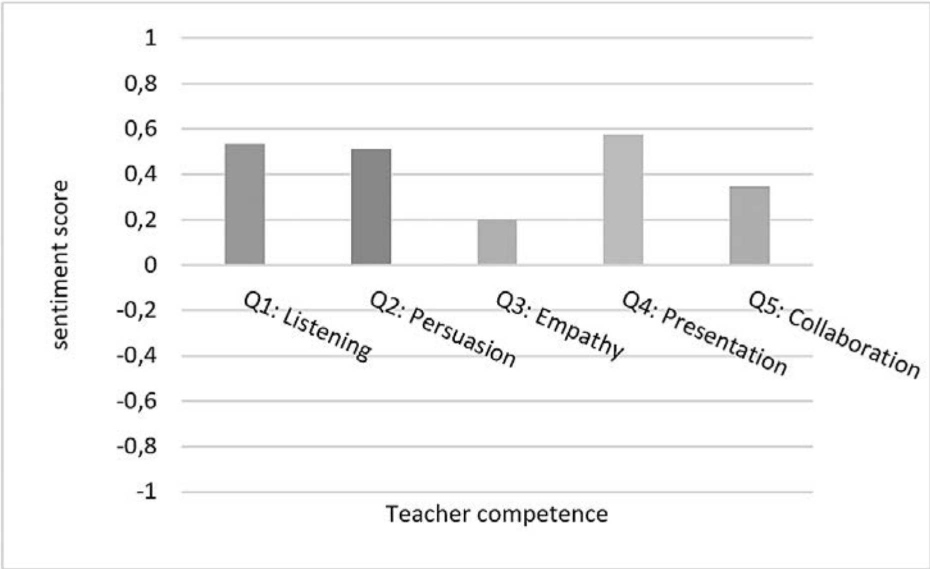


Figure 30. The mean sentiment score of student comments received, regarding their teacher's competencies of a specific course (source: Authors' own work).

More precisely, for the "Listening" competence, it judged that the students feel that their professor listens to them carefully, often displaying the willingness to answer their questions.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 16. The results received from ChatGPT based on student feedback regarding each teacher competence (source: Authors' own work).

Input from human to AI	**ChatGPT response**	**Corresponding question (was initially given to students for feedback collection)**
All student comments referring to Q1	*"It appears that you have provided a large number of statements about a professor's performance, specifically regarding their active **listening** skills and ability to engage and teach students effectively. It seems that most of the comments are positive, with students praising the professor for their approachability, eagerness to teach, and ability to explain things clearly. Some comments mention that the professor is easy to understand and keeps students' attention during lectures, while others highlight the professor's active listening skills and willingness to answer questions and provide feedback. Overall, it seems that the professor is highly regarded by the students."*	Note general comments about your professor regarding their competence in actively **listening** (i.e., making a conscious effort to hear not only the words said but, more importantly, the complete message being communicated).
All student comments referring to Q2	*"It seems that the feedback provided is generally positive towards the professor's teaching skills. Many students mention that he is very convincing and encouraging in class, and his calm way of teaching helps students pay attention. Some students also mention that he has good **persuasion** skills and presents the lessons in an easy-to-understand way. There are also a few suggestions that the professor could do more, but overall the feedback is very positive towards the professor's teaching abilities."*	Note general comments about your professor regarding their powers of **persuasion** used to inspire students and enhance / boost their interest in the subject.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Input from human to AI	ChatGPT response	Corresponding question (was initially given to students for feedback collection)
All student comments referring to Q3	*"Based on the responses, it appears that the professor in question has a lot of **empathy** and understanding towards his students. Students seem to appreciate the fact that the professor tries to put himself in their shoes and understands their needs and circumstances. Some students mentioned that he is approachable, caring, respectful, and willing to listen to their concerns. Overall, the responses are very positive towards the professor's **empathy** and understanding skills."*	Note general comments about your professor regarding the **empathy** felt for their students. (Empathy as a basic dimension in emotional intelligence is the ability to put oneself in another person's shoes and understand what they are feeling).
All student comments referring to Q4	*"Thank you for providing all of those statements about the professor's **presentation** skills. It seems that the overall consensus is that the professor presents the lesson in a clear, simple, and easy-to-understand manner, often using examples to illustrate complex topics. Additionally, it appears that the professor is willing to explain things as many times as needed and uses various tools, such as power point slides and the board, to keep students engaged and interested. Some students note that the professor is very good at explaining topics related to IT and programming, and that they appreciate his willingness to be concise while also providing enough detail to help them understand the material. A few students mention that they would like more analysis or*	Note general comments about your professor regarding their **presentation** abilities (i.e., do they deliver their presentation in a clear and concise manner?).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Input from human to AI	ChatGPT response	Corresponding question (was initially given to students for feedback collection)
	explanation in certain areas, but overall it seems that the professor is effective in communicating the material to the class."	
All student comments referring to Q5	*"Based on the responses, it is clear that the professor in question is highly regarded by the students for his willingness to* ***cooperate*** *and help them. Most of the comments suggest that he is always ready to solve problems and answer any queries, while some also mention his friendly and supportive attitude towards his students. There are a few responses that suggest it is too soon to form an opinion, but even these students note that he appears to be very willing to* ***cooperate****. Overall, the professor seems to be well-liked and respected by his students, and they appreciate his efforts to make the course engaging and understandable."*	Note general comments about your professor regarding their **collaboration** with students (i.e., are they cooperative, do they detect an unwillingness to cooperate and prevent it, do they take into consideration the smooth functioning of the course?).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Having observed the aforementioned table, we requested that the ChatGPT assign a score to correspond with the outcomes presented. Table 17 below shows the average score assigned by chatbot assigned to each teacher competence based on student feedback.

Table 17. The score assigned by ChatGPT regarding each teacher competence (source: Authors' own work).

Teacher competence	Score assigned by ChatGPT (on a scale of 0 to 100)
C22: Listening	75
C23: Persuasion	85
C24: Empathy	65
C25: Presentation	90
C26: Collaboration	80

Considering the above two approaches, we judged that it was valuable to compare them in order to understand their similarities and differences. As different scales were used in each approach, a conversion was done in order for the results to be compared using the same scale. Specifically, the model we designed for SA on the orange platform was based on a scale from -1 to +1, while the ChatGPT on a scale from 0 to 100. Therefore, we applied the following linear transformation to map the range of -1 to +1 onto the range of 0 to 100:

$$y = (x + 1) * 50 \qquad (4.2)$$

Where,

x: the value on the -1 to +1 scale (typical data mining approach)

y: the corresponding value on the 0 to 100 scale (ChatGPT approach)

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

The results are shown in the following Table 18. In addition, we calculated the percent difference between them by applying the following formula:

$$percent\ difference = \frac{|x_1 - x_2|}{[\frac{(x_1 + x_2)}{2}]} x\ 100 \qquad (4.3)$$

Where,

x_1*: the score assigned by sentiment analysis model using the Orange platform converted to a scale of 0 to 100*

x_2*: the score assigned by Score assigned by ChatGPT on a scale of 0 to 100*

Table 18. The score assigned by the sentiment analysis model, converted to a scale from 0 to 100 (source: Authors' own work).

Teacher competence	x_1 **(the score assigned by SA model using the Orange platform converted to a scale of 0 to 100)**	x_2 **(the score assigned by ChatGPT on a scale of 0 to 100)**	**Percent difference (%)**
C22: Listening	77	75	2.6 %
C23: Persuasion	76	85	11.1 %
C24: Empathy	60	65	8 %
C25: Presentation	79	90	13 %
C26: Collaboration	67	80	17.6 %

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

From the two different approaches, it is evident in both cases, that the teacher has strong skills in certain areas such as “Listening”, “Persuasion” and “Presentation” with slight differences, but may need improvement in other areas. More specifically, in the first method, the teacher's “Listening” score was 77, whereas in the second, it was slightly lower at 75. This suggests that both approaches revealed the same intensity, that the teacher demonstrates competence in active listening, as they are able to comprehend and address the needs and concerns of their students. As for “Persuasion”, the teacher scored high in both approaches, with a score of 76 in the first approach and 85 in the second, suggesting that the teacher can effectively communicate and persuade students to engage in learning activities and achieve their goals. However, in both approaches, the teacher's “Empathy” scores were lower, at 65 and 60, respectively. This indicates that the teacher might encounter difficulties in forming an emotional connection with their students, comprehending their emotions and viewpoints, and empathizing with them to understand how they feel. Obviously, according to these results it seems that the teacher has room for improvement and further development of this particular characteristic. Conversely, the teacher received a very high score in the “Presentation” competence in both methods according to student feedback, with a score of 90 in the first approach and 79 in the second. This denotes that the teacher possesses the ability to deliver in captivating and efficient ways that attract student interest and promote learning. Finally, both through the model we developed for the SA of student comments on the Orange platform as well as with ChatGPT, we reached the same conclusions regarding the competence of “Collaboration” that the professor demonstrates to their students.

Summarizing the comparison results, it is evident that the ChatGPT calculations are highly similar to those revealed by the model implemented on the Orange platform, both indicating, that the teacher is generally effective in most competencies, with the highest scores being for “Listening”, “Persuasion” and “Presentation”, and the lowest score being for “Empathy”.

4.2.3.4. Limitations

In this specific study, our data collection was constrained to gathering feedback exclusively from students at UTH for a single course and professor. However, we acknowledge that a more comprehensive collection of student feedback encompassing multiple courses with varying characteristics such as semester, subject area, and course difficulty would necessitate additional time and effort. We believe that such an expanded dataset could reveal additional insights and conclusions of value.

Furthermore, to ensure the accuracy and meaningfulness of student responses, the students expressed themselves in their native language, Greek. Although our model had the ability to rely on a multilingual sentiment lexicon, we chose the reliability of the VADER lexicon. As a result, we

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

were restricted to using the English language, and thus, all the initially collected student comments in Greek were subsequently translated into English by certified translators. Finally, we also limited our research to a specific domain of teacher competencies related to the *Communication* dimension and based on a holistic model of teacher competencies as a continuation of our previous research.

4.2.3.5. Conclusions

In this case, we analyzed data of over 700 feedback comments from students of UTH and evaluated particular competencies by performing SA in two distinct ways; a typical data mining process, using an open source data mining and visualization platform and secondly, the Chat GPT, a generative AI tool. Through our study, we explored and emphasized how textual educational data derived from student feedback can provide valuable insights to academic institutions. Surprisingly, we found that the ChatGPT took a specific and clear position, identifying with great clarity the teacher competencies that the students' comments were referring to, even though the chatbot had never been fed the questions we asked the students, thus strengthening the tool acceptance and its validity. Our research revealed that the ChatGPT calculations are highly similar to those revealed by the model implemented on the Orange platform. More specifically, in both approaches we implemented, the results were almost identical with an index of difference between them ranging on average close to 10%.

4.2.4. Study 4: Comparative SA using lexicon and ML techniques

In this forth case, we conducted a comparative analysis of ML (ML) and Lexicon-based models in order to determine the most effective and efficient method for SA of student responses in course quality questionnaires. The primary questions driving our study of this case, were as follows:

a) Is there a competitive advantage in terms of classification accuracy, multilingual support and context sensitivity of ML approaches over Lexicon-based approaches?

b) How does the inclusion of neutral sentiment class affect the performance of classifiers in each approach?

Through our investigation, we aim to provide insights and recommendations regarding the choice of approach for SA in this specific context.

4.2.4.1. Method

In this section, we provide a comprehensive overview of the stages involved in the process we followed for each SA approach. The visual depiction of these stages is presented in Figure 31. Specifically, in the context of ML algorithms, there is a requirement to incorporate additional steps

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

pertaining to feature selection and model training. Each of these stages are described in detail in the following paragraphs. We have chosen to depict the overall methodology using a circular pattern instead of a linear one to highlight the iterative nature of the procedure, especially in terms of model fine tuning. This circular representation highlights that the different stages of the methodology are iterative rather that one time tasks.

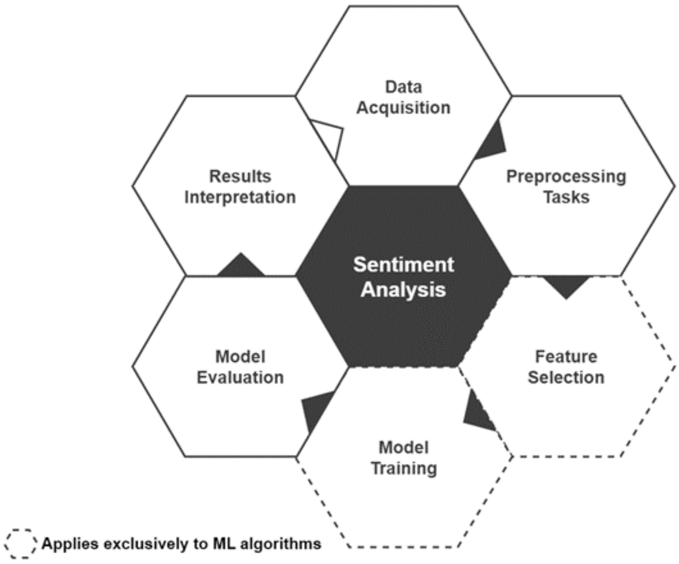


Figure 31. The overall procedure followed.

<u>Stage 1: Data Acquisition</u>

According to our University's regulations, and as part of improving the operations of the QAU, it is essential to evaluate each course and instructor based on student feedback, as well as to thoroughly assess the results of this evaluation. This assessment is conducted through an online questionnaire, which has been carefully designed by the University's QAU. The anonymity of the participating students is fully guaranteed throughout the process.

Specifically, at the end of each semester, students are asked to anonymously complete an online questionnaire for every course and instructor. A link to the questionnaire, aligned with the guidelines of the HAHE, is sent via email to students by the QAU of the UTH. The questionnaire consists of both closed and open-ended questions. The data collected from these responses was stored online and it is accessible to authorized administrators. This dataset was imported in various file formats such as Excel (.xlsx), comma-separated (.csv), native tab-delimited (.tab) files depending on the system used.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

In this case, the data collection was conducted in June 2023 adhered to GDPR to ensure students anonymity. The data used in the present study, consists of text generated by students in response to the following questions:

1) Please provide general comments and suggestions that you would like to convey to your professor.

2) Please provide general comments and suggestions you would like to make regarding the functioning of the course.

After excluding empty responses, we obtained a total of 1420 responses from the initial pool of 1680 students. Out of the 1420 responses, 652 were responses to the first question, while the remaining were responses to the second question. In order to evaluate the performance of the methods we applied, both the lexicon-based and classification algorithms, we needed to have actual sentiment labels for our data available. This allowed us to calculate various performance metrics through the corresponding confusion matrices. These actual sentiment labels, referred to as the "true sentiment" (e.g., positive, negative, neutral), must be predefined either through manual annotation from “evaluators” or reliable sources. At the same time, each method (lexicon or classification algorithms) must generate its own predictions for the sentiment of each data sample (e.g., positive, negative, neutral) to assess its performance. Next, for the calculation of metrics regarding the performance of each SA method, the predicted label is taken as the one generated by the SA algorithm, while the real label is the one pre-determined by the evaluators.

For this purpose, the obtained responses were randomly arranged and made available to two evaluators whose task was to manually classify the text of each response into one of three polarity classes namely, POS (Positive polarity), NEG (Negative polarity) and NEU (Neutral polarity). In cases of disagreement, the corresponding data was forwarded to a third evaluator, whose judgment determined the final label of the data point. The results obtained are presented in Table 19, which shows the distribution of examples across different classes. It is observed that the majority of the available examples belong to the positive class, accounting for 56% of the total. The negative class follows with 24% of the examples, and the neutral class represents 20% of the samples.

