

DEFINING PREDICTORS OF STUDENT SATISFACTION BASED ON
STUDENT EVALUATION OF TEACHING USING DECISION TREE
ANALYSIS

RABİA KÖSTEN ÇUBUKÇU

A THESIS SUBMITTED
FOR
THE DEGREE OF MASTER OF ARTS
IN
TEACHING ENGLISH AS A FOREIGN LANGUAGE

İHSAN DOĞRAMACI BILKENT UNIVERSITY
ANKARA

AUGUST 2024



Dedicated to my beloved husband...

Defining Predictors of Student Satisfaction Based on Student Evaluation of Teaching

Using Decision Tree Analysis

The Graduate School of Education

of

İhsan Doğramacı Bilkent University

by

Rabia Kösten Çubukçu

In Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in

Teaching English as a Foreign Language

The Department of English Language Teaching

Ankara
August 2024

İHSAN DOĞRAMACI BILKENT UNIVERSITY
GRADUATE SCHOOL OF EDUCATION

Defining Predictors of Student Satisfaction Based on Student Evaluation of Teaching
Using Decision Tree Analysis

Rabia Kösten Çubukçu

July 2024

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Teaching English as a Foreign Language.

Asst. Prof. Dr. İlker Kalender (Advisor)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Teaching English as a Foreign Language.

Assoc. Prof. Dr. Jennie Farber Lane (Examining Committee Member)

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Teaching English as a Foreign Language.

Prof. Dr. Çiler Hatipoğlu, Middle East Technical University (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Alipaşa Ayas (Director)

ABSTRACT

Defining Predictors of Student Satisfaction Based on Student Evaluation of Teaching
Using Decision Tree Analysis

Rabia Kösten Çubukçu

MA in Teaching English as a Foreign Language

Advisor: Asst. Prof. Dr. İlker Kalender

August 2024

Student evaluation of teaching is a prevalent method to assess instructional quality and student satisfaction in higher education all over the world. However, there is an ongoing debate as to which characteristics of instructors make them effective. This study aimed to discover which instructional characteristics can predict student satisfaction levels. To this end, a CHAID analysis, a form of decision tree analysis, was conducted on SPSS to reveal the relationships between instructional characteristics of instructors and student satisfaction level measured by a SET form. The study was conducted at an English language preparatory school of a non-profit private university in Türkiye. 4281 forms including 23 Likert-type questions were analyzed. The findings show effectiveness and being supportive are the most significant predictors of student satisfaction. Following them, enabling students to evaluate different perspectives, encouragement to share views, feedback, and positivity are highly valued by students. Less significant predictors are found to be variety of activities, asking questions to encourage students to express opinions, subject knowledge, guidance, encouraging active participation, preparedness, and recommending publications in English. All in all, 13 of 23 items were significant predictors of student satisfaction.

Keywords: Student evaluation of teaching (SET), decision tree analysis, CHAID, instructional effectiveness

ÖZET

Karar Ağacı Analizi Kullanarak Üniversite Öğrencilerinin Öğretim Elemanı

Değerlendirme Formlarındaki Memnuniyetinin Yordayıcılarını Belirlemek

Rabia Kösten Çubukçu

Yabancı Dil Olarak İngilizce Öğretimi Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi İlker Kalender

Ağustos 2024

Öğretim elemanı değerlendirme formları, öğretim kalitesini ve öğrenci memnuniyetini değerlendirmek için dünya çapında üniversitelerde kullanılan yöntemlerden en yaygınıdır. Öğretim elemanlarının hangi özelliklerinin öğrencilerin tatmini açısından önemli olduğu ise farklı çalışmalara konu olmuştur. Bu çalışma ile, öğretim elemanlarının hangi özelliklerinin öğrencilerin memnuniyet seviyelerini yordayabileceği araştırılmıştır. Bu amaçla, SPSS’de bir karar ağacı analizi formu olan CHAID kullanılmıştır. Çalışma, Türkiye’de bir vakıf üniversitesinin İngilizce hazırlık bölümünde gerçekleşmiştir. Toplamda 23 Likert tipi soru içeren 4281 form incelenmiştir. Bulgular, öğretim elemanlarının etkili ders işlemleri ile destekleyici davranışlar sergilemelerinin öğrenci memnuniyetinin en önemli yordayıcıları olduğunu göstermektedir. Sonraki önemli yordayıcılar, öğrencilerin farklı düşünceleri değerlendirmelerini ve farklı konularla ilgili yorum yapmalarını sağlamak, geri bildirim vermek ve pozitifliktir. Nispeten daha az önemli yordayıcılar, ders içi etkinliklerin çeşitliliği, derste geçen konularla ilgili öğrencilerin görüş belirtmesini gerektiren sorular sormak, alana hakimlik, yararlı çalışmalar tavsiye etmek, aktif katılımı teşvik etmek, planlı olmak ve İngilizce yayınlar tavsiye etmektir. Anketteki 23 sorunun 13’ü anlamlı yordayıcılar olarak saptanmıştır.

Anahtar kelimeler: öğretim elemanı değerlendirme anketi, karar ağacı analizi,

CHAID, öğrenci memnuniyeti, öğretimin etkililiği

ACKNOWLEDGEMENTS

First of all, I am profoundly thankful to my advisor, Asst. Prof. Dr. İlker Kalender, for his invaluable guidance, unwavering support, and insightful feedback. His expertise and positivity made this journey less daunting for me.

I would also like to thank my thesis committee members, Assoc. Prof. Dr. Jennie Farber Lane and Prof. Dr. Çiler Hatipoğlu for their valuable guidance and feedback.

My heartfelt thanks go to my beloved husband, Ali İhsan, whose constant emotional support and technical support in Word and Excel programs made this journey more manageable and less stressful for me. His encouragement was a driving force that kept me going in very challenging times.

I would like to extend my gratitude to my parents, Ömrüye and Tevfik, and my brother Muzaffer, for all the unconditional love and support that they provided throughout my life. I feel lucky to have been born in such a great family.

I am also deeply grateful to Nurgül, Mustafa, and Büşra for their kindness, support, and encouragement throughout this journey.

My sincere thanks go to my dear friend Şeyma, without whose encouragement and support I would not have probably started doing this master's degree.

Finally, I want to thank my dear budgie, Çapkın, for always standing on my shoulder when I was studying. His presence made my studying more enjoyable.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZET.....	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES	xi
LIST OF FIGURES	xii
CHAPTER 1: INTRODUCTION	1
Introduction	1
Background	3
Problem	11
Purpose.....	12
Research Questions	13
Significance.....	14
CHAPTER 2: REVIEW OF RELATED LITERATURE	15
Introduction	15
Conceptual Framework of the Study.....	15
History of SET	17
Studies Related to the Use of SET	19
Factors Affecting Students' Satisfaction Levels with the Instructors.....	23
Instructors' Age	23

Race & Nationality	24
Gender	25
Academic Performance	25
Course Type	26
Educational Level of Students and Time Spent at School	27
The Concept of Effective Teaching in Higher Education	28
CHAPTER 3: METHOD	31
Introduction	31
Research Design.....	31
Context	31
Participants.....	34
Instrumentation	36
Method of Data Collection.....	39
Methods of Data Analysis.....	40
CHAPTER 4: RESULTS	45
Introduction	45
The Most Significant Predictors of Overall Teaching Performance Based on Whole Sample.....	49
The Most Significant Predictors of Overall Teaching Performance Depending on Course Type	64
The Most Significant Predictors of Overall Teaching Performance for MC Instructors	64

The Most Significant Predictors of Overall Teaching Performance for TW	
Instructors.....	70
The Most Significant Predictors of Overall Teaching Performance for LS	
Instructors.....	76
The Most Significant Predictors of Overall Teaching Performance Depending on	
Students' Proficiency Levels.....	85
The Most Significant Predictors of Overall Teaching Performance for AF Level	
Instructors.....	85
The Most Significant Predictors of Overall Teaching Performance for A Level	
Instructors.....	91
The Most Significant Predictors of Overall Teaching Performance for B Level	
Instructors.....	95
The Most Significant Predictors of Overall Teaching Performance for C Level	
Instructors.....	104
The Most Significant Predictors of Overall Teaching Performance Depending on	
Number of Trimesters Spent by Students in Preparatory School	112
The Most Significant Predictors of Overall Teaching Performance for the	
Instructors in the 1 st Trimester.....	112
The Most Significant Predictors of Overall Teaching Performance for the	
Instructors in the 2 nd Trimester.....	120
The Most Significant Predictors of Overall Teaching Performance for the	
Instructors in the 3 rd Trimester	128
Summary of Chapter 4	132
CHAPTER 5: DISCUSSION.....	134

Introduction	134
Overview of the Study	134
Discussion of Major Findings	135
Overall Findings from 11 Decision Trees	136
Being Effective	137
Being Supportive	139
Enabling Students to Evaluate Different Perspectives	140
Encouragement to Share Views	141
Feedback	142
Positivity	143
Variety of Activities & Asking Questions to Encourage Students to Express Opinions.....	143
Subject Knowledge, Guidance, Encouraging Active Participation, Preparedness, Recommending Publications	144
Discussion of the Findings for the Whole Sample	148
Discussion of the Findings for Different Course Types and Their Comparison with the Whole Sample	149
Discussion of the Findings for Different Proficiency Levels and Their Comparison with the Whole Sample	152
Discussion of the Findings for Different Trimesters and Their Comparison with the Whole Sample	156
Summary	159
Implications for Practice	160
Implications for Further Research.....	160

Limitations	161
REFERENCES.....	164
Appendix A	181



LIST OF TABLES

Table	Page
1 Numbers of Students per Level and per Semester.....	35
2 Questions of the SET Instrument.....	37
3 Descriptive Statistics of Survey Items.....	42
4 Questions of the SET Instrument and Their Shortened Versions.....	46
5 The Most Significant Predictors of Overall Teaching Quality Obtained from all the 11 Analyses According to the Levels of the Trees.....	136
6 Predictors Appearing in Each Level of the Decision Tree for the Whole Sample.....	148
7 The Comparison of Predictors for the Whole Sample, MC, TW, and LS Instructors.....	149
8 The Comparison of the Predictors for the Whole Sample, AF, A, B, and C Level Instructors.....	155
9 The Comparison of the Predictors for the Whole Sample, the 1st Trimester, the 2nd Trimester, and the 3rd Trimester Instructors.....	156

LIST OF FIGURES

Figure		Page
1	A Framework for Understanding the Relationship Between Teaching Quality and SET Scores.....	16
2	Sample Decision Tree Diagram.....	40
3	A Glimpse into How to Interpret Decision Trees.....	46
4	Outline of the Decision Tree Produced for the Whole Group.....	51
5	Section 1 of the Decision Tree for the Whole Sample.....	52
6	Section 2 of the Decision Tree for the Whole Sample.....	54
7	Section 3 of the Decision Tree for the Whole Sample.....	57
8	Section 4 of the Decision Tree for the Whole Sample.....	60
9	Section 5 of the Decision Tree for the Whole Sample.....	62
10	Outline of the Decision Tree for MC Instructors.....	65
11	Section 1 of the Decision Tree for MC Instructors.....	66
12	Section 2 of the Decision Tree for MC Instructors.....	68
13	Outline of the Decision Tree for TW Instructors.....	71
14	Section 1 of the Decision Tree for TW Instructors.....	72
15	Section 2 of the Decision Tree for TW Instructors.....	74
16	Outline of the Decision Tree Produced for LS Instructors.....	77
17	Section 1 of the Decision Tree for LS Instructors.....	78
18	Section 2 of the Decision Tree for LS Instructors.....	80
19	Section 3 of the Decision Tree for LS Instructors.....	82
20	Section 4 of the Decision Tree for LS Instructors.....	84
21	Outline of the Decision Tree Produced for AF Level Instructors.....	86

22	Section 1 of the Decision Tree for AF Level Instructors.....	87
23	Section 2 of the Decision Tree for AF Level Instructors.....	89
24	Outline of the Decision Tree Produced for A Level Instructors.....	92
25	Section 1 of the Decision Tree for A Level Instructors.....	93
26	Section 2 of the Decision Tree for A Level Instructors.....	94
27	Outline of the Decision Tree Produced for B Level Instructors.....	96
28	Section 1 of the Decision Tree for B Level Instructors.....	97
29	Section 2 of the Decision Tree for B Level Instructors.....	99
30	Section 3 of the Decision Tree for B Level Instructors.....	101
31	Section 4 of the Decision Tree for B Level Instructors.....	103
32	Outline of the Decision Tree Produced for C Level Instructors.....	105
33	Section 1 of the Decision Tree for C Level Instructors.....	106
34	Section 2 of the Decision Tree for C Level Instructors.....	108
35	Section 3 of the Decision Tree for C Level Instructors.....	109
36	Section 4 of the Decision Tree for C Level Instructors.....	111
37	Outline of the Decision Tree Produced for the Instructors in the 1 st Trimester.....	113
38	Section 1 of the Decision Tree for the Instructors in the 1 st Trimester.....	114
39	Section 2 of the Decision Tree for the Instructors in the 1 st Trimester.....	116
40	Section 3 of the Decision Tree for the Instructors in the 1 st Trimester.....	119
41	Outline of the Decision Tree Produced for the Instructors in the 2 nd Trimester.....	121

42	Section 1 of the Decision Tree for the Instructors in the 2 nd	
	Trimester.....	122
43	Section 2 of the Decision Tree for the Instructors in the 2 nd	
	Trimester.....	124
44	Section 3 of the Decision Tree for the Instructors in the 2 nd	
	Trimester.....	127
45	Outline of the Decision Tree Produced for the Instructors in the 3 rd	
	Trimester.....	128
46	Section 1 of the Decision Tree for the Instructors in the 3 rd	
	Trimester.....	130
47	Section 2 of the Decision Tree for the Instructors in the 3 rd	
	Trimester.....	131

CHAPTER 1: INTRODUCTION

Introduction

Higher education institutions have been in our lives for about a thousand years. For instance, the University of Bologna is claimed to be the oldest university in Europe with its foundation in 1088. As universities have been the driving force for development in numerous areas, it is expected that they will continue their existence in the future and function as the facilitator of development in societies (Semerena, 2021).

However, the needs and demands of human beings in every period show difference in higher education, and higher education institutions need to adapt to those changing needs. For example, in the past only privileged people used to be involved in higher education, and their main aim was to become scholars at universities. Daily jobs did not require university degrees. University professors at that time were seen as the ultimate source of knowledge, and they could not be questioned. However, in our modern world, university education is regarded as a necessity by many people with the prevalence of universities nearly in all big cities. This prevalence of universities has led to another issue in higher education, which is quality of education and quality of research.

Although higher education institutions have existed for about a thousand years, it has been only a hundred years since student evaluation of teaching (SET) forms were first used in the beginning of 1920s (Gelber, 2020). SET was introduced as a form of evaluating teaching quality at that time. Although it was initially intended for the personal use of university professors, SET was an indicator that

teaching effectiveness of the professors could be questioned and students' beliefs about teaching quality mattered as they were the ones who were directly affected by it. In time, SET forms were not only used by professors for self-reflection, but they were also used for different reasons such as promotion. Although this was a good step towards ensuring teaching quality in higher education, there was another problem. What was effective teaching?

One definition of effective teaching is that "A teacher is effective if he/she can accomplish the planned goals and assigned tasks in accordance with school goals" (Ko et al., 2013, p.5). Goe (2007) described teacher effectiveness as the improvement in student learning, which is measured through examinations. Although the definition of the concept of effective teaching might change depending on many factors such as students' purpose to get education, their major, or whether they are at a graduate or undergraduate level, there are still certain traits of instructors which make them effective in higher education. The questions which are asked in SET forms present operational definitions about these traits related with effectiveness.

Having good subject knowledge, having pedagogical knowledge, being professional, being able to understand students' needs, being approachable, being open-minded, and being enthusiastic are some of the traits that effective instructors are expected to have according to the literature (Brown & Atkins, 2002; Kurniati & Cahyono, 2018; Su & Wood, 2012). Therefore, many higher education institutions aim to evaluate those abovementioned traits of instructors which imply effective teaching, and SET forms used in different institutions show similarities. As students are one of the main stakeholders in education, their opinion about what constitutes effective teaching is very important. However, the traits which show effective teaching cannot be expected to have the same level of importance. For instance, an

instructor's being approachable might be more important for students than his/her being punctual to class, or some students might give more weight to variety of activities during lessons rather than instructors' positivity. Therefore, it can be claimed that the items on SET forms do not have the same relationship with overall teaching quality of an instructor. An instructor's having certain teaching characteristics might make him/her more effective during instruction according to students. Therefore, determining which aspects of instructional effectiveness are more important for students may help create better learning opportunities for them. Hence, this study aims to discover which teaching characteristics of instructors measured through a SET form can predict those instructors' overall teaching quality at a university setting.