Table 19. The distribution of samples across the sentiment classes (source: Authors’ own work).

Actual Class	**Number of Instances**
POS	795
NEG	341
NEU	284

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

The agreement between the first two evaluators was recorded in 80% of the cases while it is worth noting that the majority of cases of disagreement were related to the classification of neutral sentiment, accounting for 62% of the total disagreements. This highlights the additional challenge posed by the inclusion of the neutral class, further complicating an already difficult problem.

Stage 2: Preprocessing Tasks

The SA process operates at both sentence and word levels. Therefore, a step is required to tokenize the sentences into words before proceeding further. Additionally, it is good practice to remove special characters and digits, as well as converting the text to lowercase. There are several steps involved in preparing the text for SA, and the majority of these steps are implemented in the preprocessing stage. Table 20 presents an overview of the preprocessing tasks applied to the textual data in the context of Lexicon-based approach.

Table 20. Preprocessing steps applied in lexicon-based algorithm (source: Authors' own work).

Preprocessing Step	**Description**
Basic text transformation tasks	Convert the text to lowercase or uppercase (if it is deemed that no information will be lost with this conversion) or removing accent marks from the text.
Text Tokenization	The text is segmented into sentences or words using simple rules, such as based on the presence of punctuation marks.
Exclusion of Stopwords	The exclusion of stopwords involves removing commonly used words from the text.
Lemmatization	Lemmatization is the process of mapping all different forms of a word to its base form, known as the lemma.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

In the case of ML algorithms, as part of the preprocessing phase, we employ the use of a pre-trained fastText model to embed the documents from the input corpus into a vector space. This technique enables us to capture the semantic representations of the text, facilitating efficient processing and analysis (Grave et al., 2018).

Stage 3: Feature Selection

Feature selection is a crucial step in machine learning. The quality and relevance of the features used can greatly impact the performance of the modeling algorithm. Having too many features can lead to overfitting and increased computational complexity, while having too few features may result in underfitting and a lack of representative information. Hence, in this step, our objective is to carefully select the most informative features so as to ensure optimal model performance and generalization capabilities while also reducing computational complexity. In the proposed approach, we represent the sentences as vectors using the Sentence-BERT (SBERT) technique (Seo et al., 2022).

Stage 4: Model Training

During this stage, the model learns from the labeled training data and adjusts its internal parameters to make accurate predictions or classifications on unseen data. The training process involves optimizing the model's performance by minimizing the difference between its predicted outputs and the actual labels of the training data. In our approach, we employ the technique of Stratified K-fold Cross-Validation for model training, as proposed by Reimers & Gurevych (2019), and Misra et al. (2017), and we set K=10. Cross-Validation (CV) is widely recognized and commonly utilized for training and evaluating the performance of classifiers in practical scenarios (Arlot & Celisse, 2010). It entails partitioning the dataset into two distinct subsets: the training set and the validation set. During the training process, the classifier is presented with the training data set, while the validation set is utilized for assessing the model's performance.

There are numerous variations of CV, depending on how the dataset is partitioned. One such variation is Stratified CV. In Stratified CV, the sampling of instances for each subset is not random. Instead, the samples are chosen in a way that ensures roughly equal proportions of each class in all subsets. This provides a basic safeguard against imbalanced distributions in the training and validation sets. Through the use of cross-validation techniques, such as Stratified CV, we can effectively assess the performance and generalization capabilities of the model. By training the model on one set of data and evaluating it on a separate set, we can gauge its ability to handle new and unseen instances, and mitigate issues of overfitting. In summary, the use of Stratified CV allows us to obtain reliable estimates of the model's performance and ensure its ability to generalize well to unseen data.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Stage 5: Model Evaluation

One essential step in the evaluation process is measuring the performance of the developed model. The quality of a model can have multiple interpretations, but the most common interpretation is its performance on unseen/real data. Success in this phase depends on two factors: the use of appropriate metrics and the correct interpretation of the results. To ensure the accurate evaluation and ranking of text classification algorithms, we utilize several different metrics that complement each other (Beiranvand et al., 2017). This approach aims to cover various aspects of algorithm performance, including efficiency, reliability, and overall of the solution provided. The selected metrics enable us to assess the effectiveness and robustness of the algorithms, ultimately facilitating the determination of their overall performance. Below, we briefly present the chosen metrics that are recorded throughout this study.

In the case of binary classification, this matrix represents the categories of successes and failures in a 2x2 grid as shown below in Table 21. This matrix provides information about whether certain classes tend to be quality confused with other classes and consequently enables the calculation of the performance metrics (Room, 2019), presented in Table 22.

Table 21. Confusion matrix for binary classification (source: Authors' own work).

		Predicted Category	
		Negative	**Positive**
Actual Category	**Negative**	True Negative (TN) (True negative cases: The number of true negative predictions)	False Positive (FP) (False positive cases: The number of false positive predictions)
	Positive	False Negative (FN) (False negative cases: The number of false negative predictions)	True Positive (TP) (True positive cases: The number of true positive predictions)

In our experiment, we are dealing with three classification classes, which correspond to the Positive, Negative, and Neutral sentiments. Thus, in the case of 3 classes, the confusion matrix becomes a square 3×3 matrix, where the (i,j) element represents the number of instances that

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

belong to class i but are classified as class j (Room, 2019; Grandini et al., 2020). It follows that the elements on the main diagonal correspond to the correct classifier decisions, while the elements outside the diagonal represent the errors (or confusion) of the classifier. Capitalizing on the above, we proceeded to calculate the metrics outlined in the Table 22 that follows.

It is worth note that, to achieve our goal, we needed to write Python code to create a 3x3 confusion matrix. This specific part of code is provided in the appendix section.

Table 22. Metrics for evaluating classification performance (source: Authors' own work).

Metric	**Description**
$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$	Accuracy indicates the percentage of correct predictions made by the classifier. It provides a general assessment of the model's performance across all classes. However, accuracy alone may not be sufficient if the dataset is imbalanced or if different classes have varying degrees of importance.
$\text{Precision} = \frac{TP}{TP + FP}$	A model with low precision is capable of identifying a substantial number of positive examples, but it also exhibits a high rate of false positives, incorrectly classifying many non-positive instances as positive. Conversely, a model distinguished by high accuracy may not succeed in capturing all positive examples, but the instances it categorizes as positive are highly likely to be true positives.
$\text{Recall} = \frac{TP}{TP + FN}$	Recall is a measure that reflects the ability of a classification model to correctly detect positive instances. A model with high recall performs well in detecting positive instances in the data, although it may also incorrectly classify certain negative instances as positive. Conversely, a model with low recall is unable to identify all (or at least a significant portion of) the positive instances.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Metric	Description
$\text{F1 Score} = \frac{2 * Precision\ * Recall}{Precision + \ Recall}$	It is used to balance the results of Accuracy and Recall into a single measurement of algorithm performance. Therefore, it can effectively capture the model's performance even in cases where the data exhibits significant class imbalance. When the F1 score is in the middle range, it indicates that one of the two components (precision or recall) is relatively low, but the metric does not provide a precise breakdown of which one.

4.2.4.2. Implementation

As previously mentioned, we designed our models using the Orange platform. Orange is an open-source platform that enables the implementation of data mining processes, supporting a variety of methods from the fields of statistics and machine learning (Demšar & Zupan, 2013). Additionally, the platform complements these methods with interactive visualization and data exploration tools for the data being processed.

To implement the lexicon-based approach, we start by presenting the text data in the workflow as it shown in Figure 32. Next, we choose to create the corpus which consists of student reviews, and additionally define the dependent variable.

The next steps of our workflow, following our paradigm, consist of the text preprocessing tasks that have been presented in Table 20.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

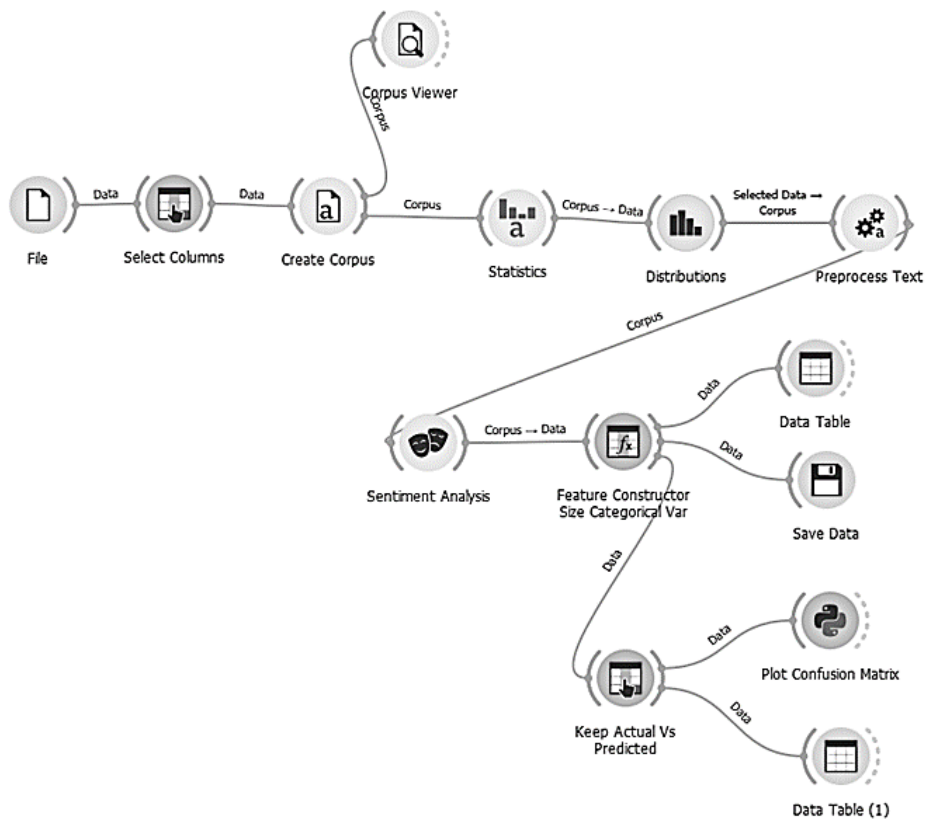


Figure 32. Lexicon-based approach (source: Authors' own work).

After the preprocessing steps the text polarity is being extracted using the VADER lexicon. We judged that we should rely on the VADER lexicon which is fully open-sourced under the MIT License, in the English language, fully validated and widely accepted across multiple distinct domains including the field of education (Navaneetan et al., 2024; Shafana & Safnas, 2022; Wook et al., 2020). Furthermore, the VADER sentiment lexicon has been compared to many other well-established SA lexicons such as Linguistic Inquiry Word Count (LIWC), General Inquirer (GI), Affective Norms for English Words (ANEW), SentiWordNet (SWN), SenticNet (SCN), WordNet and the Hu-Liu, with quite remarkable results. The comparison results revealed that the VADER lexicon performs at a similar level to that of individual human raters. In fact, the classification accuracy metric "Precision" ranged between 0.69 and 0.99 across various domain contexts (Hutto & Gilbert, 2014). Students express themselves in their mother tongue, Greek, to ensure the accuracy and clarity of their sentiments (Sharma et al., 2014). Therefore, to achieve maximum accuracy and efficiency, all the comments were initially collected in the Greek language and subsequently translated into English by accredited translators. In this way, we ensured the highest degree of accuracy both from the students' feedback in their mother tongue, as well as by utilizing the VADER

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

English dictionary which provided us with the validity needed for the purposes of our research. We then entered the final data into our system.

The compound score indicates the overall intensity of sentiment in each sentence. Figure 33 shows a representative sample of comments with the corresponding compound (sentiment) score for each. The compound score of a specific response is determined by summing up the sentiment scores of each individual word within the sentence and falls within the range of -1 to +1 (Qi & Shabrina, 2023).

Students Comment True	pos	neg	neu	compound
I believe he presents the lesson ...	0.175	0.066	0.758	0.3182
I think that students pay attenti...	0.108	0.072	0.821	0.1779
Your active listening capability ...	0.387	0.075	0.538	0.6997
His persuasion skills are satisfa...	0.148	0.076	0.776	0.2115
he has told us very politely that...	0	0.076	0.924	-0.1027
The professor uses everyday to...	0.176	0.088	0.736	0.3678
His calm way of teaching enco...	0.333	0.09	0.577	0.5859

Figure 33. Sentiment score of student feedback comments (source: Authors' own work).

In the second implemented approach (Figure 34), we extract the sentiment carried by the text by applying a ML method. Our goal here is to train a classifier based on the RF algorithm and subsequently gather and analyze the performance results. In the initial stage of the workflow, the file containing the comments is loaded into the workflow, and the data is propagated to the steps of the workflow that facilitate the specification of the dependent variable. Subsequently, the text corpus is created.

Prior to the input of the data to the ML algorithm, a transformation occurs whereby the comments in the corpus are converted into vectors using the Multilingual SBERT (Sentence BERT) approach (Feng et al., 2020). At this stage, in order to examine the significance of each feature generated by the SBERT method, we rank all features using the Gain and Gini metrics (Tangirala, 2020) and keep the top n features. This particular step allows us to exclude features that do not "explain" the dependent variable and therefore do not further enhance the quality of node separation in the decision trees generated by the method. In the experiments conducted, we did not observe any improvement in the results of the method when selecting more than ninety features. We choose to train and evaluate the RF model using the Stratified K-Fold Cross Validation technique. This choice is based on our initial observation regarding the uneven distribution of

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

sentiment classes in our data. Furthermore, we set K=10, an empirically proven value that often yields good results (Anguita et al., 2012).

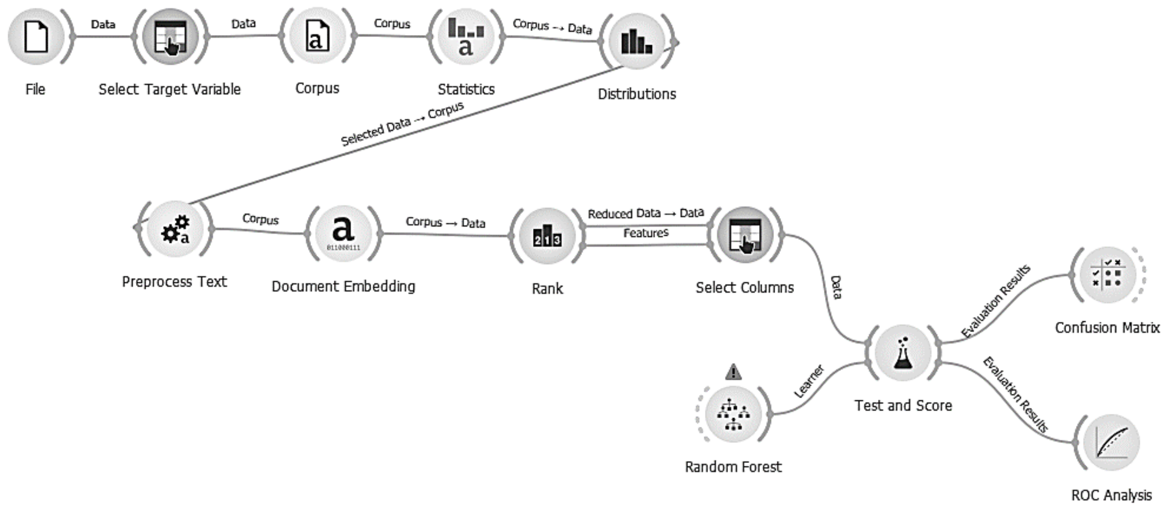


Figure 34. Random forest based approach (source: Authors' own work).

Similar, to the first ML method, the ANN-based method, follows the same preprocessing steps. After creating the corpus and transforming the comments into vectors using the S-BERT technique, the features are ranked based on the Information Gain metric using the "Rank" component. The top sixty features are selected for further analysis. This selection was determined through experimentation, aiming to find the simplest model that maintains a good balance between performance and generalization, effectively reducing the risk of overfitting.

At the final steps of the workflow, the model is trained and evaluated using the Stratified CV technique. Through an iterative experimentation process, we explored different combinations of hyperparameters to identify the best performing configuration for our ANN model. The resulting hyperparameters are presented in Table 23.

Table 23. ANN hyperparameters (source: Authors' own work).

Hyper-Parameter	Value
Neurons in hidden layers	60
Activation Function	Relu

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Hyper-Parameter	Value
Solver	Adam
Regularization, α	0.07
Maximal number of iterations	2500

4.2.4.3. Results and discussion

In this section, we comparatively examine the outcomes of the previously implemented methods with the aim of drawing final conclusions, as well as identifying limitations and prerequisites.

As a general observation (Table 24), we note the superiority of ML methods over the lexicon-based approach. This observation holds true both in terms of overall accuracy and individual evaluation metrics. However, it should be emphasized that the performance of the lexicon-based technique is directly dependent on the style of the analyzed text and the specific lexicon used.

Staying within the lexicon-based technique and delving into our analysis of the degree of granularity of each class, we observe (Table 25) that the model's performance for the POS class is relatively satisfactory, with a Recall of 71.9% and correctly classifying 572 out of 795 comments, the Precision for the same class is 76%, as out of the 752 responses classified as POS, 572 actually belong to that class. However, the model's performance is not at the same level when it comes to classifying negative (NEG) and especially neutral (NEU) responses, where we observe Recall of 57% and 38% respectively.

These values indicate the model's difficulty in recognizing patterns belonging especially to the neutral sentiment class. The same pattern can also be observed in the precision metric. Specifically, as depicted in Table 26, the precision is considerably low in the NEU class. Out of the 291 responses predicted to belong to the NEU class, only 108 were correctly classified, resulting in a precision of 37.1%.

Focusing on the two ML algorithms, we observe that they demonstrate satisfactory performance, with notable differences in capturing the neutral sentiment. Specifically, as shown from Table 28, the ANN consistently performs well in identifying responses from all classes, unlike the RF, which struggled to capture correlations in the data for neutral comments. More specifically, as observed from Table 27, the RF algorithm achieves an overall accuracy of nearly 74% (1042 out of 1420 instances) in correctly predicting the sentiment class. This performance is considered satisfactory given the difficulty of the problem (Roebuck, 2012).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Analyzing the Recall and Precision metrics for each sentiment class, we observe excellent results for the POS and NEG classes, while it is evident that there is difficulty in identifying instances belonging to the neutral class. In the case of the ANN algorithm, we can conclude that the resulting model shows overall satisfactory results. It achieves a classification accuracy of 78%, correctly assigning 1112 out of 1420 comments to their respective classes. The ANN-based classifier performs exceptionally well in identifying positive (POS) and negative (NEG) comments, with a Recall of 82% in both classes.

Table 24. Evaluation results over all classes (source: Authors' own work).

SA Method	**CA**	**F1**	**Precision**	**Recall**
Lexicon	0.616	0.552	0.549	0.556
Random Forest	0.734	0.667	0.718	0.662
Artificial Neural Network	0.782	0.759	0.763	0.759

Table 25. Evaluation results by class (source: Authors' own work).

SA Method	**Precision**			**Recall**		
	POS	NEG	NEU	POS	NEG	NEU
Lexicon	0.760	0.517	0.371	0.719	0.571	0.380
Random Forest	0.747	0.701	0.706	0.933	0.583	0.350
Artificial Neural Network	0.829	0.701	0.761	0.819	0.819	0.640

Table 26. Lexicon confusion matrix (source: Authors' own work).

		Actual Data			
		POS	**NEG**	**NEU**	**sum**
Classifier	POS	572	89	91	752
	NEG	97	195	85	377
	NEU	126	57	108	291
	sum	795	341	284	1420

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 27. RF confusion matrix (source: Authors' own work).

		Actual Data			
		POS	**NEG**	**NEU**	**sum**
Classifier	POS	742	116	135	993
	NEG	37	199	48	284
	NEU	16	26	101	143
	sum	795	341	284	1420

Table 28. ANN confusion matrix (source: Authors' own work).

		Actual Data			
		POS	**NEG**	**NEU**	**sum**
Classifier	POS	649	54	79	782
	NEG	97	281	23	401
	NEU	49	8	182	239
	sum	795	341	284	1420

In other words, the ANN correctly predicts 649 out of 795 positive cases and 281 out of 341 cases belonging to the NEG class. These results are combined with good precision for both classes. Specifically, the model achieves a Precision of 83% for the positive class, 70% for the negative class and 76% for the neutral class. Even in the recognition of neutral messages the ANN observes much better results than the previous classifiers. Specifically, while the Lexicon-based and the RF-based classifier had 38% and 35% Recall respectively, the ANN classifier achieved 64% Recall.

Within the scope of this study, the RF and ANN ML algorithms clearly outperform the lexicon-based methods, as Srivastava et al. (2022), highlights in their study, with the ANN particularly excelling in performance across all sentiment classes. Notably, as an extension of the Wankhade et al., (2022) study, the limitation in the case of the ANN is the availability of a large volume of training data and the risk of overfitting the model to the training data, leading to poor generalization ability. Furthermore, as a continuation of the Adeniyi et al., (2022) study, we observed that the integration of the ANN method combined with sentence embedding, outperformed the other approaches. Table 29 summarizes the pros and cons of each method.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 29. Pros and Cons of each method examined in our study (source: Authors' own work).

Method	**Pros**	**Cons**
Lexicon-based Approach	- Initial simplicity in implementation. - Potential usefulness under specific conditions.	- Limited in handling the introduction of the neutral sentiment class. - Struggles in recognizing patterns within this class, particularly evident in the lexicon-based approach.
Random Forest (RF)	- Demonstrates the ability to model both linear and complex non-linear relationships. - Performance improvement potential with a larger volume of data.	- Faces challenges in recognizing the patterns of the neutral sentiment class. - Interpretability might be a concern.
Artificial Neural Network (ANN)	- Superior overall performance, especially with a large volume of data. - Ability to model complex relationships.	- Interpretability is compromised. - The need for a substantial dataset for optimal performance.

These findings are fully in line with the Adeniyi et al., (2022) study, which thoroughly examined and compared ML algorithms, however, limited to 2 classes: positive and negative. Even though recent review papers (Sen et al., 2017; Leelawat et al., 2022; Mourtzis et al., 2021) have also examined various SA approaches in other domains, none have ventured into the domain of education as from our perspective. Although our study is focused on how the teaching process as well as the teachers are evaluated (identifying and attributing the intensity of sentiment that emerges from a comment regarding their teacher) from the students' point of view, it could be combined, embedding the ANN algorithm which addresses the textual data challenges, with a powerful study conducted by Del Gobbo et al., (2023). In particular, Del Gobbo et al., (2023), in a rather interesting paper, proposed a quite promising framework of the automatic grading of students' answers for exams. Their solution was conceived to offload teachers and instructors from the potentially huge task of grading exam answers to open-ended questions. In summary, our contribution extends prior research by directly applying SA to student textual data, comparing

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

different approaches in a challenging 3 class context, positive, negative and neutral, opening up new avenues for further research and discussion, offering opportunities to refine and improve the learning process, the assessment of teacher competencies etc. based on student opinions.

This study was limited to examining and comparing two different ML algorithms, Artificial Neural Network (ANN) and Random Forest (RF) to a specific lexicon (VADER) for SA on student textual data. This limited scope of algorithms and lexicon examined, may not be representative or cover the diversity of ML algorithms or other lexicon-based methods that may be utilized. However, the selection of the approaches we compared, from which the results were derived, is not exhaustive but covered a substantial range of significant algorithms. In addition, the dataset used, was generated from an educational domain, therefore, the findings may not be entirely generalized to other domains. During the data collection process, the study was also limited to gathering textual data from the student responses of UTH in 2 courses and 2 professors. Although obtaining an extensive collection of student feedback encompassing various courses with distinct characteristics, such as different semesters, cognitive areas, and varying degrees of course's difficulty, may demand extra time and effort, we are confident that it has the potential to unveil significant insights and conclusions.

4.2.4.4. Conclusions

In our research of this case study, we used student comments obtained from feedback at the end of an academic semester and implemented three methods of SA to extract the underlying polarity of each comment. Initially, we examined a Lexicon-based method, followed by corresponding implementations based on ML algorithms, RF and ANN. It was observed that the Lexicon-based SA technique faced challenges in capturing the subtle nuances of the language. It is worth noting that this method did not necessitate complex preprocessing stages, as required by ML methods, because we relied on the existence of updated lexicons that allow for accurate inference. More specifically, the lexicon-based approach struggles to correctly extract meaning in cases where word order and contextual meaning need to be taken into account. In our case, the text rarely included explicit negative words or descriptors. Instead, the negative tone was conveyed through the entirety of the sentence, making it challenging for the Lexicon-based approach to accurately detect the polarity. In contrast, the ML approaches were more effective in their analysis due to their ability to capture the whole meaning of the comment. In short, the linguistic characteristics of the text favored the ML approaches over the lexicon-based method, while the results documented, emphasize the increased difficulty introduced by the inclusion of the neutral sentiment class to the problem. All classifiers struggled to recognize patterns within this class, particularly with the lexicon-based approach.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Another key finding of the research is that the ML algorithms examined performed better in the specific domain of student feedback responses. Both ANNs and RF have the ability to model linear as well as complex nonlinear relationships. However, the ANN algorithm demonstrated better overall performance, which can be even further improved compared to RF, utilizing large volumes of data. We should also highlight the pivotal role of the SBERT in achieving more accurate results. This embedding technique allowed the ML algorithms to capture the semantic meaning and contextual relationships between the comments more precisely while also speeding up the training process. Therefore, the integration of the ANN method combined with sentence embedding, outperformed the Lexicon-based approach as well as the RF method.

We plan to conduct extensive research on various methods of measuring and leveraging student feedback using and comparing different ML algorithms and hybrid approaches adapted to the field of assessment in education. We strongly believe that through our research we can contribute to how institutions face the challenge of harnessing this untapped textual data. Ultimately, there is no intention to undermine or challenge any other perspective. Instead, we want to showcase the variety of different approaches and open up a dialogue to fully explore what's possible in this particular context.

4.3. Validation

This section focuses on validating the competency model we presented earlier. As previously mentioned, through a systematic review process, guided by the PRISMA methodology framework, we identified and categorized 33 distinct competencies within six dimensions. These competencies represent the collective insights into the skills and attributes required for effective teaching in higher education. In the current study, we aim to build on this foundation by exploring competencies from a new perspective—through the lens of student feedback. By applying TM techniques to analyze written comments from student feedback, we seek to uncover the competencies that students value in their educators. We will compare these findings with the competency framework that emerged from our previous research, providing a unique opportunity to validate and expand our understanding of what defines effective teaching in higher education.

At this stage of our thesis, the research questions that guided this case were:

RQ1: What are the key competencies that students perceive as essential for their teachers to possess?

RQ2: To what extent do the competencies identified through the analysis of student feedback align with those documented in existing literature?

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

4.3.1. Method

In this section, we provide a comprehensive overview of the stages involved in the process we followed in our study, which analyzes student feedback comments and identifies teacher competencies through the TM procedure. The initial stage concerns the collection of student comments. More specifically, at the end of each semester at our university, students are invited to answer a questionnaire which contains closed and open-ended questions. For this purpose, the QAU of our UTH, which is the central body for the coordination and support of the quality assurance procedures, sends a relevant link to the students' email. Also, specific measures were implemented to guarantee the anonymity of the students, while also complying with the GDPR. The questionnaire adheres to the requirements of the HAHE. In our study, the data was sourced from evaluations provided by students of the UTH in June 2023, wherein they provided feedback on their course and their instructor. Figure 35 illustrates the stages that we followed. Firstly, we collected a total of 11,735 student responses from our university. 2,870 of those responses -about 25% of the entire sample- included open-ended comments on the course and instructor. This relatively low percentage may reflect students' general reluctance to spend time on open-ended survey questions, as many prefer to only respond to closed-ended, Likert questions, (Baars & Arnold, 2014). Despite this, the 2,870 comments that we obtained provided a robust dataset for TM and were deemed sufficient for generating reliable insights, based on current standards in text analysis research (Jeon & Choi, 2023).

After data collection, the text preprocessing process takes place, which consists of a sequence of steps. Its objective is to prepare the raw text data for the next stage, thus acquiring a clean and consistent form that can then be fed into a model for further analysis. During this procedure, specific NLP issues such as Transformation, Tokenization, Stopwords filtering, Normalization (stemming and lemmatization) etc. should be considered. After preprocessing the student feedback data, we applied the LDA algorithm to extract meaningful topics from the student comments. LDA is a probabilistic TM technique widely used in NLP to uncover hidden structures in textual data. By analyzing the co-occurrence patterns of words across documents, LDA identifies a set of topics, each represented as a distribution of words, and assigns each document a probability distribution over these topics. This approach allowed us to explore the key themes embedded in students' feedback and gain insights into the competencies they associate with effective teaching. The use of LDA provided a systematic and interpretable framework for identifying patterns and trends in the open-ended responses. The topics identified through our approach, were labeled as Topic 1, Topic 2, ..., Topic n, with each topic representing a cluster of words that frequently appeared together, reflecting specific aspects of the competencies students valued in professors.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

In the final stage of fine-tuning the topic model, determining the optimal number of topics (k) was crucial. Selecting an appropriate k value ensured that the topics extracted were both meaningful and distinct. If k is too large, the resulting topics often become redundant, with several topics covering similar content. Conversely, a value that is too small results in overlapping or overly broad topics, failing to capture the nuances and diversity of themes within the text. Thus, finding the right balance was essential to effectively identify and represent the variety of ideas in the dataset.