Background

Education is probably the most important area to which countries should pay the utmost attention if they want to develop in our modern world. However, the ever-changing nature of our modern world brings new challenges to education. These challenges compel educators to renew their techniques, practices, methodologies, subject knowledge and teaching philosophies, so they need to revise what they teach and how they teach constantly. Therefore, educational institutions need continuous revision, monitoring, and improvement. As it could be expected, higher education is not an exception. However, making necessary changes such as revising course content and exams, or reconsidering teaching methods in educational settings is impossible without receiving necessary feedback from different parties involved in the process, who are policymakers, administrators, educators, students, and parents. The roles of these parties and how they are involved in the process may differ depending on the type of educational level and institution.

Higher education institutions need their students to be involved in decision-making processes as well as policymakers, administrators, and educators because students are the ones who receive the service provided by higher education institutions. Although there are various ways of being involved in decision-making processes for students such as student unions or student & faculty meetings out of class (Darmanin, 2020; Nadler & Nadler, 2000), the most prevalent one is SET with mostly Likert-type questions and occasionally open-ended questions. SETs are vastly used all over the world in higher education institutions along with other means of teaching evaluation such as classroom observations, peer evaluations, self-evaluations, teacher portfolios, and students' test scores. All of these instructional quality evaluation techniques have their disadvantages, which makes SET the most commonly used method of evaluating teaching quality. For instance, Karaca (2022) states that teaching portfolios can be confusing and time-consuming unless they are carefully designed. Moreover, Sayavedra (2014) claims that instructors might feel judged in peer-evaluations although that is not the purpose. Additionally, Pourjamal et al. (2018) state that instructors, especially at the beginning of their professional life, might lack the necessary skills to conduct self-assessment, which can result in misleading results.

SET forms first started to be used at the beginning of 1900s (Gelber, 2020) with the sole purpose of providing personal feedback to the instructors. However, Shevlin et al. (2000) asserted that SETs became the most prevalent form of evaluating teaching quality in the USA and the UK for various reasons, which include faculty decisions, such as tenure, annual reviews, promotion, salary, merit raises, and employment (Barth, 2008; Benton & Cashin, 2014). However, SETs are even more commonly used for giving valuable feedback to instructors about the

lessons, materials, syllabi, and form of teaching (Spooren et al., 2013). Additionally, some institutions use SET results to decide on the courses that will be assigned to instructors and to provide prospective students with information about instructors and lessons (Wang & Williamson, 2022). Moreover, Marsh (2007) states that SETs are used as a component in quality assurance practices nationally and internationally in higher education institutions.

Therefore, SETs are one of the ways which provide valuable information to higher education institutions to improve the quality of teaching. This role of SETs is also emphasized by the Quality Assurance Agency for Higher Education in the UK (Shevlin et al., 2000). Similarly, Australian universities prevalently use SETs as academicians are rewarded depending on their teaching performance, which is a result of performance-based funding in Australia. Until performance-based funding was introduced in 2004, SETs were voluntary in Australia, and the results were obtained by the instructors only. However, the act of quality assurance in higher education and performance-based funding have changed that (Shah & Sid Nair, 2012). Moreover, universities in the Arab world utilize SETs to monitor teaching effectiveness in higher education as well. For instance, National Commission for Academic Accreditation and Assessment in Saudi Arabia and other agencies use SETs and other instruments to evaluate the quality of education (Mohammed & Pandhiani, 2017).

The fact that SETs have been used all around the world since the beginning of the 20th century makes it one of the most studied research areas (Linse, 2017). For years, researchers have questioned how well SETs can measure or reflect teaching effectiveness, whether SETs yield biased results, whether course characteristics can affect results, how well students can comprehend what is asked in SET forms, and

numerous questions as such have been in the center of SET studies. For example, Kulik (2001) asserts that SETs are valid in measuring instructional effectiveness, basing their view on a literature review study, and states that students are most likely to give higher grades to instructors who can teach them more effectively. This situation refers to the validity hypothesis. The validity hypothesis suggests that students who are successful in a course get higher grades in that course; therefore, they give higher grades to their instructors in SET forms because they believe that their success is the result of their instructors' effective teaching (Wang & Williamson, 2022).

On the other hand, opponents of the validity hypothesis claim that instructors may manipulate their SET scores by giving higher grades to students. This phenomenon is called grading leniency. Grading leniency hypothesis refers to the idea that students tend to give higher scores to their instructors in SET forms when they expect or know that they will get high scores from those instructors' classes (Wang & Williamson, 2022). In a study conducted by Crumbley and Reichelt (2009), it was found that more than half of the accounting professors who were surveyed admitted knowing other professors who intentionally made course content easier to be able to give higher grades to their students so that they can get higher grades in SET forms. There are also other studies that show SETs are not trusted by instructors. Morgan et al. (2003) claim that the majority of instructors do not trust SETs for the reason that SET results can be affected by course type and workload, the grades that students get from the course, and the instructor's personality.

Although validity hypothesis and leniency hypothesis seem to be about the same thing, they suggest completely different ideas. Validity hypothesis suggests that SET forms are valid measures of effective teaching because students believe their

instructors make them successful in a course, so they are effective instructors, which in turn results in higher SET scores. However, leniency hypothesis suggests the idea that students rate their instructors high in SET forms if they believe they can get higher course grades easily without showing much effort. Therefore, instructors can manipulate their own SET scores by making the course content relatively easy, which means buying students' rating in exchange for course grades (Weinberg et al., 2009). This hypothesis implies that SET is not a valid evaluation method of teaching performance. In the SET literature, there are many studies which support or refute the validity hypothesis and the leniency hypothesis, which makes use of SET a controversial issue. For instance, Stroebe (2020) claims that SETs cannot be trusted as a form of evaluating teaching quality because students tend to rate the instructors of easier courses higher, and they also tend to do the same when they get good grades, and if given a chance, students register for courses which offer good grades in general. On the contrary, Marsh (2007) asserts that SET is a reliable form of evaluation because there is a positive correlation between students' actual learning and the SET scores that they give to their instructors. Despite these contradictions, SET continues to be one of the most commonly used methods of evaluating teaching in higher education.

There are other factors which are claimed to affect the validity of SET in the literature. The construct irrelevant factors include instructors' gender, likability, race, nationality, students' prior knowledge, motivation, class attendance, class size, and course type (Adams et al., 2021; Arnold & Versluis, 2019; Clayson, 2022; Feistauer & Richter, 2018; Subtirelu, 2015). Although these factors are not relevant to the instructor's teaching skills or under the control of the instructor, some of them might be associated with teaching effectiveness to some extent. For instance, while gender

is not supposed to be related to teaching quality, years of experience of an instructor might be indirectly related to teaching quality as experience makes a difference in how instructors conduct their lessons. Brockx et al. (2011) conducted a study to examine the relationship between SET results and student, teacher, and course characteristics. They concluded that students' attendance to classes has a positive relationship with SET results; however, they have asserted that this can be explained in two ways. First of all, students may attend effective instructors' courses more as they enjoy these classes more, and in turn, may become more motivated leading to more learning, which implies the validity of SET. On the contrary, students who attend classes more might be intrinsically more motivated to learn, which makes them prone to rank their instructors higher in SET forms, but in fact, these students might be the main reason why they perform well in class, not their instructors.

Another study conducted by Weinberg et al. (2009) concluded that students tend to give lower scores to female instructors and foreign instructors, which is a sign that SET forms are affected by construct irrelevant factors. Additionally, Uttl (2021) claims that instructors of quantitative courses such as mathematics, science, and engineering are rated lower than the instructors of non-quantitative courses such as English, history, and psychology. This shows that the instructors of quantitative courses are in a disadvantaged position if SET forms are used for promotion or merit pay at a university. Moreover, Subtirelu (2015) found out that instructors who have Chinese or Korean surnames are rated much lower than the instructors with American surnames on RateMyProfessors.com, which is a sign that foreign-born instructors are also in a disadvantaged position when it comes to SET.

Additionally, Pounder (2007) asserted that students at higher levels of a course are more lenient while evaluating their instructors as opposed to lower-level

students. Another study conducted by Jones et al. (2022) concluded that instructors' SET scores are positively correlated with the motivational climate in their class, which implies that students' motivation is a biasing factor in SET forms. The same study also revealed that students' interest and success in a class positively affect the way that they evaluate their instructors. Similarly, Grammatikopoulos et al. (2014) also discovered in their study that students who have higher levels of interest in a course evaluate their instructors more positively.

Most of the construct irrelevant factors mentioned above are out of instructors' control. An instructor cannot change his/her nationality or students' prior knowledge. Therefore, bias in these areas leads to unfair evaluation of instructors. On the other hand, some construct irrelevant factors can be controlled by instructors to some extent. For instance, the likability of an instructor is claimed to be a biasing factor in SET forms. As some instructors are aware of that, they can try to increase their likability by bringing students some gifts or by being too friendly towards students, which is not directly related to teaching quality. In that way, instructors can manipulate their SET scores. Simpson and Sigauw (2000) interviewed instructors about SET, and they found that faculty members believe SET forms evaluate instructors' likeability and personality rather than teaching quality. Clayson (2022) also reported that two-thirds of the total variance in SET results can be explained with instructors' likeability without taking other factors into account. Feistauer and Richter (2018) found that likeability causes a large amount of bias in SET forms, as well.

Nevertheless, concerns regarding SET are not limited to those mentioned above. Although SET forms are crucial in informing instructors and other stakeholders in higher education about the quality of teaching, finding the most

powerful predictors of effective teaching is another issue because the definition of effective teaching changes depending on the context, purpose of the education or even the learners themselves. As widely known, every student has a specific learning style, and the same instructor might not appeal to one student while he/she caters for the needs of another student. As effective teaching should result in effective learning, the correlation between learning and the characteristics regarding instructor and course as measured by SET forms has been widely studied. These studies are vital for instructors because if an instructor is aware of which aspects of his/her teaching contribute to better learning, he/she can make quality lesson plans by taking these aspects into consideration. As teaching is a multidimensional phenomenon, SET forms should include different types of questions related to different dimensions of teaching.

Grammatikopoulos et al. (2014) reported that since there are debates as to the dimensions of effective teaching, some universities prefer to create their own SET forms by including different dimensions; however, this makes it difficult to do thorough research in this area, and these SET forms may lack validity and reliability. To overcome this problem, some universities adopt a commonly preferred method called Students' Evaluations of Educational Quality (SEEQ), which was developed by Marsh (1982). Marsh (1982) included 9 dimensions in his instrument. These are how much learning happened in the class, how enthusiastic the instructor was, how organized the course was, whether the instructor encouraged students to interact with each other, whether the instructor had a good rapport with students, how broad the course content was, whether examinations and feedback given to students were appropriate, whether course assignments were helpful for learning, and finally whether students considered that this course was better than other courses, which is

an overall evaluation criterion. Marsh (1982) also stated that according to the instructors, these nine dimensions were found to be important to evaluate teaching quality.

Özgüngör and Duru (2015) conducted a study to discover the characteristics of instructors who have the highest and lowest student ratings. To achieve this, they collected 23814 students' responses to the SET form used in a university. The qualitative part of the study revealed that four areas were the most important to distinguish instructors with the highest student ratings from those with the lowest ratings. These four dimensions were lecturing quality, having a good rapport with students, proving to be an expert in the related area, and the quality of evaluation. These results are in line with Marsh's (1982) findings.

However, Biggs (1999) offered a different viewpoint to discover the characteristics of effective teachers. According to Biggs (1999), the concept of effective teaching also depends on educational institutions' values. To illustrate, whether teaching is student-centered or teacher-centered might change the way in which instructors are evaluated. In teacher-centered education, the characteristics that show an instructor's effectiveness focus on conveying course content. On the other hand, in student-centered settings, the focus might be on how well teachers facilitate learning by promoting active participation, learner autonomy, and critical thinking skills.

Problem

The use of SET forms has remained a controversial topic in the literature. Although many studies support the ability of SET forms to assess effectiveness, there are some research studies that question whether these instruments can truly measure the significant aspects of effective teaching (Shevlin et al., 2000).

Another concern regarding SETs is that researchers cannot agree on aspects of the SETs that show effective teaching, nor can they agree on which aspects are more important than others (Patrick & Smart, 1998). The aspects that can indicate effective teaching include having classroom management skills, giving clear instructions, demonstrating good communication skills, being organized, asking stimulating questions, and showing respect for students (Shevlin et al., 2000). Even though there are many predictors of effective teaching, they all cannot be expected to have the same level of predictive power for perceived teaching effectiveness.

Therefore, treating all the items in SET forms equally may create problems, such as low-quality learning because considering some aspects more important than necessary or ignoring some important ones can hinder the effectiveness of instruction. For this reason, it is important to determine the most significant predictors of student satisfaction with instructors in SET forms. Identifying these predictors will help make necessary improvements in curriculum and teaching at a preparatory school of a non-profit private university.

Purpose

The primary purpose of this research is to identify the instructional characteristics, which have the highest predictive power for overall student satisfaction as measured by students' ratings for instructors in a SET form. Furthermore, the findings will clarify whether these characteristics differ depending on course type, proficiency level, and time spent by students in preparatory school. The literature provides many insights as to what effective teaching means in general, but conducting a study in a specific context like this might shed light on the needs and preferences of this institution's students. The results can be used by other institutions which have similar groups of students and similar curricula. Different

from the previous studies, this study offers fine-grained information about sub-groups in the whole student sample, which enables the researcher to understand different expectations of students in different course types, proficiency levels, and trimesters.

To this end, decision tree analysis, a relational analysis method, is used in this study. Decision tree analysis can be described as “a multiple variable analysis which uses algorithms to identify a significant set of inputs in predicting the target variable, recursively splitting data into segments based on the identified inputs until the segments hold homogenous observations” (Park & Dooris, 2020, p. 780-781). Decision tree analysis is chosen for this study because being able to analyze multiple variables at the same time allows researchers to discover cause-effect relationships in a more detailed way (De Ville & Neville, 2013). This approach provides a better perspective as predictors may change within different subgroups, creating relationships with achievement. Chi-squared automatic interaction detector (CHAID), which is a decision tree technique, is used in this study because it can provide powerful predictions about the ratings of students for overall teaching quality of their instructors by looking at the responses given to the other questions in the SET form. The decision tree analyses are repeated for whole sample, course type, proficiency level, and time spent by students at preparatory school.

Research Questions

1. Which instructional characteristics of instructors are the most significant predictors of student satisfaction based on student evaluation of teaching?
 - 1.1. Do these instructional characteristics show differences depending on the course type in this preparatory school?

- 1.2. Do these instructional characteristics show differences depending on students' proficiency levels?
- 1.3. Do these instructional characteristics show differences depending on the number of trimesters spent by students in this preparatory school?

Significance

Identifying the characteristics that can predict student satisfaction is important for the reason that it can help instructors make informed decisions while planning their lessons to facilitate learning better and it can lead to more informed and fair decision-making processes in higher education institutions. Marsh (2007) claimed that administrators use SET results for personnel decisions. If administrators are aware of which characteristics of instructors predict effective teaching in general, they can make better decisions while employing new instructors and while deciding on promotion and tenure. In this manner, higher education institutions can maintain and improve their teaching quality.

Moreover, although there are some studies conducted on SET forms that are used in different departments of universities, there are only a few studies conducted in the English language preparatory schools in Türkiye. Therefore, this study will fill in the gap about the expectations of university level preparatory school students from their English language instructors in Türkiye.