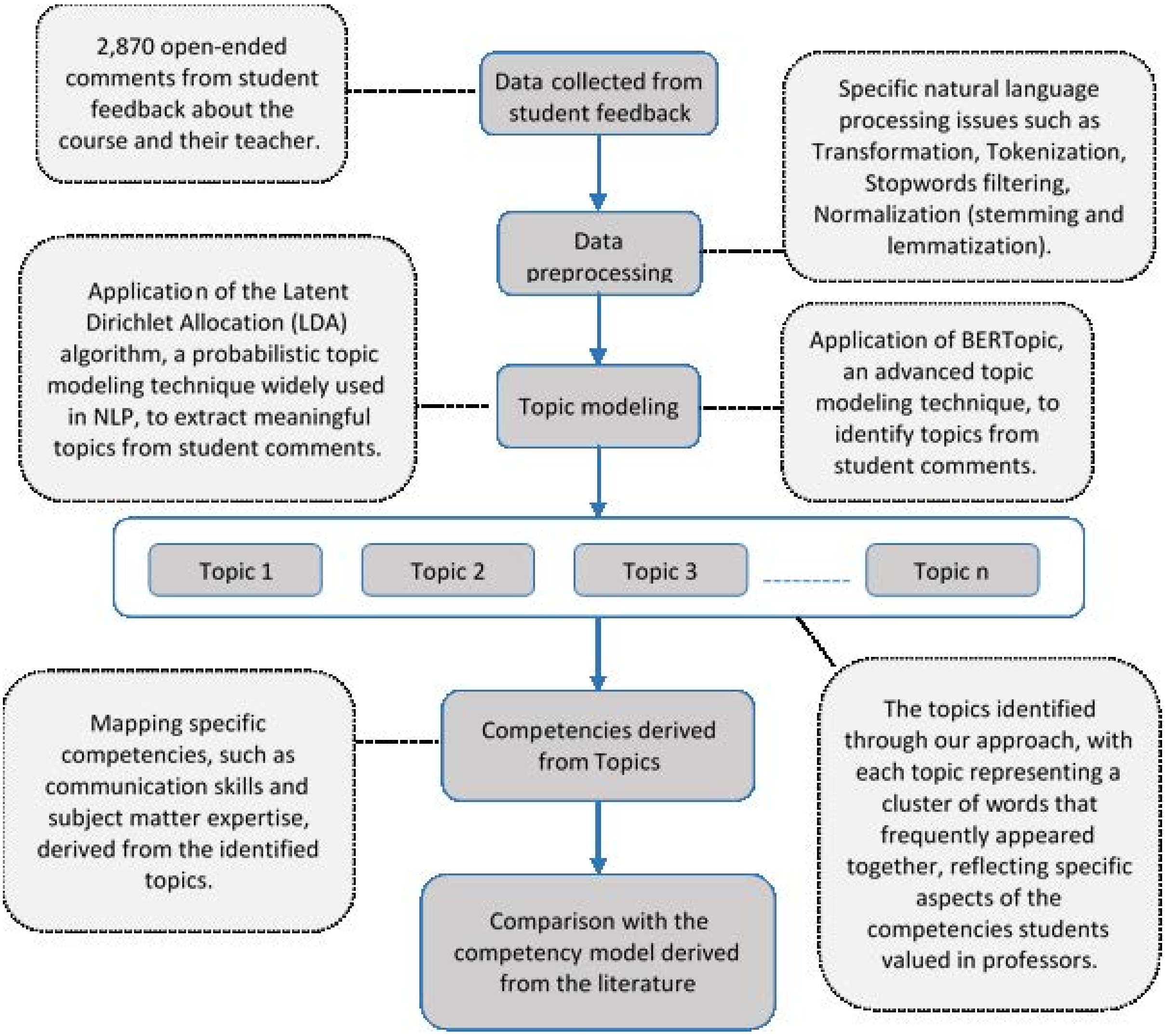


Figure 35. The overall procedure followed (source: Authors' own work).

To determine the optimal k, we experimented with multiple models, each using different values for k. These models were evaluated based on their coherence scores, which measure the semantic relatedness of the words within each topic. A high coherence score indicated that the topics were contextually meaningful and interpretable, as the words within them were strongly related to one

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

another. By comparing coherence scores across the different models, we identified the k value that provided the most stable and cohesive representation of the data (Nguyen and Hovy, 2019; Fang et al., 2016). Additionally, we examined the behavior of coherence scores to find the point where the score either peaked or leveled off, indicating diminishing returns in interpretability. This approach ensured that the selected k value struck a balance between capturing a broad range of themes and maintaining the specificity needed for meaningful insights (Rosner et al., 2014; Syed & Spruit, 2017). This method provided a rigorous framework for determining the number of topics and enabled the extraction of relevant and coherent themes.

Next, we conducted a manual review of each model by examining the keywords associated with each topic. This step allowed us to evaluate the interpretability of the topics and ensure that the selected words accurately represented their underlying themes. For instance, a group of keywords such as "teaching," "students," and "lecture notes" would ideally correspond to a theme related to the use and distribution of lecture notes in education. However, it was crucial to verify that the keywords were contextually coherent and aligned with the intended topic. In addition to this manual inspection, we compared the coherence scores across all models to identify the one that offered the optimal balance between interpretability—how intuitively understandable the topics were—and the overall alignment of the topics with the dataset. The selected model was the one that generated the most coherent and conceptually relevant topics, demonstrating a strong fit between the topics and the underlying data. This systematic process ensured the robustness and reliability of our final model.

After extracting the topics, we mapped them to specific competencies, such as communication skills, subject matter expertise, etc. This mapping process allowed us to derive a set of competencies that reflected students' perceptions of what makes a teacher effective. In the final step of our method, we compared the competencies identified through TM with those derived from our previous research (Dervenis et al., 2022), which focused on teacher competencies as outlined in the literature. This comparative analysis, building on our earlier work, helped us validate the relevance of the competencies identified in this study by placing them within the context of established frameworks from the academic literature. The comparison also highlighted any potential gaps or emerging competencies that might not have been as prominent in the existing literature, offering new insights into the evolving expectations of students in higher education.

4.3.2. Results and discussion

As a result of testing with topic numbers ranging from 1 to 27, we concluded that the optimal number of topics was k=5, where we achieved the highest coherence score (63%). Figure 36 illustrates the coherence values for the different topic numbers tested. The coherence index

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

reflects how well the words within each topic are related, with higher values indicating that the topics are more meaningful and interpretable. In our case, the coherence score of 63% suggests that the topics identified were both relevant and cohesive, effectively capturing distinct topics that were present in the data. This relatively low number of topics can be explained by the limited thematic range of student comments, which are focused specifically on educational matters. Unlike datasets such as newsgroups, where users may discuss a wide array of subjects, this dataset focuses primarily on educational issues, including teaching skills, instructional effectiveness, and the learning environment. This narrower thematic scope likely contributed to the model's improved performance with fewer topics, as fewer thematic categories were necessary to capture the main points of student feedback accurately. Consequently, the coherence score indicates that the selected number of topics offers an optimal balance between interpretability and thematic coverage, making it well-suited for understanding students' core concerns in the educational context.

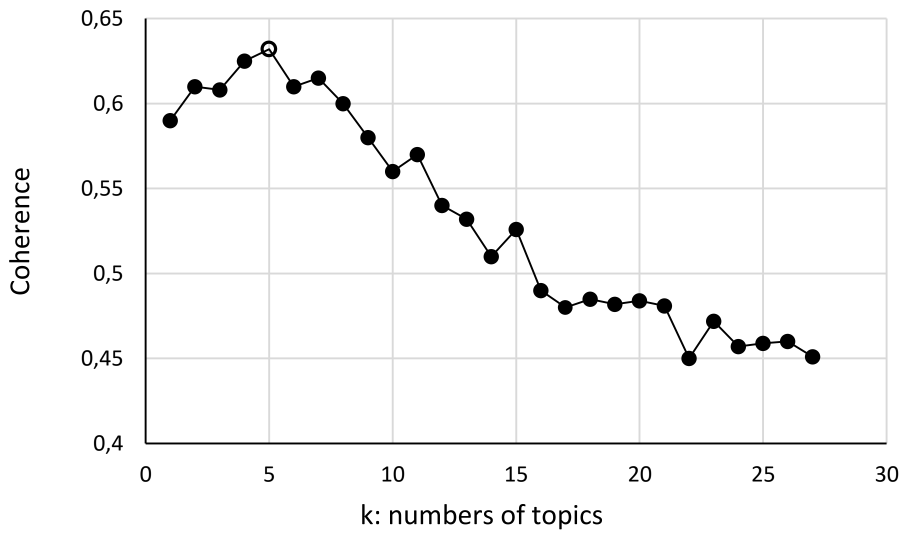


Figure 36. Topic coherence across different k values (source: Authors' own work).

During the TM analysis, the application of a Document Frequency (DF) filter within the range of 0.05–0.95 proved to be instrumental in refining the quality of our results. By excluding terms that appeared in less than 5% of the documents, we eliminated extremely rare words that were unlikely to contribute meaningfully to topic differentiation. Similarly, filtering out terms present in more than 95% of the documents removed overly common words that lacked specificity and were unlikely to help define distinct topics. This preprocessing step effectively reduced noise in the dataset and retained only those words that significantly influenced topic formation. As a result, the

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

use of this filter significantly contributed to achieving the highest coherence score of 63% when we modeled five topics (n=5), highlighting the effectiveness of this preprocessing method in refining the results.

The analysis of the 2,870 student comments revealed that each of the five topics identified, applying the TM methodology described earlier, provided specific aspects of students' experiences regarding their instructors' competencies. To gain a more comprehensive understanding and interpretation of each topic, we examined not only the associated keywords but also the comments with the highest score to each topic. This dual approach allowed us to contextualize the topics and ensure their alignment with broader aspects of student perceptions regarding instructors and coursework. Table 30 provides a detailed mapping of each set of keywords to overarching dimensions related to students' views on their teachers and the learning experience. Subsequently, Table 31 presents the representative student comments for each of the five topics derived from the topic modeling, along with their respective contribution scores to each topic.

Table 30. Identified topics in student feedback through LDA (source: Authors' own work).

Topic	Keywords	Topic interpretation	Topic label assigned
1	exercise, result, theory, lab, lesson, understanding, can, way, thing, give	This topic appears to reflect students' views on the balance between theoretical knowledge and practical application, such as through labs or exercises. It may indicate feedback on how effectively the instructor's approach helps students understand the subject matter and apply it in practice.	Balancing theory and practice
2	good, lesson, quite, interesting, lecture, professor, student, lecturer, difficult, can	This topic focuses on students' evaluation of their teachers, with references to their teaching abilities, approachability, and the difficulty or interest level of the lessons. Words like "good" and "difficult" suggest varying levels of feedback regarding how engaging and manageable the lectures are.	Teaching quality evaluation

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Topic	Keywords	Topic interpretation	Topic label assigned
3	student, lesson, professor, way, polite, question, say, knowledge, always, help	This topic focuses on the interaction between students and instructors, such as addressing student questions and providing support during lessons. It highlights comments about the instructor’s ability to communicate knowledge and their willingness to assist students.	Interactivity and support
4	lesson, assignment, can, time, lab, lecturer, little, topic, well, wanted	This topic addresses students’ perspectives on assignment workload or lab workload and whether the time allocated is sufficient. This may include feedback on whether the workload is manageable and whether time is used efficiently in the course.	Time management and workload
5	lesson, professor, interesting, think, way, understandable, subject, student, well, excellent	This topic reflects the overall quality of the course and the instructor’s ability to make the subject engaging and understandable. Words like "interesting," "understandable," and "excellent" suggest positive evaluations of both the course content and the teaching methods.	Course quality and engagement

The results of the TM analysis were examined in relation to the competencies identified in our prior research, which were derived from an extensive literature review. By analyzing the topics extracted from the students' feedback, specific themes emerged that align with various teacher competencies. These connections highlight how students’ perceptions and comments reflect aspects such as theoretical and practical knowledge, teaching quality, interaction, time management, and course delivery.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 31. Representative student comments for each topic (source: Authors' own work).

Comment ID	Representative Student Comment	Topic 1 score	Topic 2 score	Topic 3 score	Topic 4 score	Topic 5 score
#443	"Unfortunately, the course is very demanding and comes with more workload than it should. I feel that the professor lacks sufficient understanding of our time constraints and capabilities."	0.029	0.030	0.029	0.883	0.029
#601	"The theory lectures were extremely interesting and well-organized. As for the lab sessions, I would have preferred them to be much more interactive and participatory."	0.925	0.019	0.019	0.019	0.018
#1811	"The lectures are excellent, and the professor provides plenty of real-life examples, making a challenging and demanding course much easier to understand!"	0.021	0.918	0.020	0.020	0.021
#2005	"Although Professor L****** is knowledgeable and has a talent for teaching, unfortunately, his arrogance creates significant barriers to effective interaction with students. He is often highly sarcastic and dismissive, constantly adopting an air of superiority that comes across as offensive."	0.010	0.011	0.959	0.010	0.010

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Comment ID	Representative Student Comment	Topic 1 score	Topic 2 score	Topic 3 score	Topic 4 score	Topic 5 score
#2781	"This course was my favorite this semester. Ms. R******* was always enthusiastic and eager to share her knowledge with us throughout the semester. What I particularly appreciated was the relaxed and friendly atmosphere, free from the typical power dynamics between professors and students."	0.015	0.015	0.014	0.014	0.942

The first topic that emerged from the student feedback highlights the importance of balancing theoretical knowledge with practical application, particularly through exercises and lab work. This reflects key competencies identified in our earlier research, such as C6: Theoretical knowledge and C7: Practical knowledge, which emphasize the need to connect theory with real-world experience. Additionally, the C22: Effectiveness competency is apparent in how students value the integration of both theoretical and practical elements in a way that enhances their overall learning experience. Lastly, C13: Planning plays a crucial role here, as students appreciate well-organized lessons that ensure a smooth transition between theory and practice, making the learning process more cohesive and impactful. This connection highlights how thoughtful teaching strategies can support both student understanding and the application of knowledge.

The second topic from the student feedback focuses on their evaluation of the overall teaching quality, with students highlighting factors, such as how engaging, accessible, or challenging the lectures are. This aligns with several key competencies from our previous research, such as C11: Transmissibility, which reflects the professor's ability to present information in a clear and engaging way. Additionally, C10: Creativity plays a critical role, as students appreciate lectures that are designed to keep them interested and actively engaged. The importance of C12: Studiousness is also evident, as students value the professor's preparation and commitment to delivering high-quality content. Lastly, C25: Empathy is highlighted in how well professors recognize and address students' difficulties in grasping complex material, offering support where needed. This connection

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

illustrates how teaching practices that focus on engagement, creativity, and empathy can significantly enhance the students’ learning experience.

The third topic that emerged from the student feedback emphasizes the importance of teacher-student interaction, particularly in terms of answering questions and providing ongoing support. This is closely related to several competencies, such as C2: Helpfulness, which reflects the professor’s willingness to assist students and clarify any doubts they may have. The competency of C23: Listening is also key, as students appreciate it when teachers actively listen to their concerns and feedback. In addition, C25: Empathy plays a vital role, with students valuing teachers who demonstrate an understanding of their individual needs and provide tailored support. Finally, C15: Coaching is reflected in the encouragement and guidance teachers offer, helping students navigate their learning journey and overcome challenges along the way. This connection highlights how meaningful interaction and support can significantly enhance students' academic experiences.

The fourth topic from the student feedback highlights concerns about time management and the workload they are expected to manage. This connects to several competencies that emphasize the importance of structure and organization in the learning process. C13: Planning is evident here, as students value effective time management to balance the workload and ensure tasks are completed on time. C9: Discipline also plays a role, with students appreciating a clear structure and adherence to deadlines for assignments and classwork. Lastly, C22: Effectiveness is key, as students recognize when time and resources are used efficiently, enhancing their overall learning experience and helping them achieve better outcomes. This connection underscores how well-planned courses can alleviate student stress and foster a more productive learning environment.