CHAPTER 2: REVIEW OF RELATED LITERATURE

Introduction

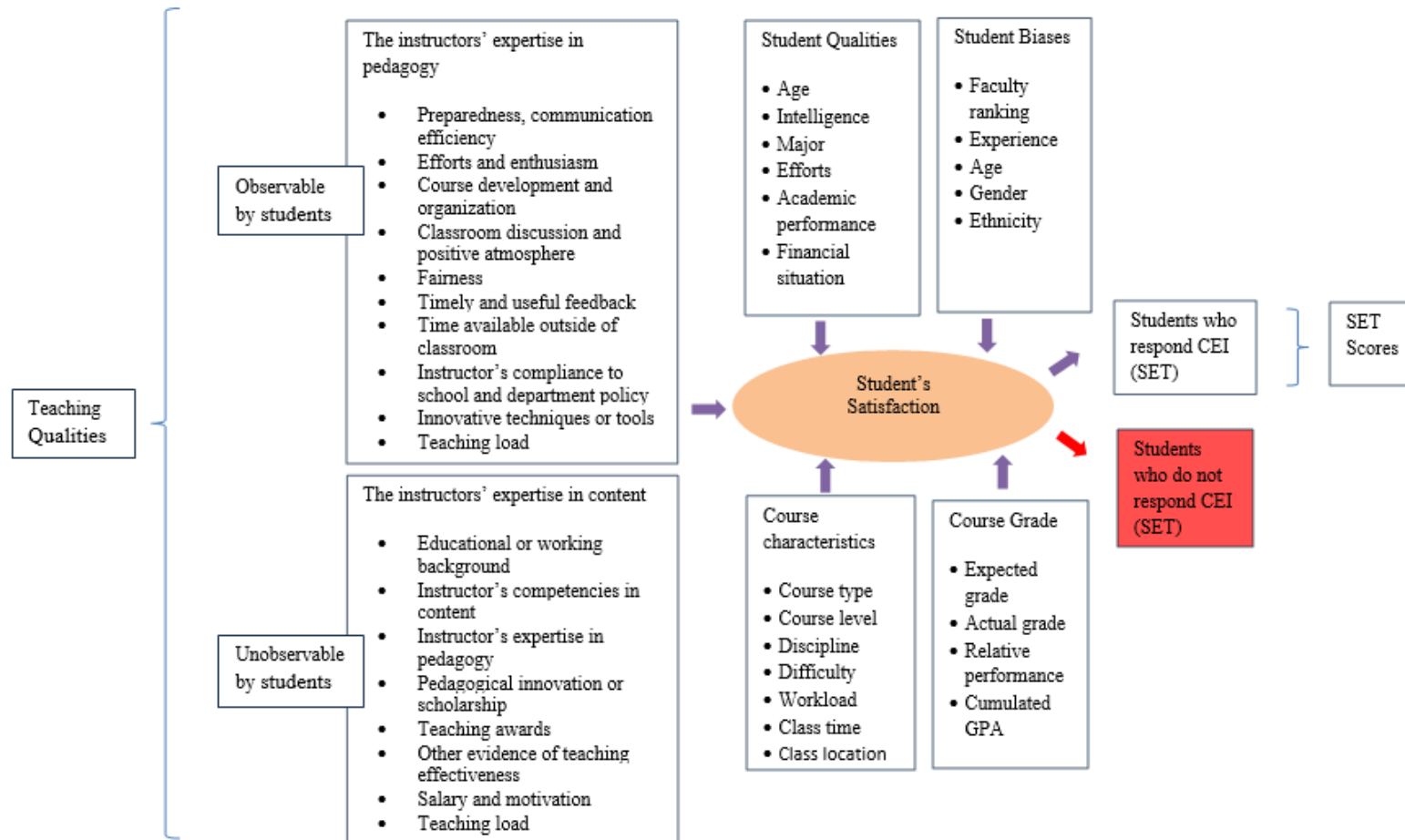
This chapter starts with the conceptual framework of the study. Then, history of student evaluation of teaching (SET), one of the most frequently used instruments to assess teaching effectiveness, is explained in chronological order. After that, studies which support the validity of SET and studies which are about the factors affecting SET forms are presented. Finally, concept of effective teaching in general terms is explained. Through these steps, the researcher aims to provide the readers with detailed background information about SET and the concept of effective teaching with the aim of explaining on which ground this study is conducted.

Conceptual Framework of the Study

SET forms are supposed to measure student satisfaction and teaching quality, and they can be affected by several factors. To illustrate these factors, Wang and Williamson (2022) created a framework. According to Wang and Williamson (2022), students' satisfaction, measured through SET forms, is affected by teaching qualities which are observable by students, teaching qualities which are unobservable by students, student qualities, student biases, course characteristics, and course grades. Figure 1 which details these factors can be seen below. This framework was chosen as the basis of this study because it was the only framework in the literature which focuses on student satisfaction and how it is affected by different factors in detail. Wang and Williamson provided a solid ground for the researchers who wanted to explore this matter in detail.

Figure 1

A Framework for Understanding the Relationship Between Teaching Quality and SET Scores



Note. Reprinted from “Course evaluation scores: valid measures for teaching effectiveness or rewards for lenient grading?” by G. Wang, A. Williamson, 2022, *Teaching in Higher Education* 27(3), p. 308. Copyright 2020 by Taylor & Francis.

This study tries to answer which observable teaching qualities in this framework have a more significant predictive power of students’ satisfaction with overall teaching quality, and whether these observable teaching qualities are evaluated differently depending on course and student characteristics, which are course type, proficiency level, and time spent by students at preparatory school.

History of SET

The use of student evaluation of teaching (SET) dates back to more than one hundred years. In his book *Grading the College: A History of Evaluating Teaching and Learning*, Gelber (2020) stated that in the beginning of 1920s, teaching in higher education in American universities was commonly considered poor quality. Therefore, the need to evaluate university instructors more systematically emerged. However, the first attempts to evaluate instructors were not successful because it was a challenging task to evaluate the teaching skills of highly respected scholars. Moreover, neither the best method to evaluate teaching nor the characteristics of a good teacher were agreed upon at that time, which made the process even more complicated. In addition, each professor and each university had different aims and values while teaching, and this made it harder to establish standards to evaluate teaching effectiveness. Gelber (2020) also stated that throughout the years, universities used different methods to evaluate teaching effectiveness such as peer observations, classroom observations, and using students’ exam results as proof of effective teaching. However, none of these methods have been used as the only

indicator of teaching effectiveness. Besides these different methods, SET has been used prevalently by higher education institutions.

Gelber (2020) wrote that SET was condemned by many in the first half of the 20th century stating that students could not be the authority to evaluate their professors. Nevertheless, SET was already gaining popularity among American universities at that time. The first comprehensive SET conducted in America was administered by the University of Washington in the end of 1920s. The first comprehensive studies on the use of SET were administered by Remmers and Brandenburg (1927) at Purdue University. According to Gelber (2020), in the beginning, SET was not used for important decisions because of its experimental nature. Instead, professors could use it for self-improvement if they wanted. Professors had the choice to refuse using SET as well. Seldin (1993) reported that 29 % of universities in America claimed to use SET to evaluate effectiveness of teaching in a 1973 survey; however, this rate rose to 68% in 1983, and finally it was reported to be 86% in 1993. This means that by the end of the first half of the 20th century, SET had become the most commonly used method of teacher evaluation to make important decisions such as salary, promotion, and tenure. In our present day, SET is commonly used by majority of universities for employing new faculty members, accreditation purposes, promotion, tenure, and teaching excellence awards, but most importantly to improve the quality of teaching (Kulik, 2001; Zhao et al., 2022). Because of that prevalent use, more research has started to be conducted to prove that SET is a reliable and valid method of teacher evaluation.

Even though SET was commonly used in the USA, its introduction to the Eastern countries was relatively late. For instance, SET started to be introduced in

China in the 1980s, but it quickly became prevalent thanks to the acts of Ministry of Education to improve undergraduate teaching in colleges (Zhao et al., 2022).

Although SET was paper based in the past, it has evolved into electronic evaluation in the recent years for reasons such as increased accessibility for students, better anonymity, easier implementation, faster analysis of the results, and decreased costs (Al-Maamari, 2015).

As mentioned above, SET has been around for about a hundred years, and during this time, considerable number of studies have been conducted related to it. In the following section, studies about the use of SET are shared.

Studies Related to the Use of SET

The first meta-analysis of studies showing the relationship between student rating results and student learning was conducted by Cohen (1981). Assuming that effective teachers help students learn more and therefore receive higher SET scores, Cohen (1981) wanted to prove that idea by analyzing forty-one studies. The results showed that a strong correlation exists between SET scores of instructors and their students' learning. Although some studies do not yield similar results, Cohen's meta-analysis is an indicator that SETs are valid tools of measuring teaching effectiveness. There are also more recent studies about the validity of SET. Stehle and Spinath (2014) discovered that there is a positive correlation between students' self-reported learning levels and the SET scores they provided for their instructors. Kalender (2013) also revealed that students' end of semester grades are positively correlated with the scores they assign to their instructors in SET forms.

Zhao and Gallant (2012) conducted a study to explore construct validity and internal reliability issues in SET forms. They analyzed 73500 SET forms in the USA and found evidence for the validity and reliability of SET.

Whether students were a reliable source of data about teacher effectiveness was questioned, and Ory et al. (1980) conducted a study about this matter. They investigated how well different measures of teaching effectiveness correlate with each other. For this purpose, they conducted SET in the form of a questionnaire, in the form of written comments to open-ended questions, and in the form of group interviews with students. Students who attended the study evaluated their instructors in these three ways, and then the results were compared. It was found that students were consistent while they were evaluating their instructors in different ways, which means all these three data collection measures had highly correlated results.

Although some experts in education assert that students do not have the necessary skills to evaluate their instructors in a consistent manner, this study proves that students can be a reliable source of information while evaluating instructors.

Arreola (2007) also claimed that if developed, conducted, and analyzed properly, SET forms are reliable and valid tools to evaluate teaching quality.

Daumiller et al. (2022) also wanted to discover whether students can judge their instructors' effectiveness properly, and they conducted a study to explore the correlation between teachers' self-reports on their teaching effectiveness and students' evaluation of teaching. The study showed a positive correlation between SET scores and teachers' self-reports on their own teaching. If teachers and students agree on the concept of effective teaching, it can be assumed that SET is a reliable form of evaluating teaching effectiveness and students can evaluate their instructors correctly.

Barnett et al. (2003) conducted a similar study, which compared the faculty members' self-evaluation and students' evaluation for the same course. The study revealed that the results of these two types of evaluation were positively correlated.

This finding indicates that student evaluation of teaching can be considered as a reliable and valid source although there are doubts as to students' capability to evaluate their instructors (Barnett et al., 2003).

Remedios and Lieberman (2006) conducted a study to reveal the effect of grades, workload, expectations, and goals on SET. They found that although these factors seem to have a minor effect on SET scores, it is not statistically significant; therefore, the study proved that SET results are a valid indicator of teaching quality.

Al-Maamari (2015) carried out a study about the validity of student evaluation of teaching. The researcher wanted to see the association between SET results in a higher education institution and instructors' gender, course type, participation percentage, and class size. The analysis showed that these factors are responsible for only 1.4 % variance in teaching effectiveness mean. This result indicates that students' responses to SET forms are not biased by these factors, which is a sign of validity of SET scores.

Although there is a substantial amount of research which confirms the validity and reliability of SET, there are also other studies which assert that SET is not a reliable form of evaluating teaching effectiveness; moreover, these studies show that there are several factors which affect the reliability and validity of SET results. For instance, in a study conducted by Cui (2021), it was found that students evaluate female instructors significantly higher than male instructors. Moreover, the same study showed that students favor white instructors over non-white instructors in SET forms. Among non-white faculty members, Latin and Black instructors were rated higher than Asian instructors. Therefore, this study points out that instructors' qualities which are not related to their teaching affect the way that they are evaluated by students.

In another study conducted by Brockx et al. (2011), a significant positive relationship between course grades and instructors' SET scores was found, and it was stated that the course grades were responsible for a 10% variance in SET scores. This study might imply that SET results' validity was affected by course grades. However, this could be an indicator that students learned a lot in this course, which resulted in higher course grades. Therefore, students might have rated their instructors high.

Clayson (2009) conducted a meta-analysis to discover the relationship between learning and SET scores. His initial assumption was that if higher SET scores are given to instructors, those instructors must be effective in facilitating learning; therefore, the students of those instructors should learn more during lessons. To prove that idea, studies conducted between the years of 1953 and 2007 were analyzed. The results revealed that although learning and SET scores seem to have a small positive relationship according to some of the studies analyzed, this relationship is situational and it is affected by many factors such as the course type, course measurement techniques, expected grades by students, and instructors' rigor.

Cox et al. (2022) questioned the trustworthiness of students while they were evaluating their instructors in SET forms. Their assumption was that many students regard themselves as customers in higher education in our modern day, and they expect to receive a service from the instructors. To understand this issue, researchers did a survey on 178 university students, and nearly half of them asserted that they think of themselves as customers in their university. Moreover, half of these students who consider themselves customers stated that they provided false information. This study also showed that SET forms are not completely valid as they are affected by bias.

Karaca (2022) also conducted a study on the dimensions of a SET form that was used in an English language preparatory school. She found that students in different proficiency levels and course types have one common expectation from their instructors. It is instructional effectiveness.

Yılmaz (2018) conducted another study on SET forms at a public university in Türkiye. She found that there is no significant correlation between students' and instructors' gender. However, she found that students' discipline and grade affect the scores they give to their professors in SET forms.

Çimen (2010) did a study on SET forms that were used in different departments of a public university in Türkiye. He discovered that students' major and their repeat & non-repeat status affected the way they evaluated their professors.

Factors Affecting Students' Satisfaction Levels with the Instructors

Instructors' Age

The age of instructors has been reported to be one of the factors affecting SET results. Joye and Wilson (2015) studied the effect of professors' age on their students' perceptions. 308 students participated in the study. These students were given an audio recording of a lecture with a picture of a professor who is a younger female, or an older female, or a younger male, or an older male. After the lecture was finished, students were asked to rate perceived rapport and attractiveness of the professor that was shown to them. Students rated younger female professors as having greater rapport and attractiveness compared with older female professors. However, age did not seem to have an effect on male professors.

Smele et al. (2021) wanted to study the female faculty members' experiences with SET forms. In total 34 interviews were conducted. The study revealed that older female faculty members reported to be criticized more harshly than younger female

faculty members in SET forms, and they felt they needed to work more to overcome this problem.

In a study conducted by Haris et al. (2022), students were surveyed about their preferences as to their instructors. They were asked questions related to instructors' age, gender, qualifications, personality, and medium of teaching. 77.5 % of the students reported that they preferred to take classes from younger instructors. Therefore, it can be concluded that if students prefer younger instructors over older instructors, they might also evaluate them in a biased way.

Race & Nationality

In a study conducted by Al-Issa and Sulieman (2007), students were given a survey to find out what kind of factors affected their judgments while they were evaluating their instructors. The study revealed that instructors' nationality is a factor which affects students' evaluation of their instructors. This finding shows that instructors' nationality is a factor that affects students' satisfaction levels.

Another study by Arnold and Versluis (2019) revealed that not the nationality on its own but the cultural values of students affect the way that they evaluate their instructors. For instance, students who come from individualistic societies and collectivist societies evaluate their instructors from different perspectives. While one group may like certain activities during classes, the other one may dislike them.

A study analyzing the effect of professors' race on SET results found that students tended to rate African American professors higher than White professors although they were provided with the exact same lecture recording with a computer-generated picture of an African American professor and a White professor. This indicates that students' ratings were affected by professors' race (Basow et al., 2013).

Gender

Gender is reported to be a factor affecting students' satisfaction levels in the literature. There are studies about how the gender of the professors and the students affect SET results. For instance, Basow et al. (2013) studied the effect of gender on student evaluations. They found that although the students were given the same lecture, male students tended to rate professors higher than did females. However, researchers found no correlation between the gender of the professors and the gender of the students.

Another study conducted by Adams et al. (2022) proved that similar classroom practices by male and female instructors are evaluated differently because students have different expectations from male and female instructors. For example, when male instructors showed authoritative and didactic behavior in class, students perceived it as a positive thing because males are expected to be authoritative in many societies in the world. However, when females exhibited similar behaviors, students evaluated them in a negative way because females are not associated with being authoritative. These findings imply that as students have different expectations from male and female instructors, their satisfaction with instructors differ accordingly.

Academic Performance

There is also evidence in literature that students' grades, which mean their academic performance in a course, affect their evaluation of teaching. For example, Hoefer et al. (2012) discovered that SET scores of instructors and their students' average grades were correlated. Likewise, Spooren and Mortelmans (2006) observed that there was a positive correlation between a student's grades and his/her evaluation of teaching. They also asserted that this finding was not an indicator of

grading leniency because they found that students whose overall grades were higher tended to give higher grades to their professors. It would be unlikely for all the professors in that case to inflate their students' grades for better SET scores. Hence, it can be stated that students who have better grades evaluate their professors higher. Briceno (2020) conducted a similar study and discovered that students who had better grades were more likely to rate their professors higher while less successful students rated the same professors lower.

Course Type

McPherson (2006) conducted a study to find out what kind of factors impact the way students evaluate their instructors. To achieve that, McPherson collected the SET data of 17 semesters at an American university. He included data from principles economics classes which are introductory level courses and upper-division economics classes which offer more detailed knowledge of the subject. He found that students evaluate the instructors of principles classes and upper-division classes differently. In principles classes, instructors were rated higher if their classroom sizes were smaller, but instructors teaching bigger-sized principles classes were rated lower. However, classroom size did not matter in upper-division classes. Moreover, in upper-division classes, instructors of certain subjects such as labor economics, sports economics, and international trade were rated higher compared to the instructors of other subjects. Moreover, it was revealed that students cared about their instructors' experience more in principles classes although this did not seem to have a significant effect in upper-division classes.

Yu et al. (2022) also did research to discover the factors that have an impact on SET results at a Chinese university. The researchers collected data from 27801 students for 1443 undergraduate courses. These courses were classified into two

categories: hard disciplines requiring quantitative data such as engineering and chemistry, and the soft disciplines requiring qualitative data such as literature and art. The findings of this study showed that students from hard disciplines were inclined to rate their professors higher. This finding supports the idea that course type impacts SET results.

Educational Level of Students and Time Spent at School

There is some evidence in the literature that the educational level of students might change the way they evaluate their professors. For instance, Yu et al. (2022) found that professors who taught in lower grades were rated significantly higher than those who taught in higher grades.

Kalender (2011) also wanted to explore the variables which may influence SET results by using discriminant analysis. He analyzed whether number of sections, credit, grade level, and class size affected SET results. He found that grade level of a course had the biggest influence on students' evaluation of that course. The higher the grade level of a course was, the lower the SET results were. The potential reason for this was reported to be the maturity level of students. When students get more mature during the years they spend at school, they can judge the level of teaching effectiveness better. However, in lower levels, students may tend to give higher grades as they are not experienced in higher education. Kalender (2013) also found similar results in his study to profile instructional effectiveness to reveal its relationship to learning. Grade level of a course was the most important variable which affected SET results. Similarly, Hoefer et al. (2012) found that graduate students rate their professors significantly lower than do undergraduate students.

The Concept of Effective Teaching in Higher Education

Teaching can be described as the act of creating opportunities intentionally for learners to gain certain information, skills, or insights (Brown & Atkins, 2002). The concept of effective teaching can be described in different ways depending on the instructors, learners, the subject, or the institution where teaching takes place. For instance, people who support teacher-centered teaching approaches may associate effective teaching only with teaching skills, subject knowledge, and professional skills of an instructor (Taye et al., 2022). However, Brown and Atkins (2002) claimed that effective teaching does not only rely on the skills of an instructor. Students should be active participants of the teaching process, or effective teaching might be hindered. Similarly, Kahn and Anderson (2019) asserted that an instructor's professional skills and experience in teaching increase his/her excellence in teaching; however, student engagement is a factor that should be taken into consideration by instructors for effective teaching. Taye et al. (2022) emphasized that collaboration and understanding between instructors and students is an important aspect of effective teaching. Okpala and Ellis (2005) conducted a study in which they asked students to describe what qualities effective instructors have. The most common qualities that were mentioned were caring for students and their learning, having professional skills that lead to learning, having good subject knowledge, being a dedicated instructor, and having good verbal communication skills to convey the topic of the lesson. On the other hand, different institutions might consider effective teaching in different ways as well. Since effective teaching entails effective learning, the intended outcomes of the curriculum in an institution might enforce certain styles of teaching on instructors. For instance, if a language institution values teaching receptive language skills, then instructors feel obliged to focus on those receptive

skills. Therefore, spending too much time on productive language skills might be seen as ineffective teaching by students in that institution. In a different institution, the complete opposite might happen. Hence, for better evaluation of effective teaching, the values and goals of an institution should be consistent, and instructors should be well-informed about them.

Although effective teaching might show differences depending on the instructors' teaching philosophy, students' attitudes toward teaching, the institution's values and goals, and the culture in which the teaching takes place, still there are certain common aspects of effective teaching in the higher education context.

The characteristics that instructors need to possess for effective teaching have been established in the literature. First of all, for effective teaching to take place, instructors should have deep subject knowledge. Hattie (2012) claimed that without deep subject knowledge, instructors cannot convey the necessary information which is suitable for students, nor can they organize effective lesson materials and tasks.

Secondly, Kurniati and Cahyono (2018) stated that effective instructors should have the necessary skills to be able to recognize students' needs and attitudes toward teaching so that they can design courses which properly address students' needs.

Moreover, in a study conducted by Su and Wood (2012), it was found that students regard instructors as effective if they have deep subject-knowledge, good communication skills, good educational technologies skills, and good sense of humor. The same study revealed that students expect effective instructors to be passionate, inspiring, enthusiastic, supportive, approachable, punctual in giving feedback. Effective instructors are also expected to promote students' independent learning skills.

Similarly, one of the oldest studies conducted in this area belongs to Cattell. Cattell (1931) found in his study that effective university instructors are expected to be intelligent, open-minded, kind, humorous, and enthusiastic. Furthermore, they are supposed to have good subject-knowledge, pedagogical and psychological knowledge, classroom management skills, and good communication with their students. Therefore, it can be concluded that although many years have passed since Cattell's study, common properties of good university lecturers have stayed the same.



CHAPTER 3: METHOD

Introduction

This study was conducted to discover which teaching characteristics of instructors are more associated with their overall teaching quality from students' perspective. In this chapter, research design, context, participants, instrumentation, data collection, and data analysis methods will be introduced respectively.

Research Design

This study is conceptualized as correlational research which has a predictive model. According to Creswell (2012), "in a prediction design, researchers seek to anticipate outcomes by using certain variables as predictors" (p. 341). Polio and Friedman (2016) claim that researchers do not have any control on variables in correlational research, and they only try to explore the relationships among variables. Moreover, variables do not have cause and effect relationships in correlational research. This is in line with the purpose of this study. In this study, the researcher will try to seek which instructional qualities of instructors at a preparatory school are the most powerful predictors of their overall teaching effectiveness and whether these predictors change depending on certain variables which are course type, proficiency level of students, and semester.

Context

The study took place at the preparatory school of a non-profit private university located in the capital of Türkiye. The university was established in 2003, and its medium is 30% English. This requires all the students -except the ones whose major is Turkish Language and Literature or Visual Communication and Design- to

study at the preparatory school before they start their education in their departments. The preparatory school is designed as three periods, each of which lasts 13 weeks. However, students have the option to finish preparatory school in only one or two trimesters depending on their proficiency level. At the beginning of each academic year, all students are given a preliminary examination to be able to place them in the appropriate level. There are mainly 5 levels, which are A Foundation (AF) for beginners, A for elementary level students, B for intermediate level students, C for upper-intermediate level students, and finally D for students who will study English Language and Literature. Moreover, students who fail AF and A levels are placed in A Repeat (AR) level in the following trimester. Similarly, there are also B Repeat (BR) and C Repeat (CR) levels for the students who fail to meet the requirements of B, C, and D levels.

Students who receive 49 or below out of 100 in the preliminary exam, which is different from the proficiency exam, are placed in AF level, and students who do not take the preliminary exam are also placed in the same level. Students with a score between 50 and 64 are placed in A level, students with a score between 65 and above take the proficiency exam called TOEFL ITP. TOEFL ITP is an exam which measures students' abilities in listening, structure and reading. In each part of the exam, there are multiple choice questions which are about variety of topics. Those who take TOEFL ITP and get a score 450 or below (out of 677) are placed in B level. Those with a score between 450 and 499 are placed in C level. Finally, those who receive 500 and above can start studying in their departments directly. Additionally, students can take TOEFL IBT, YÖKDİL (the exam designed to assess language level of academics and public servants in Türkiye including questions from different areas such as social sciences or medical sciences), YDS (the exam designed

to assess overall language level of participants), or PTE Academic exams to be exempt from preparatory school.

In AF level, there are two instructors teaching main course (MC) and skills (SK) lessons. Main course lessons include four skills combined (listening, reading, speaking, and writing) with a focus on grammar rules. Skills lessons focus on strategies and techniques used in daily life for listening, reading, speaking, and writing skills.

In A, B, C, D levels and in repeat levels, there are three different instructors for main course (MC), TOEFL & writing (TW), and listening & speaking (LS) lessons. While MC lessons focus on four skills with a focus on grammar, TW lessons focus on TOEFL strategies and writing skills. Finally, LS lessons focus on listening and speaking skills and strategies. Different from all the levels, AR levels have reading (R) lessons instead of TW lessons. While AF levels have 30 contact hours a week, other levels have 25 contact hours. MC lessons are taught for 15 hours a week in all the levels. Skills classes constitute 15 hours in AF levels. TW, LS, and R lessons are taught for 5 hours a week.

The curriculum of the preparatory school is prepared by the curriculum office. There is one responsible unit member for each level. Instructors are given a weekly program to follow, and it is mandatory that they complete all the necessary content. Apart from the lessons, students can reach their instructors in their office hours and ask for help.

Quizzes, mid-term exams, final exams, online homework assignments, and portfolio tasks related to speaking and writing skills are the evaluation methods used to measure students' performance. To successfully meet the requirements of each level, students are required to have the average of 70 out of 100 in A and AF levels,

but students in other levels are required to have the average of 65 at the end of a trimester. If they fail to do so, they study in the repeat version of the same level.

Continuous professional development of the instructors is highly valued at this institution. One way to help instructors improve themselves is the professional development activities offered by professional development unit. The members of this unit conduct needs analysis at the beginning of each academic year, and depending on the results, they organize certain events such as workshops and talks. Moreover, evaluating instructors' performance is also valued as a way of helping them enhance their teaching quality. For this purpose, PDU members pay classroom visits on a regular basis, and they give useful feedback about the points which can be improved.

The other important evaluation method for instructors is student evaluation of teaching. At the end of each trimester, students fill out a SET survey to evaluate their instructors. The results of these evaluation forms are shared with the instructors only. None of the instructors can see others' results. However, at the end of each trimester, there is a general meeting to which all the instructors and administrators attend. In this meeting, top five instructors who obtain the highest SET scores are shared with others and praised. Moreover, top five instructors who obtain the highest student achievement scores are also praised in this meeting, as well. Furthermore, at the end of each academic year, administrators meet personally with instructors to discuss the results of SET forms.

Participants

The participants in this study are the preparatory school students at a non-profit private university in 2021-2022 academic year, which included three study periods. The data that were used in this study were gathered from the results of 4281

SET forms for each instructor. However, as the students in the 2nd and 3rd trimesters were the same students who completed the 1st trimester, it can be assumed that the total number of students in this study was roughly equal to the number of students in the 1st trimester. In total, there were 2004 SET forms that were filled out in the fall trimester, but as each student filled out 3 forms to evaluate their MC, TW, and LS instructors separately, it can be estimated that the total number of students was approximately equal to 668 (2004/3). There were 1416 SET forms filled in the spring trimester, and 861 in the summer trimester. These numbers are a bit fewer than the actual number of students registered in the preparatory school because students tend to use their absenteeism towards the end of each trimester, so not all the students fill out this questionnaire. Moreover, the forms which include missing data are excluded from this study. In terms of levels, there are 463 forms filled out in AF levels, 386 forms in A levels, 102 forms in AR levels, 1235 forms in B levels, 35 forms in BR levels, 1449 forms in C levels, 510 forms in CR levels, and 101 forms in D levels. Table 1 shows these statistics.

In terms of course type, there are 1384 forms for MC instructors, 1049 forms for TW instructors, 1216 forms for LS instructors, 163 forms for reading instructors, and 154 forms for skills instructors. There are also support instructors who teach fewer hours than do the main instructors in a class. There are also 431 forms for support main course instructors and 66 forms for support skill instructors.

Table 1

Numbers of Students per Level and per Semester

Student Levels	Semester 1	Semester 2	Semester 3	Total
AF Level	463	-	-	463
A Level	363	-	23	386

Student Levels	Semester 1	Semester 2	Semester 3	Total
AR Level	-	102	-	102
B Level	394	791	50	1235
BR Level	-	-	35	35
C Level	429	428	592	1449
CR Level	317	65	128	510
D Level	38	30	33	101
Total	2004	1416	861	4281

Instrumentation

The survey instrument, which is the SET form of the institution where this study was conducted, consists of 23 Likert-type questions (see Table 2). This survey was designed by the measurement and evaluation office at the preparatory school where this study was conducted. The survey was designed in Turkish. Students anonymously answer questions (for each instructor that they have) on a scale of five (1 means *strongly disagree*, 2 means *disagree*, 3 means *not sure*, 4 means *agree*, and 5 means *strongly agree*). The results are shared with the instructors in an envelope when they are ready. None of the instructors can see others' results. Only the administration can see an instructor's SET results. At the end of an academic year, the administration meets with instructors and discusses these results.

For the administration of the survey, students are distributed optical forms to evaluate their instructors. They answer the same questions for each of their instructors in different columns. In those optical forms, option A equals *strongly agree*, B equals *agree*, C equals *not sure*, D equals *disagree*, E equals *strongly disagree*.

The survey includes questions about different dimensions of teaching such as instructors' personality in class, the methods that they use, whether they promote extracurricular activities, and whether they show professional manners in class. Depending on the changing needs of the institution and the feedback given by the instructors, the survey is revised and updated regularly. Table 2 shows the items in the questionnaire.

Table 2

Questions of the SET Instrument

Questions
1. She is positive, lively, and enthusiastic in the lesson.
2. You can easily communicate with her/him (via email or etc).
3. S/he has been supportive and encouraging in my learning process.
4. S/he manages the class well.
5. S/he ensures that the lessons are covered effectively and efficiently.
6. S/he knows the subject of the lesson well.
7. Whenever I need, she recommends studies that may be useful to me and offers the necessary guidance.
8. S/he can be reached during the specified office hours.
9. S/he encourages all students' active participation in the lesson.
10. The activities in the lessons vary.
11. S/he gets in and out of classes on time.
12. S/he comes to classes planned and prepared.
13. S/he uses the lesson time effectively till the end.
14. S/he teaches the lesson in English (except TOEFL lessons).
15. S/he uses visual tools (board, projection, internet, picture, video, e-book,

Questions

course book, etc.) effectively.

16. S/he recommends visual and printed publications (films, series, documentaries, newspapers, magazines, books, online sources, etc.) to support my learning process.
17. S/he encourages my participation in extra-curricular activities (English Friends, Foodie Club, etc.).
18. S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.)
19. S/he asks me questions which require me to express my opinions on a topic or concept discussed in class.
20. S/he enables me to evaluate different perspectives with the questions s/he asks.
21. S/he encourages me to share my views about the topics discussed in class.
22. S/he gives me feedback about my progress in class.
23. I am glad that I have taken lessons from this instructor.

Note. Reprinted from “Dimensions of Student Evaluation of Teaching Forms at University Level” by R. C. Karaca, 2022, Master's thesis, p. 31-32. Reprinted with permission.

Note. This survey instrument is normally in Turkish to make sure that students understand the questions easily. Karaca (2022) translated the questions into English for her master's thesis. After Karaca's study, the survey instrument was revised, and some questions were altered, or some new questions were added. Therefore, I used Karaca's translations of the questions that stayed the same, and I translated the revised or new questions.

Method of Data Collection

The data used in this research were collected in 2021-2022 academic year in three trimesters. During the administration of this survey, students used optical forms anonymously to answer questions about each of their instructors. While students were answering questions, none of their own instructors were available in class. Instead, instructors were randomly assigned to classes where they did not normally teach to make sure that students did not feel intimidated by the presence of their own instructors while completing the survey. This process was to ensure the reliability of the results. Invigilators also gave instructions as to how to fill out the survey to make sure the process was completed effectively. Students completed the first column of the optical form for their MC instructors, the second column for their TW instructors, and the third column for their LS instructors. If they had support MC instructors as well, they filled out the fourth column for them. This survey was conducted in the 12th week of each trimester (each trimester is 13 weeks in total) before the final exam, and it took about 15 minutes to complete. Upon completion of the forms, they were taken to the measurement and evaluation office for analysis. When all the forms were scanned through an optical mark scanner, the results were analyzed by the measurement and evaluation office and shared with the instructors individually in an envelope. At the end of that academic year, instructors had a meeting with the administration to discuss their results personally.