The fifth topic from the student feedback focuses on the overall quality of the course and the instructor’s ability to make the subject both engaging and understandable. This topic connects strongly to several competencies, starting with C11: Transmissibility, as students appreciate it when complex topics are simplified and made accessible. C18: Innovativeness is also reflected in how students value creative teaching methods that make the course more engaging and dynamic. Additionally, C25: Empathy comes into play, as students recognize when instructors understand their challenges and adapt their teaching strategies to better meet their needs. Finally, C26: Presentation is key, as students appreciate well-structured, clear, and captivating presentations that make the learning process more enjoyable and effective. This connection emphasizes how the quality of the course and the teaching approach can significantly influence students' engagement and overall satisfaction. Table 32, illustrates the mapping of the identified topics to the corresponding competencies, offering a clear representation of how the textual data aligns with the conceptual framework established in our previous study.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 32. Mapping results: Topic modeling (LDA) insights vs. literature-derived competency framework (source: Authors' own work).

Topic	Topic label assigned (from textual data)	Mapped competencies (LDA-based topic modeling vs. literature-based framework)	Not mapped competencies
1	Balancing theory and practice	C6: Theoretical knowledge C7: Practical knowledge C22: Effectiveness C13: Planning	C1: Energy C3: Personal integrity and reliability C4: Multiculturalism C5: Morality C8: Experience C14: Adaptability C16: Inventiveness C17: Resourcefulness C19: Imagination C20: Diligence C21: Persistence C24: Persuasion C28: Digital literacy C29: Digital pedagogy C30: Digital communication and cooperation C31: Creating digital content C32: Safety C33: Basic troubleshooting
2	Teaching quality evaluation	C11: Transmissibility C10: Creativity C12: Studiousness C25: Empathy	
3	Interactivity and support	C2: Helpfulness C23: Listening C25: Empathy C15: Coaching	
4	Time management and workload	C13: Planning C9: Discipline C22: Effectiveness	
5	Course quality and engagement	C11: Transmissibility C18: Innovativeness C25: Empathy C26: Presentation	

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

The next step was the application of BERTopic algorithm. BERTopic is an advanced topic modeling technique, to identify topics from student comments. The algorithm was developed in Python using Google Colab online environment[2], which provides easy access to powerful computational resources and supports the BERTopic library (Bisong & Bisong, 2019). Google Colab allowed us to perform the necessary text analysis tasks in a flexible and user-friendly environment, enhancing the efficiency of the topic modeling process and ensuring effective management of our data. The developed code is presented in in the appendix of this thesis. We used BER Topic algorithm to extract topics from the student comments, providing us with a deeper understanding of the key competencies that students value in their teachers. Table 33 provides a detailed mapping of each set of keywords to overarching dimensions related to students' views on their teachers and the learning experience.

Table 33. Identified topics in student feedback through BERTopic (source: Authors' own work).

Topic	Keywords	Topic interpretation	Topic label assigned
1	teacher, subject, interest, course, students, subject matter, good, professor	This topic highlights the importance of the instructor's ability to engage students with the subject matter. It focuses on the teacher's role in creating an interesting and relevant learning experience. Students seem to appreciate teachers who are able to make the course content engaging and meaningful, demonstrating the connection between teaching and student involvement.	Student Engagement and Teaching Effectiveness
2	good, lesson, quite, interesting, lecture, professor, student,	This topic underscores the teacher's involvement in managing the course content and engaging with students. It reflects students' recognition of a well-structured course, with teachers making the material both accessible and interesting. There is an emphasis on the	Effective Teaching Practices

[2] https://colab.research.google.com/

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Topic	Keywords	Topic interpretation	Topic label assigned
	lecturer, difficult, can	quality of the course delivery and its impact on student engagement.	
3	student, lesson, professor, way, polite, question, say, knowledge, always, help	This topic revolves around the use of online platforms like e-class for sharing materials and resources. It highlights the importance of digital tools in teaching, with a focus on how the teacher organizes and provides resources such as slides and course material to enhance learning. Students appear to value easy access to learning materials online, as well as the clarity and thoroughness of these resources.	Digital Literacy and Resource Availability
4	lesson, assignment, can, time, lab, lecturer, little, topic, well, wanted	This topic reflects students' perspectives on the teacher's approach to course difficulty and class management. The keywords suggest that some students perceive certain aspects of the course or instructor's teaching style as challenging. The topic touches on both the teacher’s ability to manage the course effectively and the students’ experiences with difficult or demanding courses.	Classroom Management and Teaching Difficulty

Subsequently, Table 34 presents the representative student comments for each of the five topics derived from the topic modeling, along with their respective contribution scores to each topic.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Table 34. Mapping results: Topic modeling (BERTopic) insights vs. literature-derived competency framework (source: Authors' own work).

Topic	**Topic label assigned (from textual data)**	**Mapped competencies (BERTopic-based topic modeling vs. literature-based framework)**	**Not mapped competencies**
1	Student Engagement and Teaching Effectiveness	C2: Helpfulness C11: Transmissibility C23: Listening C25: Empathy C27: Collaboration	C1: Energy C3: Personal integrity and reliability C4: Multiculturalism C5: Morality C6: Theoretical knowledge C7: Practical knowledge C8: Experience C10: Creativity C15: Coaching C16: Inventiveness C18: Innovativeness C19: Imagination C24: Persuasion C26: Presentation C32: Safety C33: Basic troubleshooting
2	Effective Teaching Practices	C12: Studiousness C13: Planning C22: Effectiveness C28: Digital literacy C29: Digital pedagogy	
3	Digital Literacy and Resource Availability	C28: Digital literacy C29: Digital pedagogy C31: Creating digital content C30: Digital communication and cooperation	
4	Classroom Management and Teaching Difficulty	C14: Adaptability C9: Discipline C20: Diligence C21: Persistence C17: Resourcefulness	

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Both approaches, LDA and BERTopic, revealed a significant overlap in identifying key teacher competencies. Both methods successfully captured a range of competencies that were consistent with those identified in our previous research, which was based on an in-depth literature review. The following Table 35, clearly illustrates these connections, demonstrating how the topics derived from both TM approaches align with the competency framework established earlier. This alignment not only strengthens the validity of our model but also validates the effectiveness of the competencies identified through the literature review, further confirming their relevance and importance in enhancing teaching practices and educational quality. However, although our analysis successfully identified competencies of teachers that students are interested in or simply observe, as well as those documented in the literature, some competencies from the original framework—such as “Multiculturalism,” “Experience,” and “Inventiveness”—were not identified in the students' comments. This is not a concern but can be attributed to the nature of open-ended feedback data, which often reflects immediate and tangible aspects of teaching rather than less obvious characteristics. Clearly, additional competencies, previously missing, could emerge from other data sources beyond student feedback.

Table 35. Comparison of teacher competencies identified through LDA and BERTopic with the Literature-based framework (source: Authors’ own work).

Teacher competencies identified in Literature review	**Competencies identified from LDA approach**	**Competencies identified in BERTopic approach**
C1: Energy		
C2: Helpfulness	✓	✓
C3: Personal integrity and reliability		
C4: Multiculturalism		
C5: Morality		
C6: Theoretical knowledge	✓	
C7: Practical knowledge	✓	

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Teacher competencies identified in Literature review	Competencies identified from LDA approach	Competencies identified in BERTopic approach
C8: Experience		
C9: Discipline	✓	✓
C10: Creativity	✓	
C11: Transmissibility	✓	✓
C12: Studiousness	✓	✓
C13: Planning	✓	✓
C14: Adaptability		✓
C15: Coaching	✓	
C16: Inventiveness		
C17: Resourcefulness		✓
C18: Innovativeness	✓	
C19: Imagination		
C20: Diligence		✓
C21: Persistence		✓
C22: Effectiveness	✓	✓
C23: Listening	✓	✓
C24: Persuasion		
C25: Empathy	✓	✓
C26: Presentation	✓	

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Teacher competencies identified in Literature review	Competencies identified from LDA approach	Competencies identified in BERTopic approach
C27: Collaboration		✓
C28: Digital literacy		✓
C29: Digital pedagogy		✓
C30: Digital communication and cooperation		✓
C31: Creating digital content		✓
C32: Safety		
C33: Basic troubleshooting		

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

5. Discussion, Conclusions and Recommendations

5.1. Introduction

This chapter presents the conclusions of this research through a detailed discussion of its key findings. It then highlights the limitations encountered during the study, as well as the implications for both academia and society. Finally, suggestions for future research are provided to inspire the research community.

5.2. Conclusions

In this study, we aimed to evaluate and enhance academic teaching excellence through a competency-based teacher model. The research revealed a multifaceted perspective on how both students and established theoretical frameworks perceive key teaching competencies. One critical conclusion we reached is the alignment between student feedback and competencies identified in the literature, particularly in certain areas.

Our analysis also underscored the importance of a multidimensional approach to teacher competencies. Specifically, 33 distinct competencies were identified, categorized into six main dimensions: *Personality*, *Professionalism*, *Pedagogical*, *Scientific*, *Communication* and *Digital*. These competencies reflect the broad spectrum of skills that define effective teaching.

Our results highlighted that, despite the increasing importance of digital competence—which became particularly urgent following the COVID-19 pandemic—it remains insufficiently prioritized within the educational community. While education systems are transitioning toward online and hybrid learning, especially in recent years, a surprising number of studies continue to treat digital competence as secondary to more traditional aspects of teaching. This finding was particularly evident in our study, where areas such as *Personality*, *Communication*, and *Professionalism* received more attention, while digital competence was often relegated to the background. This gap in educational practices suggests a broader issue: the slow adaptation of educational communities to the rapid pace of technological advancement.

Along with the multidimensional framework for teacher competencies, we revealed that SA can assist in capturing student perceptions of these competencies. A key aspect of our work involved applying SA techniques to analyze open-ended student feedback. Unlike traditional closed-ended feedback, which tends to oversimplify responses into predefined categories, we found that open-ended responses provide rich, nuanced insights that reflect the complexity of student experiences with teaching. The analysis of student feedback revealed specific competencies that students

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

perceive in their instructors. Through this analysis, we concluded that students recognize particular competencies differently depending on the course's level of difficulty. For instance, in easier courses, competencies related to "Presentation" and "Persuasion" were more pronounced among students. On the other hand, in more demanding courses, these competencies were less emphasized, while others such as "Listening" and "Empathy" gained significance across all course types, regardless of difficulty. This suggests that these specific competencies are widely valued by students, irrespective of the subject matter or the course's complexity. We therefore concluded that students tend to appreciate teachers who can communicate effectively and be persuasive in simpler courses, while in more demanding ones, understanding and empathy take precedence.

Additionally, one of the most intriguing findings of our research emerged from the comparison between SA using traditional lexicon-based methods and ML algorithms, such as Random Forest (RF) and Artificial Neural Networks (ANN). While lexicon-based methods were easy to implement, they struggled to capture the full meaning of student comments, particularly when the feedback conveyed nuances or implicit sentiments. In contrast, ML algorithms, particularly ANN, performed significantly better, capturing both the explicit and subtler aspects of student feedback. This highlights the advantages of ML techniques in handling complex educational data, where context and sentiment often play a critical role in analysis.

By applying an AI-generative tool to analyze the same dataset, we found that the results closely aligned with those of the lexicon-based approach, with only a small margin of difference between the two methods. This consistency underscores the potential of AI tools in educational research, as they can process and analyze large volumes of qualitative data with high accuracy. Moreover, AI's ability to provide cohesive insights from feedback reinforces the value of integrating AI into educational research and data analysis processes.

Another critical aspect of this research was the validation process, where we assessed the robustness and reliability of the competency model derived from the literature, by comparing it with findings from data analysis using TM algorithms. Specifically, the use of TM techniques such as LDA and BERTopic allowed us to validate our model, by examining whether the competencies identified in the literature are genuinely reflected in student perceptions. Through the validation process, our model proved highly effective in capturing student sentiments regarding key competencies such as *Empathy*, *Presentation*, *Communication*, and others. Furthermore, ML and TM techniques enhanced our understanding of how these competencies are perceived in practice, providing additional validation for the competency framework and its applicability in real-world teaching conditions.

Overall, our findings confirm the importance of teaching competencies and how they can be measured through innovative methods such as SA and tools based on ML. We observed that a wide

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

range of competencies plays a critical role in shaping students' educational experiences. However, it is also evident that certain competencies, such as digital competence, still require further attention and development in both research and practice. The findings of this study highlight the need for ongoing efforts to redefine teaching competencies in HE aiming to adapt to the ever-evolving landscape of education and technology. In conclusion, this study provides valuable insights into how students perceive teacher competencies and how these perceptions can be captured and analyzed through advanced data analysis techniques. The findings emphasize the importance of leveraging student feedback, particularly open-ended comments, to gain a more comprehensive understanding of teaching effectiveness.

As educational institutions continue to navigate the complexities of education in contemporary society, it is essential to recognize the value of competencies beyond traditional frameworks and to integrate emerging competencies, such as digital literacy, into teacher development programs. By doing so, institutions can ensure that their educators are well-equipped to address the challenges of a rapidly evolving educational environment.

5.3. Limitations and implications

While this study provides valuable insights into teacher competencies through the analysis of student feedback, several limitations must be acknowledged. First, the data collection process was limited to student feedback within a single university setting. Although this dataset yielded significant findings, expanding the scope to include multiple courses at a number of educational institutions with varying characteristics—such as semester, subject area, and course difficulty—would allow for more reliable conclusions. A broader dataset would also enhance the generalizability of the results, offering a more comprehensive understanding of teacher competencies across diverse educational environments.

Second, the reliance on open-ended student feedback introduces an inherent limitation, as not all competencies can be easily articulated or identified by students. Competencies such as "Diligence" or "Digital Literacy," which involve consistent effort, persistence, or technological proficiency, may not be explicitly expressed unless students directly observe or experience their presence (or absence) in teaching practices. This highlights the potential constraints of relying solely on student perspectives to determine competencies. Future research could address this gap by incorporating complementary data sources, such as interviews with educators, academic administrators, or policymakers, to establish a more holistic framework.

Another limitation relates to the language choices in our study. The data, consisting of student comments, were initially collected in their native language, Greek, to ensure the authenticity of

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

sentiment and meaning. However, due to the use of the VADER lexicon, which is optimized for English text, all responses had to be translated into English by certified translators. While every effort was made to ensure accuracy in translation, nuances of sentiment and meaning may have been misrepresentedduring this process. Although multilingual sentiment lexicons could be explored in future studies, the decision to use VADER was based on its proven reliability and accuracy in SA.

Regarding methodology, this study dealt with the challenge of employing specific ML algorithms, such as Artificial Neural Network (ANN) and Random Forest (RF), alongside the VADER lexicon-based SA approach and the LDA and BERTopic algorithms for topic modeling, as well as the K-Nearest-Neighbors (KNN) and SVM classification algorithm. While these approaches provide meaningful insights, they do not exhaust the full spectrum of ML techniques or available lexicon-based methods.

From a practical perspective, the study focused on a framework of general teacher competencies, excluding other areas of expertise that are equally important. In this study, however, in an effort to maintain a homogeneous base of teacher competencies, inclusion of domains of expertise was avoided.

Despite the limitations outlined above, this study has significant implications for research in the rapidly developing field of education, as well as practical and social implications, which are summarized as follows:

- Research Implications: The findings of this study highlight the importance of student feedback as a valuable source for evaluating and improving teaching practices. By identifying competencies that students value, educators and institutions can enhance their approaches to improve teaching effectiveness and align instructional strategies with student needs. However, the study also emphasizes the need for a multifaceted approach to competency evaluation, incorporating diverse perspectives to ensure a more comprehensive assessment.