To obtain this data, the researcher has officially asked for permission from the institution where this study is conducted. She has been granted with permission by the administration, and the measurement and evaluation office has provided the researcher with the data after removing the names of instructors not to violate their privacy.

Methods of Data Analysis

This study utilized quantitative data, which were collected through a survey called instructor evaluation form at a preparatory school of a non-profit private university in the capital of Türkiye. A total of 4281 evaluation forms were analyzed on SPSS by using decision tree analysis, specifically CHAID analysis. Decision tree analysis, which is a multiple variable analysis method, was conducted on SPSS to discover which characteristics of instructors were the most important indicators of their overall teaching effectiveness. For this purpose, the researcher considered the 23rd question of the SET instrument (see Table 2) “*I am glad that I have taken lessons from this instructor*” as the dependent variable because this was a question which indicated instructors’ overall teaching effectiveness from students’ perspective. Then, she tried to find any possible relationship between this question and the other questions which focused on more specific teaching characteristics of instructors. In other words, decision tree analysis helped estimate the overall teaching effectiveness by looking at the results of other individual questions in this SET form. In the context of this study, relationships between variables, which were instructor, course, and student characteristics, were discovered.

Figure 2

Sample Decision Tree Diagram

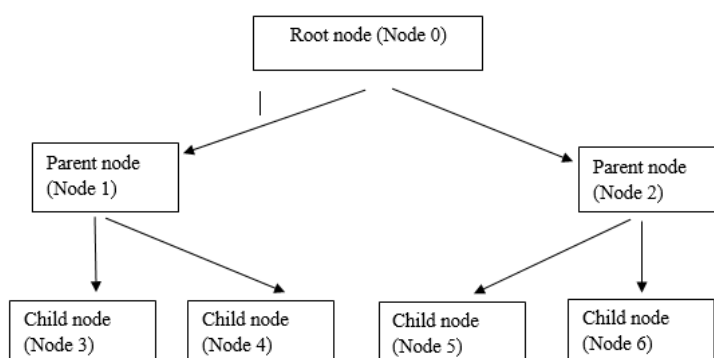


Figure 2 demonstrates how SPSS creates different nodes, which collectively look like an upside-down tree. These nodes are categorized into three. The first node in the beginning of the analysis is the root node. The nodes which produce other nodes under the root node are called parent nodes. However, nodes which do not produce other nodes are called child nodes. The depth of the trees may change according to the data, the number of independent variables, sample size, and the constraints imposed during the analysis.

Decision tree analysis was chosen for this study because decision trees allow researchers to discover the relationships between multiple variables effectively, and the results can be interpreted easily and visually (De Ville & Neville, 2013). Among the alternative decision tree algorithms, CHAID (Chi-Squared Automatic Interaction Detector) algorithm was selected as the method to develop a decision tree in this study. Gordon V. Kass created this data mining method in 1980, and he described it saying, “CHAID partitions the data into mutually exclusive, exhaustive subsets that best describe the dependent variable” (Kass, 1980, p. 119). Kass (1980) also stated that CHAID uses chi-square statistics, Bonferroni method, and category merger algorithm to create a tree diagram, through which the relationship between the most important independent variables and the dependent variable is exhibited. Chi-square statistics is used to show the dependency among different variables (Kirk, 2008). CHAID holds several benefits. First of all, CHAID provides outputs which are simple to comprehend because they are highly visual (You et al., 2014). Moreover, CHAID can analyze variables of all scale types at the same time and presents the relationships between dependent variable and independent variables starting from the most statistically significant one (Yüner, 2023). Additionally, Karakaya et al. (2018) assert that homogeneity assumption is violated when data are collected from a large

sample size; therefore, using CHAID, which does not require homogeneity, is a better option than traditional regression models if sample size is large. Önder and Uyar (2017) also claim that inasmuch as CHAID is nonparametric and nonlinear, missing data do not threaten the results, and normality and homogeneity assumptions are not needed although they are needed in more traditional techniques. All in all, CHAID can produce very powerful estimations while working with large samples. Since this study includes 4281 instructor evaluation forms, CHAID is selected as the preferred data mining method.

Furthermore, student evaluations of teaching generally produce skewed results, providing justification for using CHAID for this data set. Table 3 demonstrates the mean, median, mode, standard deviation, skewness and kurtosis values of each item in the SET form used in this study.

Table 3

Descriptive Statistics of Survey Items

Items	<i>M</i>	Median	Mode	SD	Skewness	Kurtosis
1. Positivity	4.39	5.00	5.00	1.02	-1.76	2.36
2. Communication	4.51	5.00	5.00	0.87	-1.96	3.62
3. Being supportive	4.28	5.00	5.00	1.05	-1.45	1.35
4. Class management	4.13	5.00	5.00	1.15	-1.24	0.60
5. Effectiveness	4.26	5.00	5.00	1.05	-1.43	1.29
6. Subject knowledge	4.61	5.00	5.00	0.79	-2.41	6.12
7. Guidance	4.39	5.00	5.00	0.98	-1.66	2.12
8. Availability	4.49	5.00	5.00	0.86	-1.78	2.85
9. Encouraging active participation	4.18	5.00	5.00	1.14	-1.33	0.84
10. Variety of activities	4.10	5.00	5.00	1.17	-1.14	0.27
11. Punctuality	4.66	5.00	5.00	0.77	-2.74	7.76

Items	<i>M</i>	Median	Mode	SD	Skewness	Kurtosis
12. Preparedness	4.66	5.00	5.00	0.71	-2.59	7.52
13. Using time effectively	4.48	5.00	5.00	0.90	-1.97	3.60
14. Teaching in English	4.57	5.00	5.00	0.84	-2.32	5.38
15. Using visual tools effectively	4.54	5.00	5.00	0.84	-2.13	4.54
16. Recommending publications in English	4.21	5.00	5.00	1.10	-1.31	0.82
17. Encouraging extra-curricular activities	3.84	4.00	5.00	1.29	-0.76	-0.63
18. Applying classroom rules	4.43	5.00	5.00	0.94	-1.80	2.88
19. Asking questions to encourage students to express opinions	4.41	5.00	5.00	0.95	-1.74	2.50
20. Enabling students to evaluate different perspectives	4.30	5.00	5.00	1.02	-1.44	1.33
21. Encouragement to share views	4.35	5.00	5.00	1.00	-1.60	1.95
22. Feedback	4.04	5.00	5.00	1.21	-1.05	-0.04
23. Glad to take lessons	4.32	5.00	5.00	1.10	-1.64	1.77

During the analyses conducted for the course type, SET forms used to evaluate support instructors and skill instructors were not included. The reason for this is that support MC instructors have less contact hours than main MC instructors with a class, and the organization of the course is mainly determined by the main MC instructor. Support MC instructors follow the content as planned by the main MC

instructors. Therefore, they have less control over how the lessons are planned. This rule applies to support skill instructors, as well.

The reason why skill instructors' SET forms were not included in the analyses for the course type is that they had the duties of both TW and LS instructors. Actually, skill courses were a pilot study in which skill instructors had the responsibility of TW and LS instructors combined. They covered both courses' books and content in their lessons. As it was not well received by the instructors, this system was abandoned after the fall semester. Therefore, their results are not included in the analyses for the course type.

Moreover, in this study, parent node values were set to 100. Child node value was 50 for the whole sample, but it was changed to 25 for all the other analyses conducted because the sample sizes were smaller in the sub categories of students when they were compared to the whole sample. This allowed the researcher to be able to conduct deeper analyses as decreased child node values mean increased number of nodes. Finally, no constraint was imposed in terms of the depth of the decision trees.

CHAPTER 4: RESULTS

Introduction

This chapter presents the results of the data analysis. In the beginning, the most significant predictors of instructors' overall teaching performance were discovered through decision tree analysis on SPSS using the whole sample. Next, the same analysis was conducted repeatedly to find the predictors varied depending on course type, student proficiency, trimester, and students' repeat vs. non-repeat status separately. For all the decision tree analyses, the dependent variable was chosen as *glad to take lessons* question (Q23 – See Table 2 for the full list of items), as mentioned before.

Decision tree analyses produce root nodes, parent nodes, and child nodes. The root node shows the distribution of responses provided by students to each response category of the dependent variable. Similarly, each node under the root node shows response distribution of, again, dependent variable conditioning on the previous variables responses. For example, in Figure 3, it can be seen in Node 0, which is the root node, that 64.3% of students in total answered *strongly agree* to question 23 (the dependent variable), which is “*I am glad that I have taken lessons from this instructor*”. The most significant predictor of this question is question 5, which is *effectiveness*. In Node 1, it can be seen that the percentage of the students who choose *strongly agree* for *effectiveness* question is 57.1% (n=2444), and 91.4% (n=2234) of those students also answered *strongly agree* for the dependent variable, which is a higher percentage than the percentages of other answers. This means that

if a student answers *strongly agree* to *effectiveness* question, this student will most likely answer *strongly agree* to *glad to take lessons* question, as well.

Figure 3

A Glimpse into How to Interpret Decision Trees

23. I am glad that I have taken lessons from this instructor.

Node 0		
Category	%	n
Strongly disagree	4,1	177
Disagree	4,9	210
Not sure	9,7	417
Agree	16,9	723
Strongly agree	64,3	2754
Total	100,0	4281

5. S/he ensures that the lessons are covered effectively and efficiently.

Adj. P-value=0,000, Chi-square=3631,285, df=16

Strongly agree

Node 1		
Category	%	n
Strongly disagree	0,2	4
Disagree	0,2	6
Not sure	1,8	45
Agree	6,3	155
Strongly agree	91,4	2234
Total	57,1	2444

While reporting the results of the analyses, following short expressions will be used to refer to each question in the student evaluation form used in this research (see Table 4).

Table 4

Questions of the SET Instrument and Their Shortened Versions

Questions	Shortened Versions
1. She is positive, lively, and enthusiastic in the	1. Positivity

Questions	Shortened Versions
lesson.	
2. You can easily communicate with her/him (via email or etc).	2. Communication
3. S/he has been supportive and encouraging in my learning process.	3. Being supportive
4. S/he manages the class well.	4. Class management
5. S/he ensures that the lessons are covered effectively and efficiently.	5. Effectiveness
6. S/he knows the subject of the lesson well.	6. Subject knowledge
7. Whenever I need, she recommends studies that may be useful to me and offers the necessary guidance.	7. Guidance
8. S/he can be reached during the specified office hours.	8. Availability
9. S/he encourages all students' active participation in the lesson.	9. Encouraging active participation
10. The activities in the lessons vary.	10. Variety of activities
11. S/he gets in and out of classes on time.	11. Punctuality
12. S/he comes to classes planned and prepared.	12. Preparedness
13. S/he uses the lesson time effectively till the end.	13. Using time effectively
14. S/he teaches the lesson in English (except TOEFL lessons).	14. Teaching in English
15. S/he uses visual tools (board, projection,	15. Using visual tools

Questions	Shortened Versions
internet, picture, video, e-book, course book, etc.) effectively.	effectively
16. S/he recommends visual and printed publications (films, series, documentaries, newspapers, magazines, books, online sources, etc.) to support my learning process.	16. Recommending publications in English
17. S/he encourages my participation in extra-curricular activities (English Friends, Foodie Club, etc.).	17. Encouraging extra-curricular activities
18. S/he studiously applies the classroom rules (lesson entry-exit times, mobile phone use, course materials, etc.)	18. Applying classroom rules
19. S/he asks me questions which require me to express my opinions on a topic or concept discussed in class.	19. Asking questions to encourage students to express opinions
20. S/he enables me to evaluate different perspectives with the questions s/he asks.	20. Enabling students to evaluate different perspectives
21. S/he encourages me to share my views about the topics discussed in class.	21. Encouragement to share views
22. S/he gives me feedback about my progress in class.	22. Feedback

Questions	Shortened Versions
23. I am glad that I have taken lessons from this instructor.	23. Glad to take lessons

As the main aim of this research is to identify which instructional characteristics are the most significant predictors of overall teaching effectiveness according to students, focusing on the terminal nodes which have higher percentages of positive answers (*strongly agree* and *agree*) than the root node itself can help understand the most significant predictors. Thus, while reporting the results of all the analyses, the researcher describes only the terminal nodes which produce higher percentages of *strongly agree* and *agree* (combined) answers than the percentage of the same answers given to *glad to take lessons* question. After these nodes are explained, the nodes which produce extremely low positive answer ratios compared with the dependent variable are presented.

The Most Significant Predictors of Overall Teaching Performance Based on Whole Sample

Figure 4 shows the outline of the decision tree produced for the whole data including a total of 4281 evaluation forms. This is the number of the forms after missing data were excluded from the study. The analysis produced 37 nodes in total, except Node 0. Node 0 shows the distribution of the dependent variable (Q23). 12 of the nodes are parent nodes, and 25 of them are child nodes in total at three levels in depth. Among 22 items of the survey in total, only seven of them are used in this decision tree analysis. At the top level, *effectiveness* is the most predictive item. This item created five nodes in the 1st level. *Being supportive*, *encouragement to share views*, and *enabling students to evaluate different perspectives* are the most

significant predictors in the 2nd level. In this level, there are eight child nodes, and seven parent nodes, which produce additional 17 child nodes in the 3rd level. The most significant predictors of the 3rd level are *feedback*, *positivity*, *enabling students to evaluate different perspectives*, *being supportive* and *asking questions to encourage students to express opinions*.

As the whole decision tree is too big to present in only one page, it is divided into five sections. Figure 5, Figure 6, Figure 7, Figure 8, and Figure 9 respectively present each section. After each figure, explanations are given for the related part.



Figure 4

Outline of the Decision Tree Produced for the Whole Group

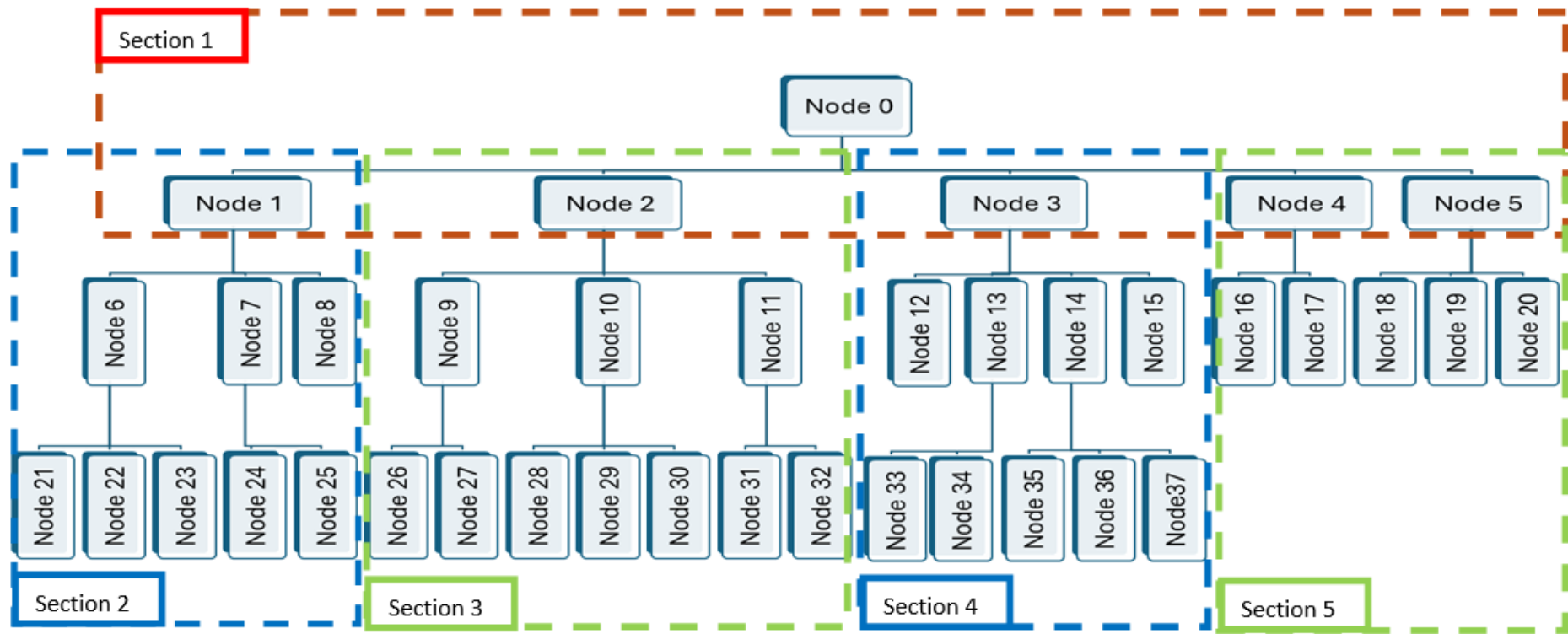


Figure 5

Section 1 of the Decision Tree for the Whole Sample

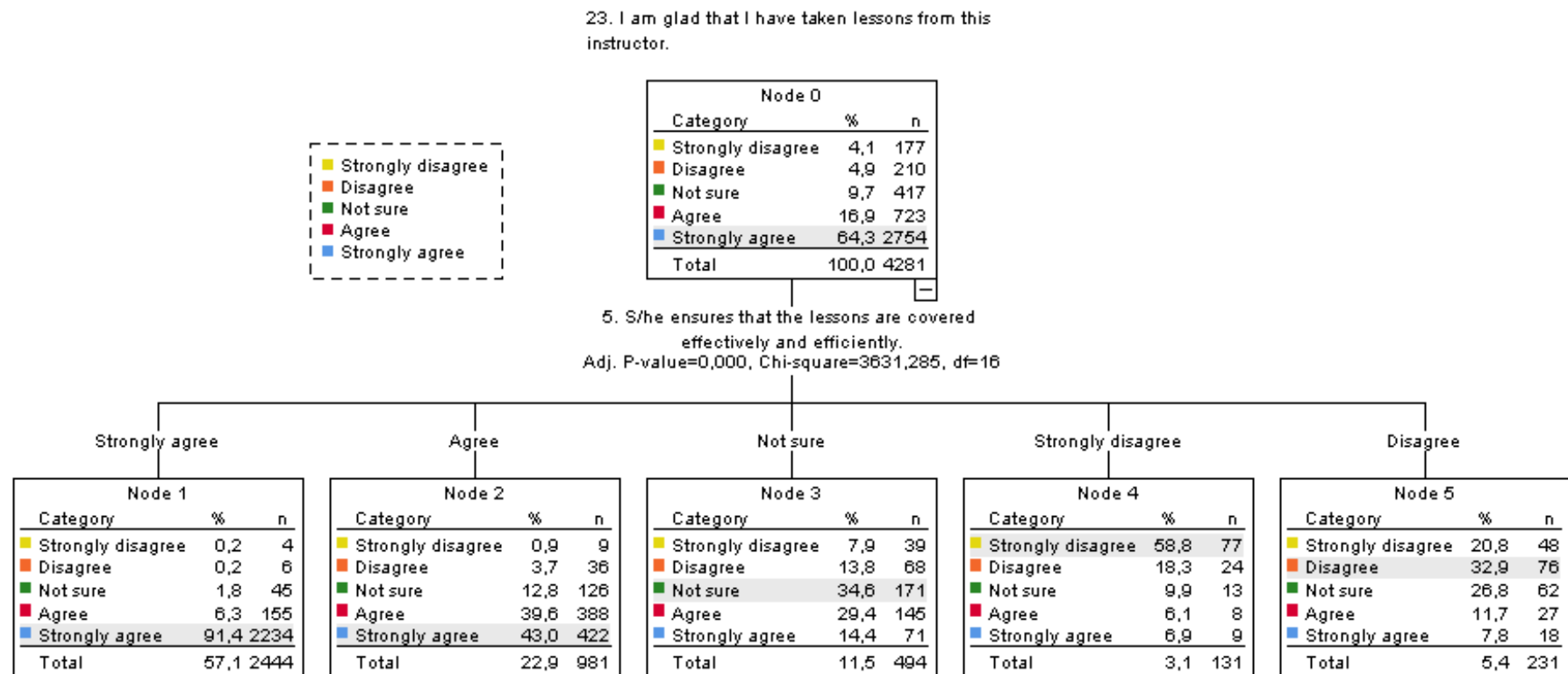


Figure 5 represents the first section of the decision tree which is produced for the whole sample. Node 0 shows that there were 4281 students who took part in this study. 64.3% of them ($n = 2754$) replied *strongly agree* for *glad to take lessons* question. 16.9% of them ($n = 723$) replied *agree*. The most significant predictor of the target question is *effectiveness*. It created 5 nodes (nodes 1 to 5). Node 1 demonstrates that 57.1% ($n = 2444$) of all the students in this sample answered *strongly agree* to *effectiveness* question. 91.4% ($n=2234$) of this group answered *strongly agree* and 6.3% ($n = 155$) answered *agree* to the target question, which is 97.7% positive answers in total. This indicates that there is a strong relationship between questions 23 and 5. Under Node 1, there are more child nodes, but they will be interpreted under Figure 6.

Node 2 shows that if a student answers *agree* to *effectiveness* question, this student is expected to answer *strongly agree* to the target question with a ratio of 43% and *agree* with a ratio of 39.6%. Node 3 presents the students who answered *not sure* to *effectiveness* question ($n = 494$, 11.5%). 14.4% of this group answered *strongly agree* to the target question, and 29.4% of them answered *agree*, 34.6% answered *not sure*, which is the greatest portion of this group. Node 4 shows the students who replied *strongly disagree* to the *effectiveness* question ($n = 131$, 3.1%). These students were highly likely to reply *strongly disagree* to the target question with a ratio of 58.8%, and they also replied *disagree* to the target question with a ratio of 18.3%. Node 5 demonstrates the students who replied *disagree* for *effectiveness* question ($n = 231$, 5.4 %). They had 32.9% chance to answer *disagree* and 20.8% chance to answer *strongly disagree* to the target question (53.7% negative responses).

Figure 6

Section 2 of the Decision Tree for the Whole Sample

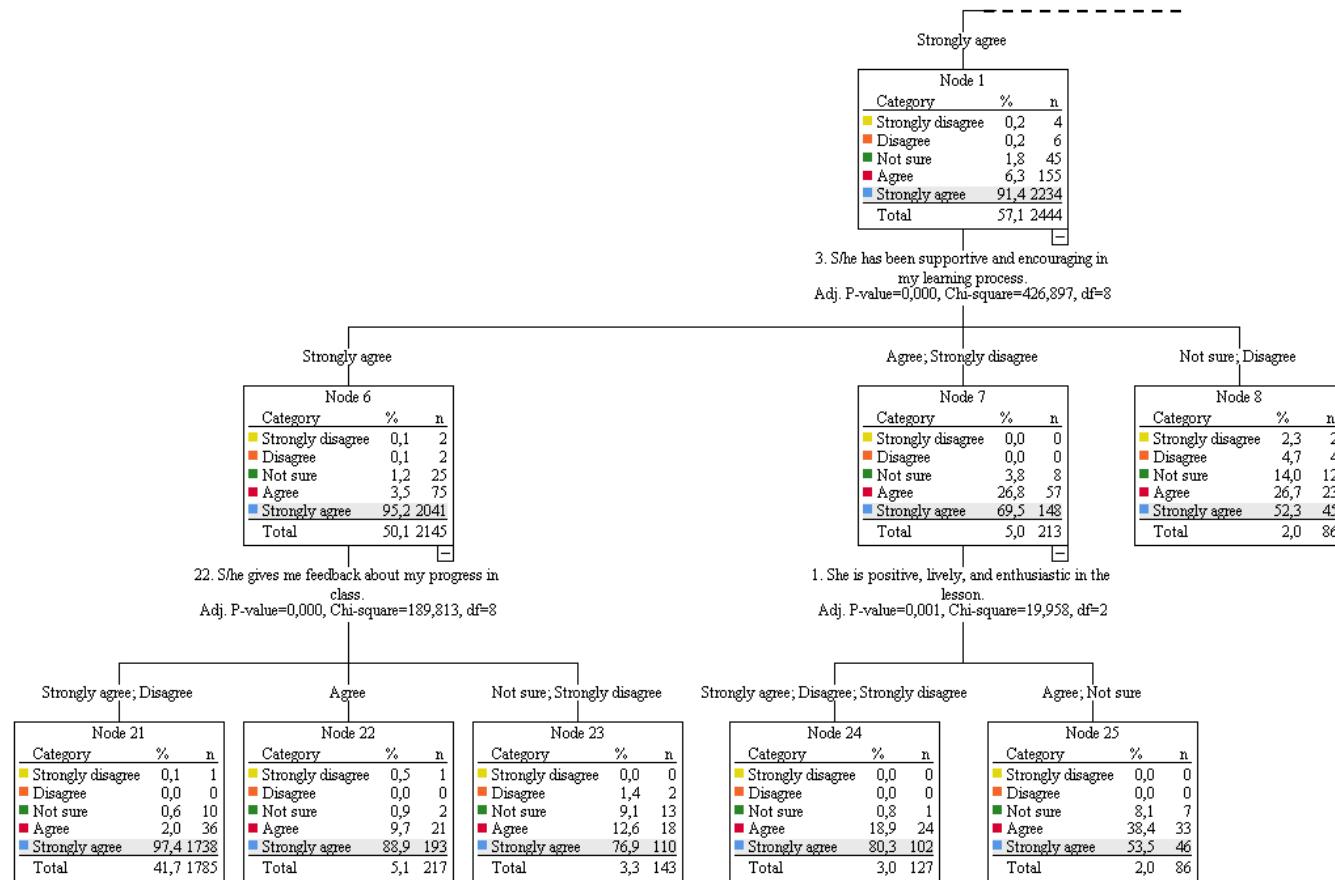


Figure 6 demonstrates the second section of the decision tree produced for the whole group. After the *effectiveness* question, the most powerful predictors are *being supportive*, *encouragement to share views*, and *enabling students to evaluate different perspectives* in the second level. However, Figure 6 only shows the nodes created under *being supportive*. The other predictors will be discussed under Figure 7, Figure 8, and Figure 9. Node 6 ($n = 2145$) shows that most students ($n = 2444$) answered *strongly agree* to *effectiveness* question, and 2145 of them also answered *strongly agree* to *being supportive*. And 95.2% of those students replied *strongly agree* to *glad to take lessons* question. This shows students' possibility of providing a positive answer for the target question increased when they answered *strongly agree* for both *effectiveness* and *being supportive* questions. Node 6 produced three terminal nodes (21, 22, and 23) based on *feedback* question.

In Node 21, it can be realized that when students answered *strongly agree* for *effectiveness* question and *being supportive* question, and when they also chose *strongly agree* or *disagree* for *feedback* question, 97.4% of them answered *strongly agree* to the target question. The same group answered *agree* to the target question with a ratio of 2%. In total 99,4% of these students had positive answers to the target question.

In Node 22, it can be viewed that most of the students (88.9%), who replied *strongly agree* to both *effectiveness* and *being supportive* questions and *agree* to *feedback* question, chose *strongly agree* for the target question. They also chose *agree* for the same question with a ratio of 9.7%.

Curiously, Node 23 presents the idea that students who answered *strongly agree* to *effectiveness* and *being supportive* questions but *not sure* and *strongly*

disagree to the *feedback* question still had a ratio of 76.9% to answer *strongly agree* to the target question and 12.6% possibility to answer *agree*.

Node 7 produces two child nodes, which are Node 24 and Node 25. Node 24 presents that those students who chose *strongly agree* for *effectiveness question*, and *agree* or *strongly disagree* for *being supportive* question, and *strongly agree* or *disagree* or *strongly disagree* for *positivity* question were 99.2% likely to choose a positive answer for the target question (80.3% *strongly agree*, and 18.9% *agree*).

Node 25 shows that if the same group of students chose *agree* or *not sure* for the *positivity* question, they were 53.5% likely to choose *strongly agree* for the target question, and 38.4% likely to choose *agree* for the target question. This means that there is a total of 91.9% chance for them to provide a positive answer for the target question.

As Node 8 shows fewer students with positive responses as compared to Node 0, it was not considered for interpretation.

Figure 7

Section 3 of the Decision Tree for the Whole Sample

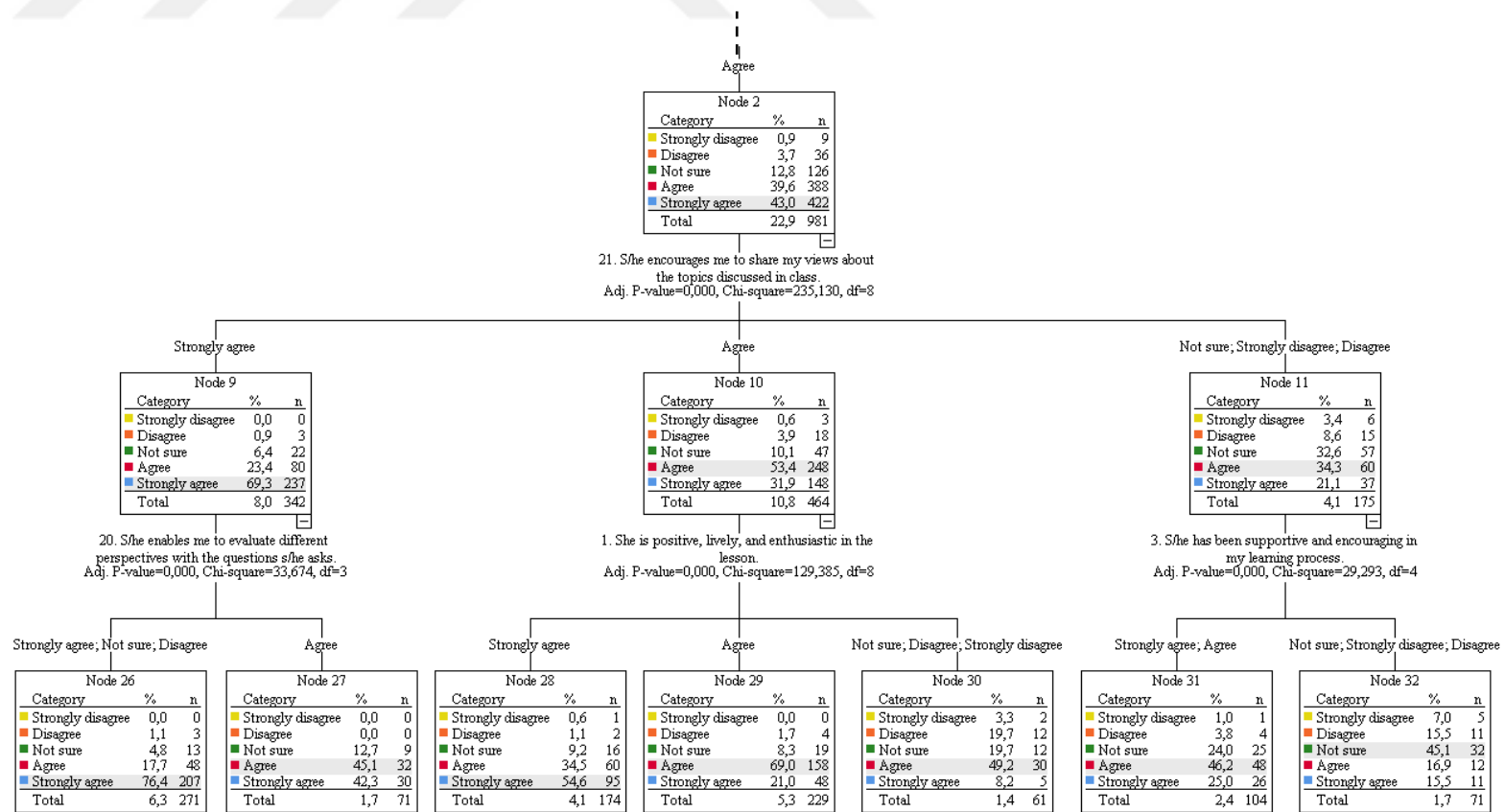


Figure 7 shows section 3 of the decision tree which is created for the whole data. The first node on the top is Node 2. It is described under Figure 5. Nodes under Node 2 were created based on question 21 (*encouragement to share views*).