- Practical Implications: The development of a structured teacher competency framework provides a useful reference for improving the quality of HE in multiple domains. Such a framework can inform teacher recruitment, professional development initiatives, and national education policies aimed at ensuring teaching excellence. Moreover, it facilitates dialogue among stakeholders—including educators, policymakers, and academic administrators—on quality assurance and evaluation processes in education.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

- Social Implications: By focusing on the competencies not only valued by students but those integral to the learning process in a rapidly evolving world, this research supports the development of engaging and supportive learning environments that positively influence educational outcomes. In addition, teacher competencies play a central role not only in fostering effective teaching and learning environments but also in preparing individuals to contribute meaningfully to the labor market and society. Subsequently, a community equipped with essential skills and values is better positioned to proactively address challenges, driving social progress and achieving optimal outcomes.

5.4. Recommendations and future research

Based on the findings of this study, potential research that could further contribute to enhancing teaching quality and learning outcomes emerge. First, future research should focus on designing and implementing a more holistic framework that captures a broader range of teacher competencies and explores their relationship with the students' academic achievements. This expanded approach could address competencies that may currently be underrepresented or overlooked due to the absence of their identification. By incorporating them into academic frameworks, a more comprehensive and nuanced understanding of teaching effectiveness can be developed.

Additionally, further exploration of methods for measuring teaching competencies and utilizing student feedback is necessary. While this study demonstrated the value of using ML algorithms and TM techniques to analyze open-ended textual data, future research should incorporate and compare additional advanced methods. More specifically, this includes hybrid approaches, aspect-based SA, and AI-based systems specifically tailored to the field of education. By evaluating the performance of the various techniques that emerged from this research, scholars can identify the most effective and accurate tools for extracting information from untapped textual data.

Furthermore, the integration of AI chatbots represents a promising avenue for simplifying the analysis process. These tools could automate the collection, categorization, and interpretation of large volumes of student feedback, allowing institutions to gain valuable insights in a timely and systematic manner. It should also be mentioned that future studies need to explore the feasibility and reliability of such systems, assessing their ability to streamline feedback analysis while maintaining accuracy and consistency.

A further critical recommendation is to broaden the scope of student feedback analysis to include different education contexts, encompassing various courses, disciplines, and institutions. By analyzing larger and more diverse datasets, researchers can uncover patterns and relationships

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

that may vary across different cognitive domains, course difficulties, and student demographics. This broader perspective would not only validate existing findings but also highlight potential opportunities for improving teaching practices.

Lastly, raising awareness within the educational community remains a key priority. It is vital that institutions and educators recognize that vast amounts of textual data from student feedback are often collected but remain underutilized. Promoting the value of tapping into this resource will encourage collaboration between researchers, policymakers, and educators, fostering a culture of evidence-based decision-making aimed at enhancing the quality of teaching and learning.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

References

Adeniyi, J. K., Adeniyi, A. E., Oguns, Y. J., Egbedokun, G. O., Ajagbe, K. D., Obuzor, P. C., & Ajagbe, S. A. (2022). Comparison of the performance of machine learning techniques in the prediction of employee. ParadigmPlus, 3, pp. 1–15.

Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. Procedia Computer Science, 127, pp. 511–520.

Ali, W. (2019). The Efficacy of Evolving Technology in Conceptualizing Pedagogy and Practice in Higher Education. Higher Education Studies, Vol. 9 No. 2, pp. 81–90.

Ali, W. (2020). Online and Remote Learning in Higher Education Institutes: A Necessity in light of COVID-19 Pandemic. Higher Education Studies, Vol. 10 No. 3, pp. 16–24.

Al-Rahmi, W., Aldraiweesh, A., Yahaya, N., Kamin, Y. B., & Zeki, A. M. (2019). Massive open online courses (MOOCs): Data on higher education. Data in brief, 22, pp. 118–125.

Altrabsheh, N., Gaber, M. M., & Cocea, M. (2013). SA-E: sentiment analysis for education. In Intelligent Decision Technologies. IOS Press, pp. 353-362.

Alvarez, I., Guasch, T. and Espasa, A. (2009). University teacher roles and competencies in online learning environments: A theoretical analysis of teaching and learning practices. European Journal of Teacher Education, Vol. 32 No. 3, pp. 321–336.

Amin, A., Hossain, I., Akther, A., & Alam, K. M. (2019, February). Bengali vader: A sentiment analysis approach using modified vader. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, pp. 1–6.

Anderson, L. (2004). Increasing teacher effectiveness. FUndamentals of Educational Plannning, Vol. 57, pp. 1–171.

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012, April). The'K'in K-fold Cross Validation. In ESANN, pp. 441-446.

Arlot, Sylvain, and Alain Celisse (2010). A survey of cross-validation procedures for model selection. Statist. Surv. 4, pp. 40–79.

Asghar, M. Z., Ullah, I., Shamshirband, S., Kundi, F. M., & Habib, A. (2019). Fuzzy-based sentiment analysis system for analyzing student feedback and satisfaction.

Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN computer science, 3(1), pp. 74–81.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Babu, S., & Mendro, R. (2003). Teacher accountability: HLM-based teacher effectiveness indices in a state assessment program. In annual meeting of the American Educational Research Association, Chicago.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec Vol. 10, No. 2010, pp. 2200–2204.

Bakhru, K.M. (2013). A Principal Component Analysis of Teaching Competencies Required for Management Education. APJEM Arth Prabhand: A Journal of Economics and Management, Vol. 2 No. 7, pp. 1–15.

Bakhru, K.M. (2017). Personal competencies for Effective teaching: A Review Based Study. Educational Quest- An International Journal of Education and Applied Social Sciences, Vol. 8 No. spl, pp. 297- 308.

Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: A comparison of retrieval performances.

Barth, M., Godemann, J., Rieckmann, M. and Stoltenberg, U. (2007). Developing key competencies for sustainable development in higher education. International Journal of Sustainability in Higher Education, Vol. 8 No. 4, pp. 416–430.

Beckwith, R., Fellbaum, C., Gross, D., & Miller, G. A. (2021). WordNet: A lexical database organized on psycholinguistic principles. In Lexical Acquisition, Psychology Press, pp. 211-232.

Beiranvand, V., Hare, W., & Lucet, Y. (2017). Best practices for comparing optimization algorithms. Optimization and Engineering, 18(4), pp. 815-848.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. Information Sciences, 191, 192-213.

Bergsmann, E., Schultes, M.T., Winter, P., Schober, B. and Spiel, C. (2015). Evaluation of competence-based teaching in higher education: From theory to practice. Evaluation and Program Planning, Elsevier Ltd, Vol. 52, pp. 1–9.

Bisong, E., & Bisong, E. (2019). Google colaboratory. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners, pp. 59–64.

Blašková, M. (2014). Influencing Academic Motivation, Responsibility and Creativity. Procedia - Social and Behavioral Sciences, Elsevier B.V., Vol. 159, pp. 415–425.

Blašková, M., Blaško, R. and Kucharčíková, A. (2014). Competences and Competence Model of University Teachers. Procedia - Social and Behavioral Sciences, Vol. 159, pp. 457–467.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), pp. 993–1022.

Bryman, A. (2016). Social research methods. Oxford university press.

Caena, F. (2011). Literature review Teachers’ core competences: requirements and development. Education and Training Hematic Working Group ‘Professional Development of Teachers, No. January 2011, pp. 1–28.

Caena, F. (2013). Supporting Teacher Competence Development for Better Learning Outcomes. European Commission, Education and Training, pp. 41–49.

Caena, F. and Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (Digcompedu). European Journal of Education, Vol. 54 No. 3, pp. 356–369.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. The Journal of Machine Learning Research, 11, pp. 2079-2107.

Cebrián, G., Junyent, M. and Mulà, I. (2020). Competencies in education for sustainable development: Emerging teaching and research developments. Sustainability (Switzerland).

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, pp. 408-419.

Chakraverty, S., & Saraswat, M. (2017). Review based emotion profiles for cross domain recommendation. Multimedia Tools and Applications, 76, pp. 25827-25850.

Chen, C. C., & Tseng, Y. D. (2011). Quality evaluation of product reviews using an information quality framework. Decision Support Systems, 50(4), pp. 755-768.

Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. Tertiary education and management, 11(1), pp. 19-36.

Cochran-Smith, M., Feiman-Nemser, S., McIntyre, D.J. and Demers, K.E. (2008). Handbook of Research on Teacher Education: Enduring Issues in Changing Contexts, London: Routledge Press., available at:https://doi.org/10.5860/choice.28-1672.

Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers-a tutorial. ACM computing surveys (CSUR), 54(6), pp. 1-25.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Del Gobbo, E., Guarino, A., Cafarelli, B., & Grilli, L. (2023). GradeAid: a framework for automatic short answers grading in educational contexts—design, implementation and evaluation. Knowledge and Information Systems, pp. 1-40.

Delaney, J.G., Johnson, A., Johnson, T.D. and Treslan, D. (2010), “Students’ Perceptions of Effective Teaching in Higher Education.”

Demšar, J., & Zupan, B. (2013). Orange: Data mining fruitful and funa historical perspective. Informatica, 37(1).

Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., ... & Zupan, B. (2013). Orange: data mining toolbox in Python. the Journal of machine Learning research, 14(1), 2349-2353.

Dervenis, C., Fitsilis, P., & Iatrellis, O. (2022). A review of research on teacher competencies in higher education. Quality assurance in education, 30(2), pp. 199-220.

Dervenis, C., Fitsilis, P., Iatrellis, O., & Koustelios, A. (2024). Assessing Teacher Competencies in Higher Education: A Sentiment Analysis of Student Feedback. International Journal of Information and Education Technology, 14(4).

Dervenis, C., Kyriatzis, V., Stoufis, S., & Fitsilis, P. (2022, September). Predicting Students' Performance Using Machine Learning Algorithms. In Proceedings of the 6th International Conference on Algorithms, Computing and Systems, pp. 1-7.

European Association for Quality Assurance in Higher Education (2009). Standards and Guidelines for Quality Assurance in the European Higher Education Area.

European Centre for the Development of Vocational Training (2008). Terminology of European education and training policy: a selection of 100 key terms. Office for Official Publ. of the Europ. Communities.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), pp. 82-89.

Feng, F., Yang, Y., Cer, D., Arivazhagan, N., & Wang, W. (2020). Language-agnostic BERT sentence embedding. arXiv preprint arXiv:2007.01852.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. Teachers college record, 107(1), pp. 186-213.

Ferragina, P., & Manzini, G. (2000, November). Opportunistic data structures with applications. In Proceedings 41st annual symposium on foundations of computer science, IEEE, pp. 390-398.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Ferrari, A., & Punie, Y. (2013, June). DIGCOMP: A framework for developing and understanding digital competence in Europe.

Fitsilis, P. (2024). Navigating the Skills Revolution: The Essential Role of Competence Frameworks. Qeios.

Fitsilis, P., Damasiotis, V., Dervenis, C., Kyriatzis, V., Tsoutsa, P. (2024). Effective Data Stewardship in Higher Education: Skills, Competences, and the Emerging Role of Open Data Stewards. Manuscript under review in Journal of Changes in Education.

Flodén, J. (2017). The impact of student feedback on teaching in higher education. Assessment & Evaluation in Higher Education, 42(7), pp. 1054-1068.

Garg, K. (2020). Sentiment analysis of Indian PM's "Mann Ki Baat". International Journal of Information Technology, 12(1), pp. 37-48.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. arXiv preprint arXiv:1802.06893.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. Handbook of qualitative research, 2, 105, pp. 163-194.

Hammerstrøm, K., Wade, A., Hanz, K., Jørgensen, A.K. and Klint, A. (2009). Searching for studies: Information retrieval methods group policy brief, available at: https://doi.org/10.4073/pb.2009.1.

Hardeniya, T., & Borikar, D. A. (2016). Dictionary based approach to sentiment analysis-a review. International Journal of Advanced Engineering, Management and Science, 2(5), 239438.

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of educational research, 77(1), pp. 81-112.

Heron, J., & Reason, P. (1997). A participatory inquiry paradigm. Qualitative inquiry, 3(3), pp. 274-294.

Heshusius, L., & Ballard, K. (Eds.). (1996). From positivism to interpretivism and beyond: Tales of transformation in educational and social research (the mind-body connection). Teachers College Press.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. Educational research review, 12, pp. 45-58.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Ho, A. S. (2000). A conceptual change approach to staff development: A model for programme design. International Journal for Academic Development, 5(1), pp. 30-41.

Hoffmann, F. and Oreopoulos, P. (2009). Professor qualities and student achievement. Review of Economics and Statistics, Vol. 91 No. 1, pp. 83–92.

Hoidn, S. and Kärkkäinen, K. (2014). Promoting Skills for Innovation in Higher Education: A Literature Review on the Effectiveness of Problem-based Learning and of Teaching Behaviours. OECD Education Working Papers, No. 100, pp. 177–182.

Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media, Vol. 8, No. 1, pp. 216-225.

Iatrellis, O., Kameas, A. and Fitsilis, P. (2017). Academic advising systems: A systematic literature review of empirical evidence. Education Sciences, Vol. 7 No. 4, available at:https://doi.org/10.3390/educsci7040090.

Iqbal, A., Hussain, S., Mushtaq, S. and Javaid, Z. (2019). Effect of Teachers’ Competencies on Scholars’ Academic Achievement and Satisfaction. European Online Journal of Natural and Social Sciences, Vol. 8 No. 1, pp. 9–16.

Kabir, M., Kabir, M. M. J., Xu, S., & Badhon, B. (2021). An empirical research on sentiment analysis using machine learning approaches. International Journal of Computers and Applications, 43(10), pp. 1011-1019.

Kang, H., Yoo, S. J., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. Expert Systems with Applications, 39(5), pp. 6000-6010.

Kausar, S., Huahu, X. U., Ahmad, W., & Shabir, M. Y. (2019). A sentiment polarity categorization technique for online product reviews. IEEE Access, 8, pp. 3594-3605.

Khyani, D., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2021). An interpretation of lemmatization and stemming in natural language processing. Journal of University of Shanghai for Science and Technology, 22(10), pp. 350-357.

Knight, S., & Shum, S. B. (2017). Theory and learning analytics. Handbook of learning analytics, 1, pp. 17-22.

Kohavi, R. (1998). Glossary of terms. Machine learning, 30, pp. 271-274.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Krumsvik, R.J. (2014). Teacher educators' digital competence. Scandinavian Journal of Educational Research, Vol. 58 No. 3, pp. 269–280.

Kugel, P. (1993). How Professors Develop. Studies in Higher Education, Vol. 18 No. 3, pp. 315–328.

Kumar, A., & Jain, R. (2015). Sentiment analysis and feedback evaluation. In 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE), IEEE, pp. 433-436.

Lane, M., & Meth, D. (2021). Exploring impacts on students as givers of teaching feedback. Quality Assurance in Education, 29(2/3), pp. 225-237.

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.) (2017). Handbook of learning analytics. New York: SOLAR, Society for Learning Analytics and Research, pp. 23-31.

Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., ... & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. Heliyon, 8(10).

Lester, J., Klein, C., Rangwala, H., Johri, A., & Lester (2017). Learning analytics in higher education. John Wiley and Sons.

Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2018). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research, 5th ed., Sage, pp. 108–150.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), pp. 1-167.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., ... & Ge, B. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. Meta-Radiology, 100017.

Long, C.S., Ibrahim, Z. and Kowang, T.O. (2014). An analysis on the relationship between lecturers competencies and students satisfaction. International Education Studies, Vol. 7 No. 1, pp. 37–46.