Under Node 9, Node 26 and Node 27 were produced. Nodes 26 and 27 were created based on question 20 (*enabling students to evaluate different perspectives*), which is the next most significant predictor of effective teaching on both the second level and the third level.

However, Node 26 presents intriguing results. In Node 26, it can be seen that of all the students who replied *agree* to *effectiveness* question and *strongly agree* to *encouragement to share views* question, 6.3% answered *strongly agree* or *not sure* or *disagree* to *enabling students to evaluate different perspectives* question. However, 76.4% of this group of students answered *strongly agree* to the target question, and 17.7% of them answered *agree* to the target question. In total, positive answers with a possibility of 94.1% are observed in this group of students.

On the other hand, Node 27 presents that students who chose *agree* to *effectiveness* question and *strongly agree* to *encouragement to share views* question and *agree* to *enabling students to evaluate different perspectives* question had the possibility of 42.3% to choose *strongly agree* to the target question and 45.1% to choose *agree* to the target question.

Node 10 demonstrates the students who chose *agree* for both *effectiveness* question and *encouragement to share views* question. Those students had the ratio of 31.9% to choose *strongly agree* and 53.4% to choose *agree* to the target question. Node 10 produced Nodes 28, 29, and 30.

Node 28 shows that if this same group of students in Node 10 also answered *strongly agree* to *positivity* question, they answered *strongly agree* to the target

question with a ratio of 54.6% and *agree* with a ratio of 34.5%, which means these students were 89.1% likely to give a positive answer to the target question in total.

However, Node 29 shows that if these students in Node 10 answered *agree* instead of *strongly agree* to *positivity* question, they chose *strongly agree* for the target question with a ratio of 21% and *agree* with a ratio of 69%. In total, their chance to provide a positive answer to the target question was 90%. This indicates that being positive is another significant predictor of effective teaching.

Node 11, Node 30, Node 31, and Node 32 are not detailed here as they do not produce positive answer ratios higher than the target variable.

Figure 8

Section 4 of the Decision Tree for the Whole Sample

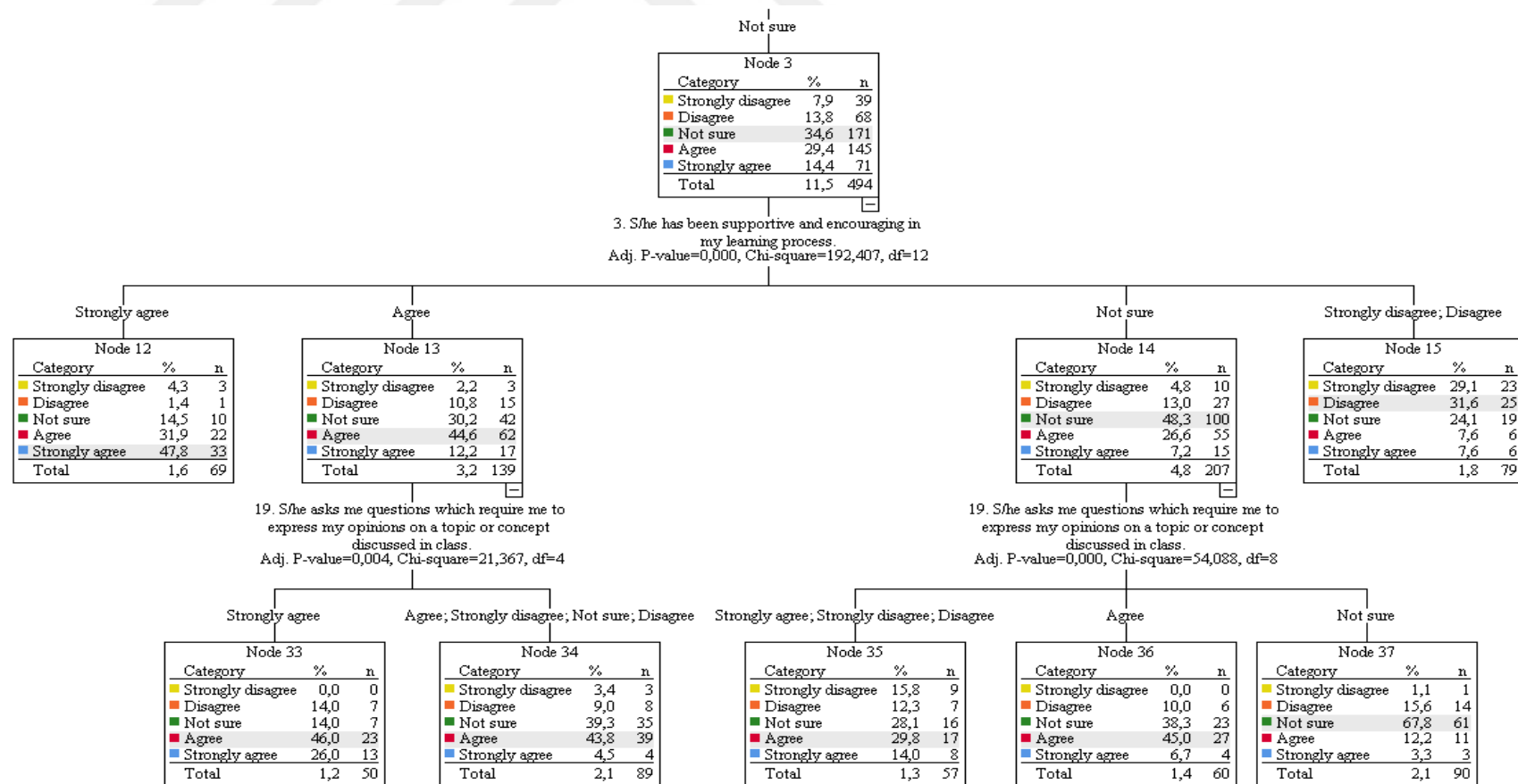


Figure 8 illustrates section 4 of the decision tree. At the top-level, Node 3 is available as previously described under Figure 5.

However, this section will not be described in detail because none of the nodes can produce a higher combination of *strongly agree* and *agree* answer percentages than the root node itself. Root node has a ratio of 81.2% positive answers in total. The reason for this choice is that this study focuses on the instructional characteristics that make instructors effective in the eyes of students. However, if a node cannot produce more positive answer percentages than the root node itself, this means the related item does not increase students' possibility of giving a positive answer to the target question. These results might still be important, but because of the focus of this research, they will not be included here.

In this part, it might be useful to mention the nodes which produce the lowest positive answer percentages. Node 15 demonstrates that students who answered *not sure* to *effectiveness* question and *strongly disagree* or *disagree* to *being supportive* question had a ratio of 7.6% to answer *strongly agree* and again a ratio of 7.6% answer *agree* to the target question. In total, there was only 15.2% chance that these students provided positive answers to the target question.

Additionally, Node 37 also produces a very similar result. According to Node 37, students who answered *not sure* to *effectiveness* and *being supportive* and *asking questions to encourage students to express opinions* questions had 3.3% ratio of answering *strongly agree* and 12.2% ratio of answering *agree* to the target question.

Figure 9

Section 5 of the Decision Tree for the Whole Sample

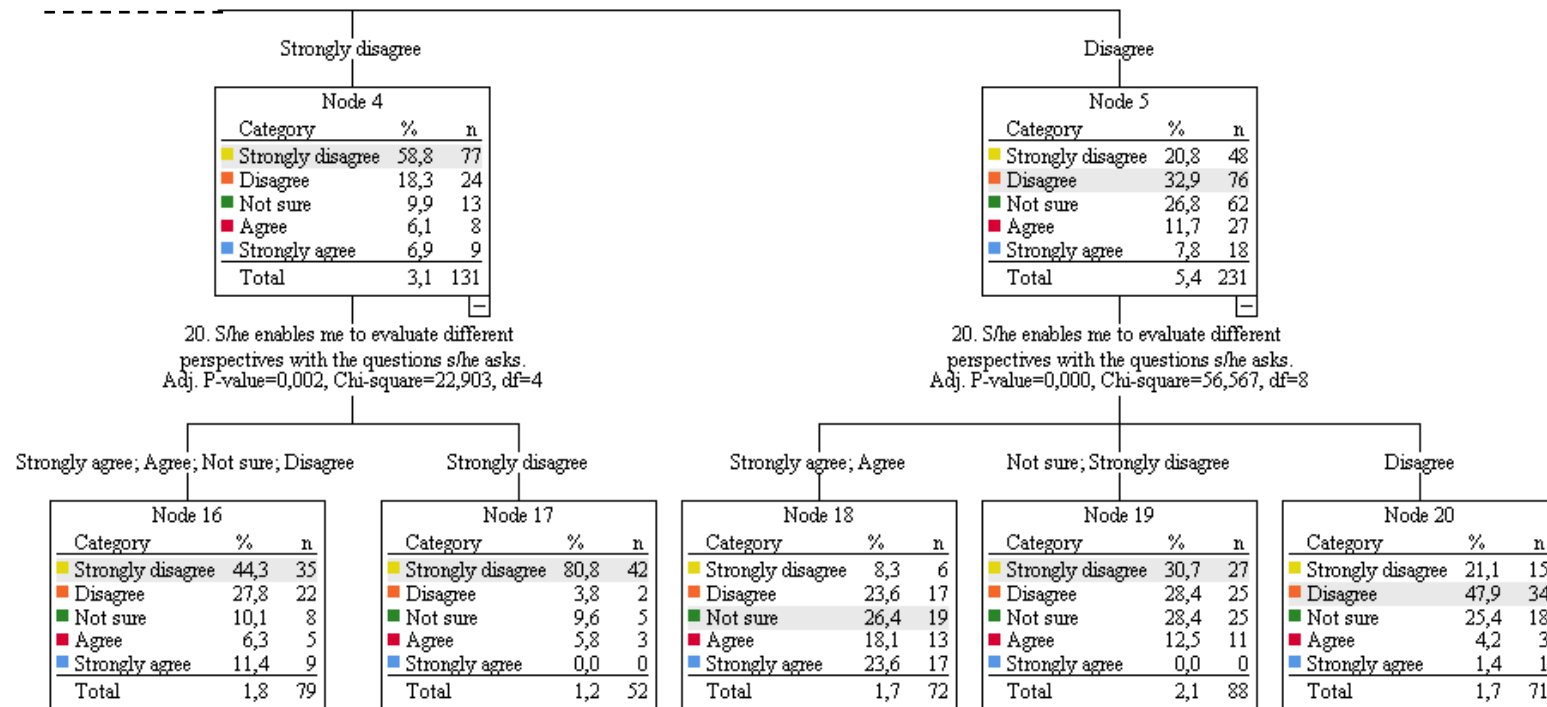


Figure 9 represents the students who answered *strongly disagree* and *disagree* to the *effectiveness* question. However, none of the nodes in this section can produce positive answers with a higher percentage than the root node itself.

Therefore, each node will not be described separately.

On the other hand, two nodes seem to produce the lowest positive answer scores compared with the root node. Node 17 shows that students who *strongly disagreed* with *effectiveness* question and at the same *strongly disagreed* with *enabling students to evaluate different perspectives* question never provided *strongly agree* answer to the target question, which is striking. And only 5.8% of them answered *agree* to the target question.

Node 20 has similar results. Students who replied *disagree* to both *effectiveness* question and *enabling students to evaluate different perspectives* question had a 1.4% ratio to choose *strongly agree* and only a 4.2% ratio to choose *agree* for the target question.

Before moving on with the new analysis, this part can be summarized as follows: After question 5 (*effectiveness*), the most significant predictors seem to be question 3 (*being supportive*), question 21 (*encouragement to share views*), and question 20 (*enabling students to evaluate different perspectives*) in the second level. In the third level, question 22 (*feedback*), question 1 (*positivity*), question 19 (*Asking questions to encourage students to express opinions*) along with question 20 and question 3 again are the most significant predictors.

The Most Significant Predictors of Overall Teaching Performance Depending on Course Type

This part will demonstrate the results of decision tree analyses created for main course (MC), TOEFL&writing (TW), and listening&speaking (LS) instructors separately. TOEFL ITP is the proficiency exam that students take at the end of their preparatory school education, which is why there is a course type named TOEFL&writing.

In the following pages, Figure 10 presents the outline of the decision tree for MC lessons. Figure 11 and Figure 12 show the results for MC instructors.

The Most Significant Predictors of Overall Teaching Performance for MC Instructors

Figure 10 presents the outline of the decision tree for MC instructors. There are 11 nodes except the root node. In the first level, *effectiveness* is the most significant predictor of teaching quality, and it creates four nodes. In the second level, there are five nodes, and *positivity* and *enabling students to evaluate different perspectives* are the most important predictors. *Encouragement to share views* is the most important predictor in the third level, and it produces two nodes.