López-Cámara, A.B., Eslava-Suanes, M.D., González-López, I., González, H.G. and De León-Huertas, C. (2018). Skills of university professor and their evaluation. ACM International Conference Proceeding Series, pp. 175–179.

Mahulae, A.V., Lumbanraja, P. and Siahaan, E. (2020). Effect of Professionalism and Competence of Teachers on Teacher Performance and Its Impact on Student Learning Outcomes at Harapan Mandiri College. International Journal of Research and Review, Vol. 7 No. 11, pp. 38–46.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Mandouit, L. (2018). Using student feedback to improve teaching. Educational action research, 26(5), pp. 755-769.

Marley, C., Faye, A. D., Hurst, E., Moeller, J., & Pinkerton, A. (2021). Moving beyond 'you said, we did': Extending an ethic of hospitality to the student feedback process. Online Postgraduate Education in a Postdigital World: Beyond Technology, pp. 1-19.

Marr, B. (2012). Key Performance Indicators (KPI): The 75 measures every manager needs to know. Pearson UK.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. The scholarship of teaching and learning in higher education: An evidence-based perspective, pp. 319-383.

Mayordomo, R. M., Espasa, A., Guasch, T., & Martínez-Melo, M. (2022). Perception of online feedback and its impact on cognitive and emotional engagement with feedback. Education and Information Technologies, 27(6), pp. 7947-7971.

McCulloch, W. S., & Pitts, W. (1990). A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biology, 52, pp. 99-115.

McDonald, J., Moskal, A. C. M., Goodchild, A., Stein, S., & Terry, S. (2020). Advancing text-analysis to tap into the student voice: a proof-of-concept study. Assessment & Evaluation in Higher Education.

Mejova, Y. (2009). Sentiment analysis: An overview. University of Iowa, Computer Science Department, pp. 5-17.

Metzler, J. and Woessmann, L. (2010). The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within- Teacher Within-Student Variation The Impact of Teacher Subject Knowledge on Student Achievement: Evidence from Within-Teacher Within-Student Variation Abstract. The Institute for the Study of Labor, No. 4999, pp. 36-48.

Misra, Amita, Brian Ecker, and Marilyn A. Walker. Measuring the similarity of sentential arguments in dialog. arXiv preprint arXiv:1709.01887 (2017).

Misuraca, M., Scepi, G., & Spano, M. (2021). Using Opinion Mining as an educational analytic: An integrated strategy for the analysis of students' feedback. Studies in Educational Evaluation, 68, 100979.

Mohamad Sham, N., & Mohamed, A. (2022). Climate change sentiment analysis using lexicon, machine learning and hybrid approaches. Sustainability, 14(8), pp. 4723-4732.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. BMJ (Online), Vol. 339 No. 7716, pp. 332–336.

Mourtzis, D., Angelopoulos, J., Siatras, V., & Panopoulos, N. (2021). A Methodology for the Assessment of Operator 4.0 Skills based on Sentiment Analysis and Augmented Reality. Procedia CIRP, 104, pp. 1668-1673.

Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: a review. Artificial Intelligence Review, 52(2), pp. 857-900.

Nasim, Z., Rajput, Q., & Haider, S. (2017, July). Sentiment analysis of student feedback using machine learning and lexicon based approaches. In 2017 international conference on research and innovation in information systems (ICRIIS), IEEE, pp. 1-6.

Navaneetan, M., Tharagesh, G., & Sabitha, A. S. (2024, April). Unveiling Sentiments: Analyzing Learner's Experience Using VADER and RoBERTa Models. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), IEEE, pp. 01-05.

Nguyen, L. A. T., & Habók, A. (2023). Tools for assessing teacher digital literacy: a review. Journal of Computers in Education, pp. 1-42.

Nguyen, L. A. T., & Habók, A. (2024). Tools for assessing teacher digital literacy: a review. Journal of Computers in Education, 11(1), pp. 305-346.

Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.

Nitin, G. I., Swapna, G., & Shankararaman, V. (2015, October). Analyzing educational comments for topics and sentiments: A text analytics approach. In 2015 IEEE frontiers in education conference (FIE), IEEE, pp. 1-9.

OECD, O. (2018). Future of education and skills 2030: OECD learning compass 2030. Research in Social and Administrative Pharmacy, Vol. 14 No. 1, pp. 6–17.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in human behavior, 31, pp. 527-541.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj, p. 372.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Pânişoară, I.O., Duţă, N., Pânişoară, G. and Tomoaica, E. (2014). Teacher Profile in Romania: Defining Features and Priorities of the Training Programs in Teaching Career. Procedia - Social and Behavioral Sciences, Vol. 140, pp. 396–400.

Park, S. and Oliver, J.S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. Research in Science Education, Vol. 38 No. 3, pp. 261–284.

Pham, S. T., & Sampson, P. M. (2022). The development of artificial intelligence in education: A review in context. Journal of Computer Assisted Learning, 38(5), pp. 1408-1421.

Potrac, P., Jones, R. L., & Nelson, L. (2014). Interpretivism. In Research methods in sports coaching. Routledge, pp. 31-41

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. Assessment & Evaluation in Higher Education, 33(2), pp. 143-154.

Pradhan, A., Senapati, M. R., & Sahu, P. K. (2023, January). Comparative analysis of lexicon-based emotion recognition of text. In Machine Learning, Image Processing, Network Security and Data Sciences: Select Proceedings of 3rd International Conference on MIND 2021, Singapore: Springer Nature Singapore, pp. 671-677.

PRISMA 2020. (n.d). http://prisma-statement.org/prismastatement/flowdiagram.aspx. Accesed Feb 2021.

Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. Social Network Analysis and Mining, 13(1), pp. 31.

Raman, R., Pattnaik, D., Hughes, L., & Nedungadi, P. (2024). Unveiling the dynamics of AI applications: A review of reviews using scientometrics and BERTopic modeling. Journal of Innovation & Knowledge, 9(3), 100517.

Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using educational data mining techniques to increase the prediction accuracy of student academic performance. Information and Learning Sciences, 120(7/8), pp. 451-467.

Ravenda, F., Bahrainian, S. A., Raballo, A., Mira, A., & Crestani, F. (2024). A self-supervised seed-driven approach to topic modeling and clustering. Journal of Intelligent Information Systems, pp. 1-21.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Redecker, C. (2017), European Framework for the Digital Competence: DigComEdu, available at: https://moodle.ktu.edu/pluginfile.php/428841/mod_resource/content/1/pdf_digcomedu_a4_final.pdf.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Rhode, J. (2017). Understanding faculty use of the learning management system. Online Learning, 21(3), pp. 68-86.

Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. Assessment & evaluation in higher education, 30(4), pp. 387-415.

Roebuck, K. (2012). Sentiment analysis: High-impact strategies-What you need to know: Definitions, adoptions, impact, benefits, maturity, vendors. Emereo Publishing.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International journal of artificial intelligence in education, 26, pp. 582-599.

Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data mining and knowledge discovery, 3(1), pp. 12-27.

Room, C. (2019). Confusion Matrix. Mach. Learn, 6, pp. 27-42.

Roumeliotis, K. I., & Tselikas, N. D. (2023). Chatgpt and open-ai models: A preliminary review. Future Internet, 15(6), pp. 192-216.

Rybinski, K. (2023). Exploring the influence of student emotions and professor behaviour on course ratings: a quantitative analysis. Quality Assurance in Education, (ahead-of-print).

Salazar, C., Montoya-Múnera, E., & Aguilar, J. (2022). Sentiment analysis in learning resources. Journal of Computers in Education, pp. 1-26.

Salazar, C., Montoya-Múnera, E., & Aguilar, J. (2023). Sentiment analysis in learning resources. Journal of Computers in Education, 10(4), pp. 637-662.

Sanders, W.L., Rivers, J.C. and Hall, M. (1996). Cumulative and Residual Effects of Teachers on Future Student Academic Achievement Summary of Findings.

Sangeetha, K., & Prabha, D. (2021). RETRACTED ARTICLE: Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. Journal of Ambient Intelligence and Humanized Computing, 12(3), pp. 4117-4126.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Saraswat, M., Chakraverty, S., & Kala, A. (2020). Analyzing emotion based movie recommender system using fuzzy emotion features. International Journal of Information Technology, 12, pp. 467-472.

Saunders, M. (2014). Research Methods for Business Students, 6th edn.

Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. English Language Teaching, 5(9), pp. 9-16.

Selvi, K. (2010). Teachers' Competencies. Cultura International Journal of Philosophy of Culture and Axiology, Vol. 7 No. 1, pp. 167–175.

Sen, T., Ali, M. R., Hoque, M. E., Epstein, R., & Duberstein, P. (2017, October). Modeling doctor-patient communication with affective text analysis. In 2017 seventh international conference on affective computing and intelligent interaction (ACII), IEEE, pp. 170-177.

Seo, J., Lee, S., Liu, L., & Choi, W. (2022). TA-SBERT: Token attention sentence-BERT for improving sentence representation. IEEE Access, 10, pp. 39119-39128.

Shafana, A. R. F., & Safnas, S. M. (2022). Does technology assist to continue learning during pandemic? A sentiment analysis and topic modeling on online learning in south asian region. Social Network Analysis and Mining, 12(1), pp. 65-79.

Shaik, A. B., & Srinivasan, S. (2019). A brief survey on random forest ensembles in classification model. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2, Springer Singapore, pp. 253-260.

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. Natural Language Processing Journal, 2, p. 100003.

Sharma, R., Nigam, S., & Jain, R. (2014). Opinion mining in Hindi language: a survey. arXiv preprint arXiv:1404.4935.

Singer, F.M. and Moscovici, H. (2008). Teaching and learning cycles in a constructivist approach to instruction. Teaching and Teacher Education, Vol. 24 No. 6, pp. 1613–1634.

Slater, S., Joksimović, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. Journal of Educational and Behavioral Statistics, 42(1), pp. 85-106.

Song, Y., Wang, J., Liang, Z., Liu, Z., & Jiang, T. (2020). Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference. arXiv preprint arXiv:2002.04815.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Srinivasa, K. G., & Kurni, M. (2021). A beginner's guide to learning analytics. Berlin/Heidelberg, Germany: Springer.

Srivastava, R., Bharti, P. K., & Verma, P. (2022). Comparative analysis of Lexicon and machine learning approach for sentiment analysis. International Journal of Advanced Computer Science and Applications, 13(3), pp. 71-77.

Su, F., Zou, D., Wang, L., & Kohnke, L. (2024). Student engagement and teaching presence in blended learning and emergency remote teaching. Journal of Computers in Education, 11(2), pp. 445-470.

Tangirala, S. (2020). Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm. International Journal of Advanced Computer Science and Applications, 11(2), pp. 612-619.

Tashakkori, A. (2010). SAGE handbook of mixed methods in social & behavioral research. Sage.

Theall, M. and Franklin, J. (2001). Looking for Bias in All the Wrong Places: A Search for Truth or a Witch Hunt in Student Ratings of Instruction? New Directions for Institutional Research, Vol. 2001 No. 109, pp. 45–56.

Tigelaar, D.E.H., Dolmans, D.H.J.M., Wolfhagen, I.H.A.P. and Van Der Vleuten, C.P.M. (2004). The development and validation of a framework for teaching competencies in higher education. Higher Education, Vol. 48 No. 2, pp. 253–268.

Tripathi, P. (2015). Competence Based Management in Academics through Data Mining Approach. Computer Engineering and Intelligent Systems, Vol. 6 No. 2, pp. 50–57.

Tsinidou, M., Gerogiannis, V., & Fitsilis, P. (2010). Evaluation of the factors that determine quality in higher education: an empirical study. Quality assurance in Education, 18(3), pp. 227-244.

E.U. (2012). Investing in skills for better socio-economic outcomes. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. –Strasbourg: European Commission.

Vogt, F. and Rogalla, M. (2009). Developing Adaptive Teaching Competency through coaching. Teaching and Teacher Education, Elsevier Ltd, Vol. 25 No. 8, pp. 1051–1060.

Voudoukis, N., & Pagiatakis, G. (2022). Massive open online courses (MOOCs): practices, trends, and challenges for the higher education. European Journal of Education and Pedagogy, 3(3), pp. 288-295.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Wang, W., & Wu, J. (2011, July). Notice of Retraction: Emotion recognition based on CSO&SVM in e-learning. In 2011 Seventh International Conference on Natural Computation, Vol. 1, IEEE, pp. 559-570.

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), pp. 5731-5780.

Welch, C., & Mihalcea, R. (2016, December). Targeted sentiment to understand student comments. In Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers, pp. 2471-2481.

Wen, M., Yang, D., & Rose, C. (2014, July). Sentiment Analysis in MOOC Discussion Forums: What does it tell us?. In Educational data mining 2014.

Willig, C. (2019). What can qualitative psychology contribute to psychological knowledge?. Psychological methods, 24(6), pp. 796-809.

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational linguistics, 35(3), pp. 399-433.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Data, M. (2005, June). Practical machine learning tools and techniques. In Data mining. Amsterdam, The Netherlands: Elsevier, Vol. 2, No. 4, pp. 403-413.

Wong, B. T. M., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). Journal of Computers in Education, 7(1), pp. 7-28.

Wook, M., Vasanthan, S., Ramli, S., Razali, N. A. M., Hasbullah, N. A., & Zainudin, N. M. (2020). Exploring students' feedback in online assessment system using opinion mining technique. International Journal of Information and Education Technology, 10(9), pp. 664-668.

Xing, W., Lee, H. S., & Shibani, A. (2020). Identifying patterns in students' scientific argumentation: content analysis through text mining using Latent Dirichlet Allocation. Educational Technology Research and Development, 68(5), pp. 2185-2214.

Xun, L. S., Gottipati, S., & Shankararaman, V. (2015, June). Text-mining approach for verifying alignment of information systems curriculum with industry skills. In 2015 International Conference on Information Technology Based Higher Education and Training (ITHET), IEEE, pp. 1-6.

Zeng, L., Tan, Z., Xia, L., Xiang, Y. A., & Ke, Y. (2023). Behavior analytics, sentiment analysis, and topic detection of Danmaku from online electronics courses on Bilibili. International Journal of Information and Education Technology, 13(2).

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for knn classification. ACM Transactions on Intelligent Systems and Technology (TIST), 8(3), pp. 1-19.

Zhang, Y., & Wallace, B. (2015). A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1, pp. 1420-1430.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Appendix A – Publications

1. **Dervenis, C., Fitsilis, P., & Iatrellis, O. (2022). A review of research on teacher competencies in higher education. *Quality assurance in education*, *30*(2), 199-220.**

   Publisher: Emerald Publishing

   Status: Published

   Abstract: The purpose of this paper is to thoroughly assemble, analyze and synthesize previous research in order to investigate and identify teaching staff competencies derived from the roles and tasks attributed to university professors. In this literature review, we looked at both the conceptual framework exploring the educational concepts and the learning theories focusing on teaching staff roles and competencies in higher education. Thirty-nine (39) scientific papers were studied in detail from a total of 102 results, which were eligible based on the PRISMA Statement. A multi-dimensional approach to teacher competencies in higher education was proposed, which consists of 6 main dimensions with their respective characteristics. Thirty-three (33) discrete teaching staff competencies were identified and distributed in the aforementioned dimensions. The research revealed that specific competencies, such as the digital competence of teachers, which have lately become of high importance worldwide due to the COVID-19 pandemic implications, surprisingly, until recently they were considered secondary in the educational process.

2. **Dervenis, C., Kyriatzis, V., Stoufis, S., & Fitsilis, P. (2022, September). Predicting students' performance using machine learning algorithms. In *Proceedings of the 6th International Conference on Algorithms, Computing and Systems* (pp. 1-7).**

   Publisher: ACM

   Status: Published

   Abstract: Learning analytics (LA), defined as "the measurement, collection, analysis and reporting of data about learners and their contexts for the purposes of understanding and optimizing learning and the environments in which it occurs", is a research topic that has gained significant importance and visibility among researchers during the past few decades. It is a research domain where the use of modern machine-learning (ML) algorithms and big

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

data management provide timely and actionable information that can transform the overall learning experience for both students and educational institutions. In this paper we use ML algorithms in order to predict the performance of students, taking into account both past semester grades and socioeconomic factors. We run two models; a 2-class one predicting a "pass" or "fail" result and then we expanded this to a 5-class model, where we predict in which grading group the student will fall in the next semester. The results acquired indicate that it is possible to accurately predict the student's performance in both cases, with the 2-class model performing better than the 5-class one, which opts in providing more fine grain results.

3. **Dervenis, C., Kanakis, G., & Fitsilis, P. (2024). Sentiment analysis of student feedback: A comparative study employing lexicon and machine learning techniques. *Studies in Educational Evaluation*, *83*, 101406.**

Publisher: Elsevier

Status: Published

Abstract: One of the most pressing concerns in contemporary education examines the integration of big data and artificial intelligence methodologies to enhance the educational learning outcome. Towards that purpose, it is imperative to leverage the unstructured data originating from student feedback, particularly in the form of comments to open-ended questions aiming to extract emotions and opinions conveyed within their messages. Our research goal is to ascertain the most efficient approach to tackle this difficult task by conducting a comparison of SA methods, including Machine Learning (ML) and Lexicon-based models. Both lexicon-based and ML approaches were implemented using an open source data mining platform while also utilizing student comments submitted at the end of academic semesters. Our study reveals a promising approach that effectively addresses the issue at hand, particularly within the domain of educational data. Additionally, it emphasizes the key aspects that led to the selection of this approach effectively highlighting the weaknesses and strengths inherent in each method.

4. **Dervenis, C., Fitsilis, P., Iatrellis, O., & Koustelios, A. (2024). Assessing Teacher Competencies in Higher Education: A Sentiment Analysis of Student Feedback. *International Journal of Information and Education Technology*, *14*(4).**

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Publisher: IJIET

Status: Published

Abstract: One of the most important issues concerning education nowadays is that of mapping the quality of teaching, and teacher competencies. Simultaneously, the need to exploit the huge amount of data derived from student feedback, and in particular the comments on open-ended questions, constitutes a huge challenge for both universities and researchers. In this study, we use SA methods to measure teachers' communication competencies. Utilizing the data of over 700 feedback comments from students of our university, we assessed specific competencies through the sentiment intensity that these comments contained. The model designed for the SA, as well as the entire experimental phase, were implemented using an open source data mining and visualization platform. Our research revealed that certain competencies are highlighted by the nature of the course while others do not depend on it. In addition, findings indicate both a homogeneous and convergent view among students, thus strengthening the validity of the students' opinion and their valuable contribution to the mapping of teaching quality in general with the ultimate goal of enhancing education.

5. **Dervenis, C., Fitsilis, (2024). Chat GPT vs. Typical Data Mining Process: A Comparative Case Study on Teaching Competencies via Sentiment Analysis. *European Education.***

Publisher: Taylor & Francis

Status: Under review

Abstract: Nowadays, a critical matter in the field of education revolves around assessing the quality of teaching and teacher competencies. Universities and researchers, however, face a significant challenge in effectively utilizing the vast amount of data generated from student feedback, particularly in the form of open-ended comments written in natural language. Simultaneously, the importance of chatbot AI usage in our daily lives has been increasingly recognized in recent times due to its ability to provide convenient and efficient solutions in various domains. In this paper, we utilize the data of over 700 feedback comments from students of UTH and assess specific teacher competencies through the sentiment intensity that these comments contained. This paper aims to perform and compare two different sentiment analyses approaches; a typical data mining procedure and a generative AI tool, the ChatGPT. Through our research we intend to contribute to introducing readers to chatbot AI usage in education, presenting a path to computing the sentiment polarity of students'

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

comments. A discussion about our research findings offers valuable insights into the potential of our approach regarding assessing teacher competencies in the teaching domain.

6. **Dervenis, C., Fitsilis, P., Iatrellis, O., & Koustelios, A. (2025). Analysis of Students' Perspectives of Teaching Quality Through Topic Modeling: A Comparative Study. *Quality assurance in education*.**

Publisher: Emerald Publishing

Status: Submitted

Abstract: One of the most pressing concerns in contemporary education examines the integration of big data and artificial intelligence methodologies to enhance the educational learning outcome. Towards that purpose, it is imperative to leverage the unstructured data originating from student feedback, particularly in the form of comments to open-ended questions aiming to extract emotions and opinions conveyed within their messages. Our research goal is to ascertain the most efficient approach to tackle this difficult task by conducting a comparison of SA methods, including Machine Learning (ML) and lexicon-based models. Both lexicon-based and ML approaches were implemented using an open source data mining platform while also utilizing student comments submitted at the end of academic semesters. Our study reveals a promising approach that effectively addresses the issue at hand, particularly within the domain of educational data. Additionally, it emphasizes the key aspects that led to the selection of this approach effectively highlighting the weaknesses and strengths inherent in each method.

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Appendix B – Generation of a 3x3 Confusion Matrix using Python Code

```
# Script to plot a 3x3 confusion matrix using the seaborn library
import seaborn as sn
import pandas as pd
import matplotlib.pyplot as plt

# Copy the incoming data into a new DataFrame
my_data = in_data.copy()
df = pd.DataFrame(my_data)

# Initialize variables for the classification results (confusion matrix cell values)
# POS = Positive, NEU = Neutral, NEG = Negative
# The variables represent the count of each classification outcome
POS_x_POS = 0  # True Positive: Predicted POS and Actual POS
POS_x_NEU = 0  # False Positive to Neutral: Predicted NEU but Actual POS
POS_x_NEG = 0  # False Positive to Negative: Predicted NEG but Actual POS
NEU_x_NEU = 0  # True Neutral: Predicted NEU and Actual NEU
NEU_x_POS = 0  # False Neutral to Positive: Predicted POS but Actual NEU
NEU_x_NEG = 0  # False Neutral to Negative: Predicted NEG but Actual NEU
NEG_x_NEU = 0  # False Negative to Neutral: Predicted NEU but Actual NEG
NEG_x_POS = 0  # False Negative to Positive: Predicted POS but Actual NEG
NEG_x_NEG = 0  # True Negative: Predicted NEG and Actual NEG

# Traverse the DataFrame row by row and calculate confusion matrix values
for index, row in df.iterrows():
    # Check the actual and predicted values and update the corresponding cell
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

```
    if row[0] == 2:  # Actual is POS
        if row[1] == 2:  # Predicted is POS
            POS_x_POS += 1
        elif row[1] == 1:  # Predicted is NEU
            POS_x_NEU += 1
        else:  # Predicted is NEG
            POS_x_NEG += 1
    elif row[0] == 1:  # Actual is NEU
        if row[1] == 1:  # Predicted is NEU
            NEU_x_NEU += 1
        elif row[1] == 2:  # Predicted is POS
            NEU_x_POS += 1
        else:  # Predicted is NEG
            NEU_x_NEG += 1
    elif row[0] == 0:  # Actual is NEG
        if row[1] == 0:  # Predicted is NEG
            NEG_x_NEG += 1
        elif row[1] == 2:  # Predicted is POS
            NEG_x_POS += 1
        else:  # Predicted is NEU
            NEG_x_NEU += 1

# Create a 2D array representing the confusion matrix
array = [
    [POS_x_POS, POS_x_NEG, POS_x_NEU],  # Row for actual POS
    [NEG_x_POS, NEG_x_NEG, NEG_x_NEU],  # Row for actual NEG
    [NEU_x_POS, NEU_x_NEG, NEU_x_NEU]   # Row for actual NEU
]
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

```
# Convert the confusion matrix array into a DataFrame for visualization
df_cm = pd.DataFrame(array, index=["POS", "NEG", "NEU"],  # Actual values
                     columns=["POS", "NEG", "NEU"])       # Predicted values

# Plot the confusion matrix as a heatmap using seaborn
sn.heatmap(df_cm, annot=True, cmap="crest", linewidth=.9, fmt=".0f")

# Initialize variables for performance metrics
Accuracy_ALL = 1.0
Precision_ALL = 1.0
Recall_ALL = 1.0

Accuracy_POS = 1.0
Precision_POS = 1.0
Recall_POS = 1.0
F1_POS = 1.0

Accuracy_NEG = 1.0
Precision_NEG = 1.0
Recall_NEG = 1.0
F1_NEG = 1.0

Accuracy_NEU = 1.0
Precision_NEU = 1.0
Recall_NEU = 1.0
F1_NEU = 1.0
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

```
# Calculate metrics for the Positive class (POS)
if (POS_x_POS + POS_x_NEG + POS_x_NEU) != 0:
    Accuracy_POS = POS_x_POS / (POS_x_POS + POS_x_NEG + POS_x_NEU)
    Precision_POS = POS_x_POS / (POS_x_POS + POS_x_NEG + POS_x_NEU)
    Recall_POS = POS_x_POS / (POS_x_POS + NEG_x_POS + NEU_x_POS)
    F1_POS = (2 * Precision_POS * Recall_POS) / (Precision_POS + Recall_POS)

# Calculate metrics for the Negative class (NEG)
if (NEG_x_NEG + NEG_x_POS + NEG_x_NEU) != 0:
    Precision_NEG = NEG_x_NEG / (NEG_x_NEG + NEG_x_POS + NEG_x_NEU)
    Accuracy_NEG = NEG_x_NEG / (NEG_x_NEG + NEG_x_POS + NEG_x_NEU)
    Recall_NEG = NEG_x_NEG / (NEG_x_NEG + POS_x_NEG + NEU_x_NEG)
    print(Precision_NEG)

if (Precision_NEG + Recall_NEG) != 0:
    F1_NEG = (2 * Precision_NEG * Recall_NEG) / (Precision_NEG + Recall_NEG)

# Calculate metrics for the Neutral class (NEU)
if (NEU_x_NEU + NEU_x_POS + NEU_x_NEG) != 0:
    Precision_NEU = NEU_x_NEU / (NEU_x_NEU + NEU_x_POS + NEU_x_NEG)
    Accuracy_NEU = NEU_x_NEU / (NEU_x_NEU + NEU_x_POS + NEU_x_NEG)
    Recall_NEU = NEU_x_NEU / (NEU_x_NEU + NEG_x_NEU + POS_x_NEU)
    F1_NEU = (2 * Precision_NEU * Recall_NEU) / (Precision_NEU + Recall_NEU)

# Calculate overall metrics
Accuracy_ALL = (NEU_x_NEU + NEG_x_NEG + POS_x_POS) / \
            (POS_x_POS + POS_x_NEU + POS_x_NEG + NEU_x_NEU +
             NEU_x_POS + NEU_x_NEG + NEG_x_NEU + NEG_x_POS + NEG_x_NEG)
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

```
Recall_ALL = (POS_x_POS + NEU_x_NEU + NEG_x_NEG) / \
          (POS_x_POS + NEU_x_NEU + NEG_x_NEG + NEG_x_NEU + POS_x_NEU +
           NEG_x_POS + NEU_x_POS + NEG_x_NEU + POS_x_NEU)

# Generate a result string containing the metrics
res = "Overall Accuracy: " + str('%.4f' % Accuracy_ALL) \
    + ", Overall Recall=" + str('%.4f' % Recall_ALL) + "\n" \
    + "POS: Accuracy=" + str('%.4f' % Accuracy_POS) \
    + ", Recall=" + str('%.4f' % Recall_POS) \
    + ", F-1=" + str('%.4f' % F1_POS) + "\n" \
    + "NEG: Accuracy=" + str('%.4f' % Accuracy_NEG) \
    + ", Recall=" + str('%.4f' % Recall_NEG) \
    + ", F-1=" + str('%.4f' % F1_NEG) + "\n" \
    + "NEU: Accuracy=" + str('%.4f' % Accuracy_NEU) \
    + ", Recall=" + str('%.4f' % Recall_NEU) \
    + ", F-1=" + str('%.4f' % F1_NEU)

# Add the result string as text to the plot
plt.text(x=0, y=0, s=res)

# Display the plot
plt.show()
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Appendix C – BERTopic Implementation Using Python in Google Colab

```
pip install bertopic[flair,gensim,spacy,use]

pip install advertools

pip install pandas openpyxl

# Importing libraries

from bertopic import BERTopic

import pandas as pd

import advertools as adv

from transformers import AutoModel, AutoTokenizer

from sklearn.feature_extraction.text import CountVectorizer

from bertopic.vectorizers import ClassTfidfTransformer


# Reading the Excel file

df = pd.read_excel('/content/drive/MyDrive/07 PhD/0_RESEARCH/Topic Modeling Student Perspective for Teaching excellence/Real_Comments.xlsx')


# Reading my data from the 'data_column' in the Excel file provided to me.

docs = df['comments'].tolist()


# Loading multilingual BERT model

multilingual_model = AutoModel.from_pretrained("bert-base-multilingual-cased")

tokenizer = AutoTokenizer.from_pretrained("bert-base-multilingual-cased")


# Creating BERTopic model

greek_stopwords = sorted(adv.stopwords['greek'])

vectorizer_model = CountVectorizer(stop_words=greek_stopwords)
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

```
# Guiding the analysis model towards specific topics.
# I am using specific words (seed words) that act as initial indicators to
# direct the model to identify topics that are more relevant to what we are
# interested in.

ctfidf_model = ClassTfidfTransformer(
Πολύ λίγες έχεις ως seeds!
    seed_words=["teacher", "late", "difficult", "communicative", "capable", "digital", "posting"],
    seed_multiplier=10
)

# Creating BERTopic model with a specific number of topics
topic_model = BERTopic(embedding_model=multilingual_model,
vectorizer_model=vectorizer_model, language="greek", ctfidf_model=ctfidf_model)

# Fitting the model to the documents
topics, probs = topic_model.fit_transform(docs)
topic_model.get_topic_info()

from gensim.models import CoherenceModel
from gensim.corpora.dictionary import Dictionary

# Extracting the topics in the form of a list of words
topics = topic_model.get_topics()
topic_words = [[word for word, _ in topics[topic]] for topic in topics]

# Converting the documents to tokens for Gensim
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

```
tokenized_docs = [doc.split() for doc in docs]  # Splitting the comments into words


# Creating a dictionary and corpus for Gensim
dictionary = Dictionary(tokenized_docs)
corpus = [dictionary.doc2bow(doc) for doc in tokenized_docs]


# Calculating the Coherence Score using c_v
coherence_model = CoherenceModel(topics=topic_words, texts=tokenized_docs, dictionary=dictionary, coherence='c_v')
coherence_score = coherence_model.get_coherence()


print(f"Topic Coherence Score: {coherence_score}")
```


Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250

Appendix D – Turnitin report

ORIGINALITY REPORT

3%	3%	0%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	2%
2	www.researchgate.net Internet Source	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

Institutional Repository - Library & Information Centre - University of Thessaly
11/09/2026 19:56:01 EEST - 216.73.216.250