Figure 10

Outline of the Decision Tree for MC Instructors

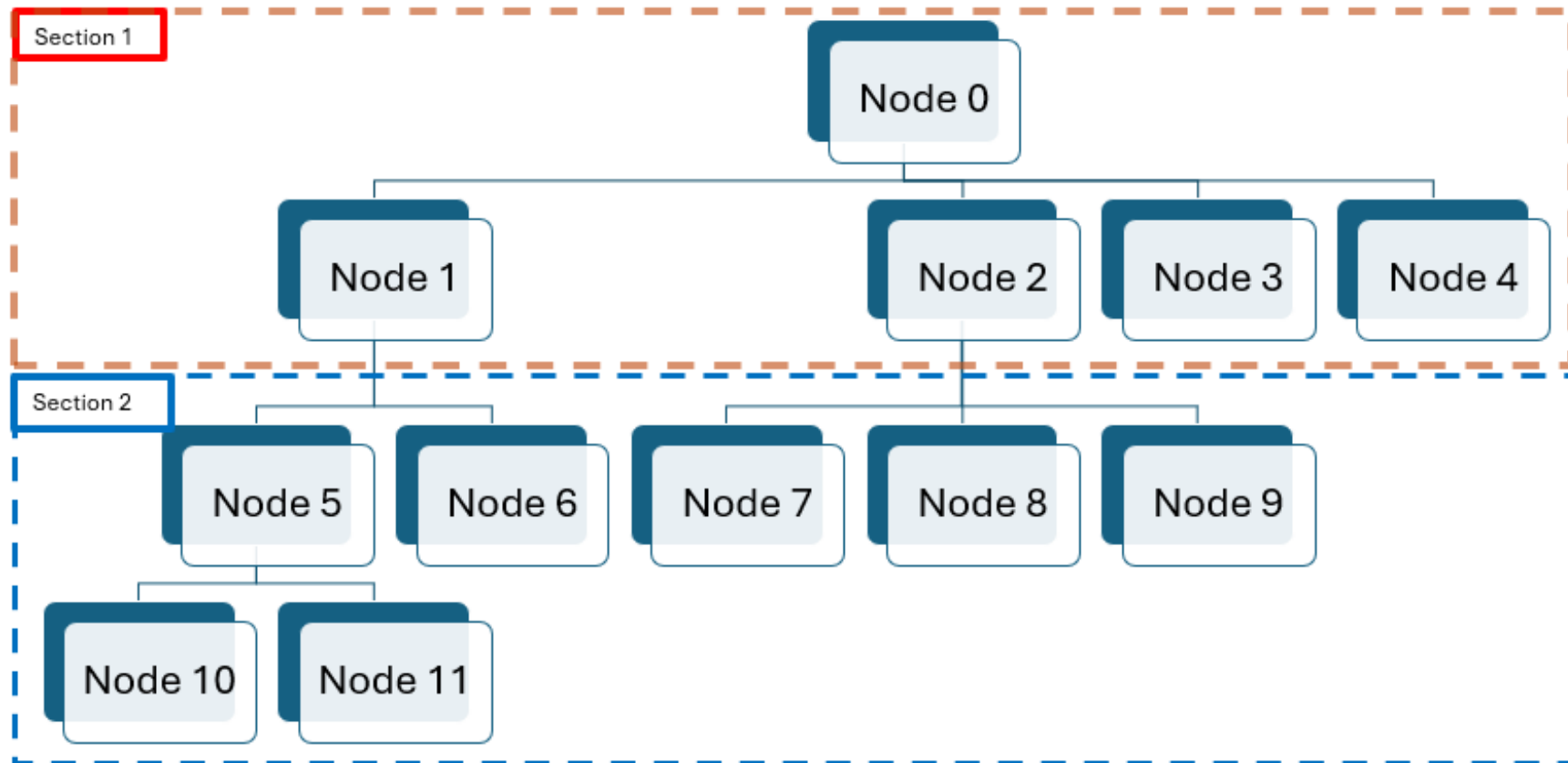


Figure 11

Section 1 of the Decision Tree for MC Instructors

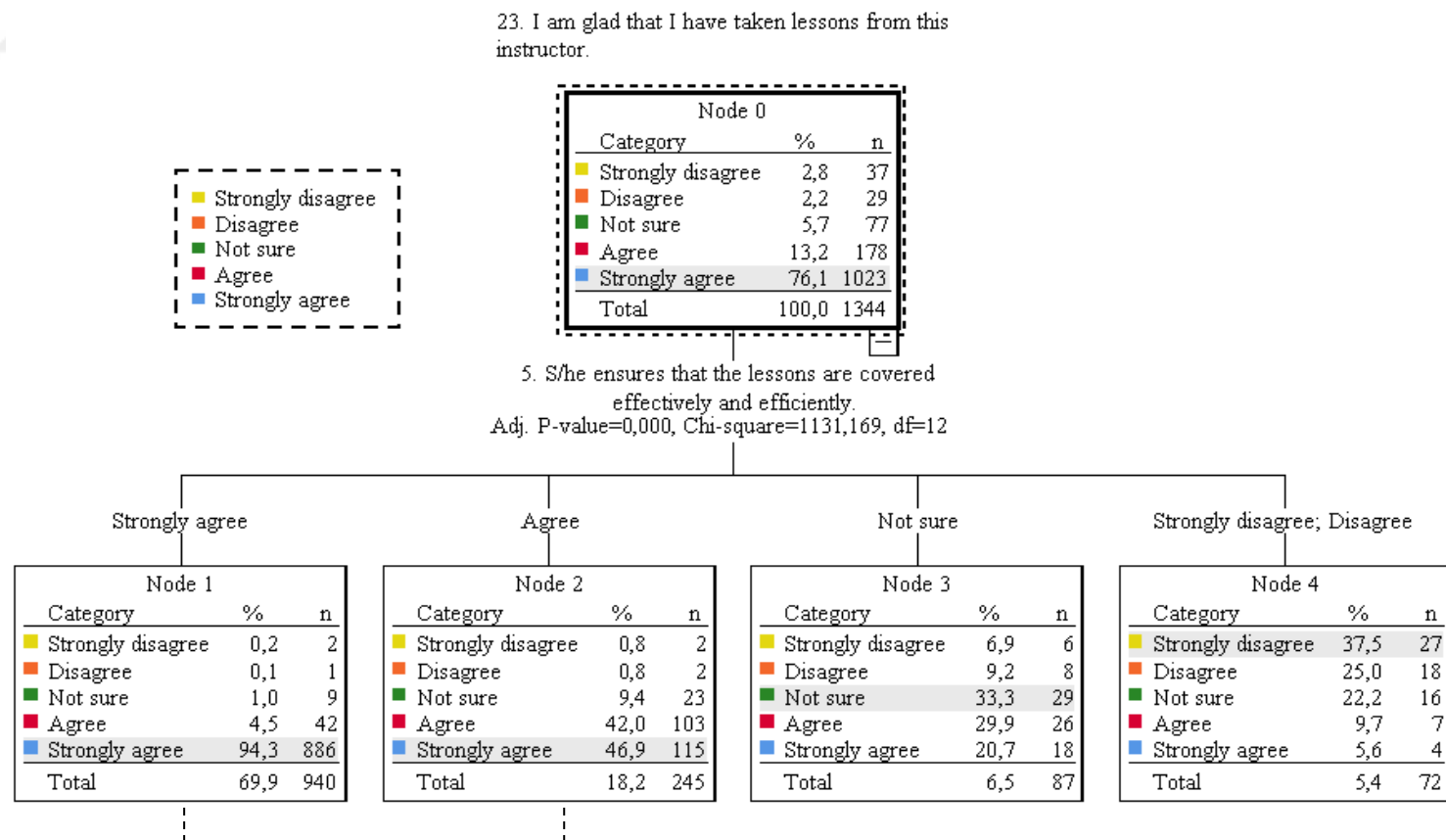


Figure 11 demonstrates the first section of the decision tree created for MC instructors. Node 0 exhibits the students who provided answers for *glad to take lessons* question for their MC instructors. Of all 1344 students, 76.1% (n = 1023) answered *strongly agree* and 13.2% (n = 178) answered *agree* to the target question. Similar to the findings of the whole group (see Figure 5), the most significant predictor is found to be *effectiveness* question in the first level for MC instructors.

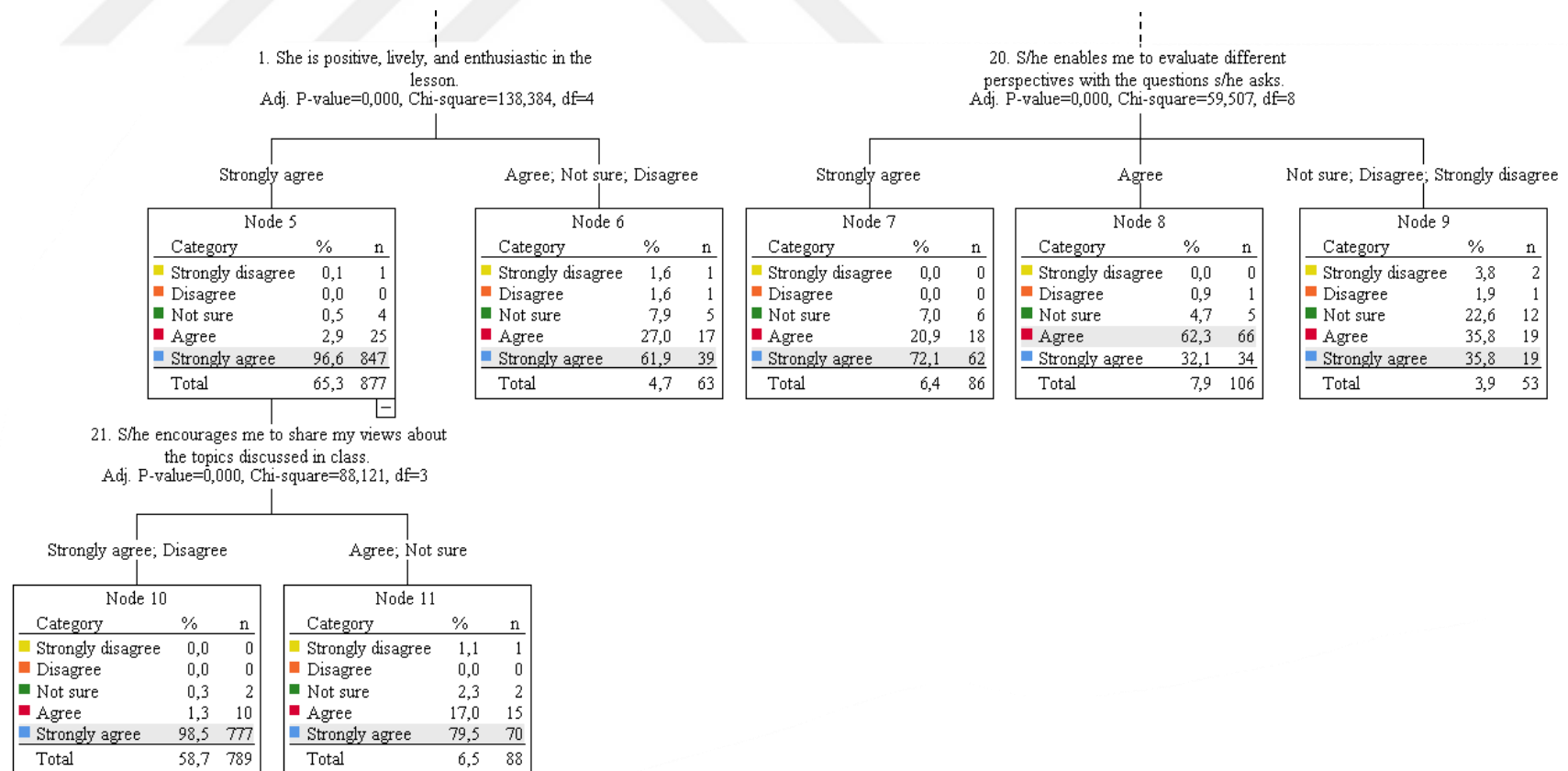
Node 1 shows that students who replied *strongly agree* to *effectiveness* question answered *strongly agree* to the target question with a ratio of 94.3% and *agree* with a ratio of 4.5%.

Node 2 demonstrates that 18.2 % of all the students replied *agree* to *effectiveness* question. However, with that answer, students' ratio of replying *strongly agree* to the target question decreased from 94.3 % (Node 1) to 46.9% (Node 2), which is a drastic change. The same group replied *agree* to the target question with a ratio of 42%. This node created three child nodes.

Node 3 and Node 4 show the students who replied *not sure*, *disagree*, and *strongly disagree* to *effectiveness* question. None of these answers increase students' possibility to provide a positive answer to the target question. Therefore, they are not detailed here.

Figure 12

Section 2 of the Decision Tree for MC Instructors



In Figure 12, which shows the second section of Figure 10, Node 5 shows that students who replied *strongly agree* to both *effectiveness* and *positivity* questions replied *strongly agree* again to the target question with a ratio of 96.6% and *agree* with a ratio of 2.9%, which is 99.5% positive answers in total.

In the third level under Node 5, *encouragement to share views* question is the most significant predictor of overall teaching quality. Node 10 presents that when students answered *strongly agree* to both *effectiveness* and *positivity* questions and *strongly agree* or -interestingly- *disagree* to *encouragement to share views* question, students still answered *strongly agree* to the target question with a ratio of 98.5% and *agree* with a ratio of 1.3%, which is in total 99.8% chance to provide a positive answer. *Encouragement to share views* item seemed to function abnormal as both positive and negative responses provided an increase.

Moreover, Node 11 shows that when the same group of students answered *agree* or *not sure* to *encouragement to share views* question, they had a ratio of 79.5% to say *strongly agree* to the target question and 17% ratio to say *agree*, which is 96.5% positive answer ratio. Similar to Node 10, the responses *agree* and *not sure* were associated with large increase in the target variable.

On the other hand, Node 6 shows that students who replied *strongly agree* to *effectiveness* question but interestingly *agree* or *not sure* or *disagree* to *positivity question* still had a ratio of 61.9% to choose *strongly agree* and a ratio of 27% to choose *agree* to the target question.

Additionally, Node 7 presents that students who replied *agree* to *effectiveness* question and replied *strongly agree* to *enabling students to evaluate different perspectives* question had 72.1% ratio to reply *strongly agree* and 20.9% ratio to

reply *agree* to the target question. In total, these students had a ratio of 93% to provide a positive answer to the target question.

Node 8 shows that when the same group of students replied *agree* to *enabling students to evaluate different perspectives* question, they had 32.1% ratio to choose *strongly agree* to the target question. Their ratio to choose *agree* for the target question was 62.3%, making a total of 94.1% positive responses.

To sum up, it can be stated that there are four significant predictors of overall teaching quality for MC instructors. The most statistically significant one is question 5 (*effectiveness*), and then question 1 (*positivity*) and question 20 (*enabling students to evaluate different perspectives*) share similar statistical significance. Finally, question 21 (*encouragement to share views*) is the next most important predictor of overall teaching quality for MC instructors from students' point of view.

Figure 13 shows the outline of the decision tree analysis conducted for TW instructors. Figure 14 and Figure 15 demonstrate the two sections of this tree.

The Most Significant Predictors of Overall Teaching Performance for TW Instructors

Figure 13 shows the outline of the decision tree for TW instructors. There are 15 nodes except the root node. In the first level, the most significant predictor of teaching quality is *effectiveness*, and based on this question, there are four nodes. In the second level, the most significant predictors of teaching quality are *positivity* and *being supportive*. There are seven nodes in total in this level. In the third level, there are two predictors of teaching quality, which are *encouraging active participation* and *subject knowledge*. These questions produce four nodes in total in the third level. The terminal nodes which produce higher rates of positive answers are described under Figure 14 and Figure 15.

Figure 13

Outline of the Decision Tree for TW Instructors

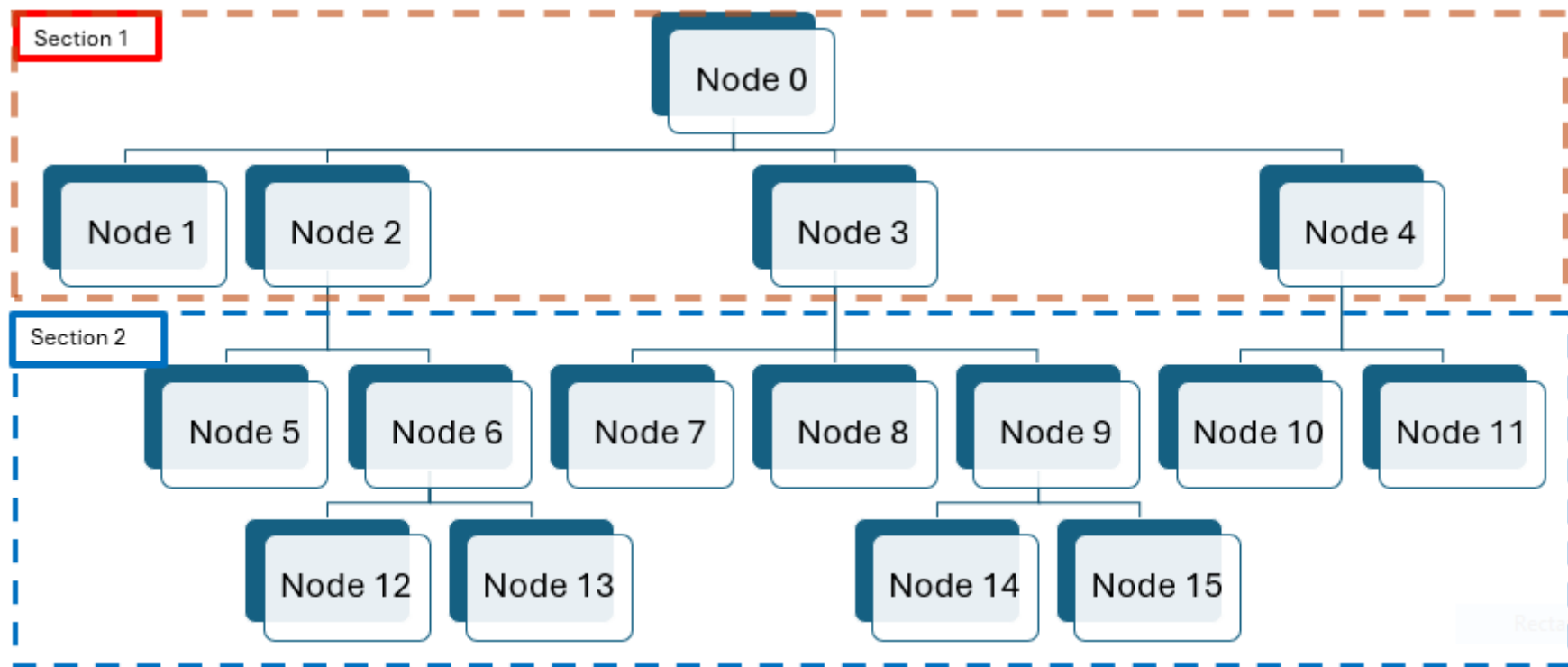


Figure 14

Section 1 of the Decision Tree for TW Instructors



Figure 14 presents the first section of the decision tree for TW instructors. Node 0 presents students' responses to the target variable. Students answered *strongly agree* to *glad to take lessons* question with a ratio of 61.7% (n = 613) and *agree* with a ratio of 16.7% (n = 166). The most significant predictor of this question is *effectiveness* question. However, under Node 0, only Node 2 shows an increase in the percentage of positive answers. Students who replied *strongly agree* to *effectiveness* question provided positive answers to the target question with a total ratio of 97.4% in total.

On the contrary, Node 1 presents the lowest positive answer scores. According to Node 1, students who opted for *strongly disagree* or *disagree* for *effectiveness* question had only a ratio of 7.1% to choose *strongly agree* for the target question and 10.6% chance to choose *agree*.

Figure 15 demonstrates the second section of the decision tree for TW instructors.

Figure 15

Section 2 of the Decision Tree for TW Instructors

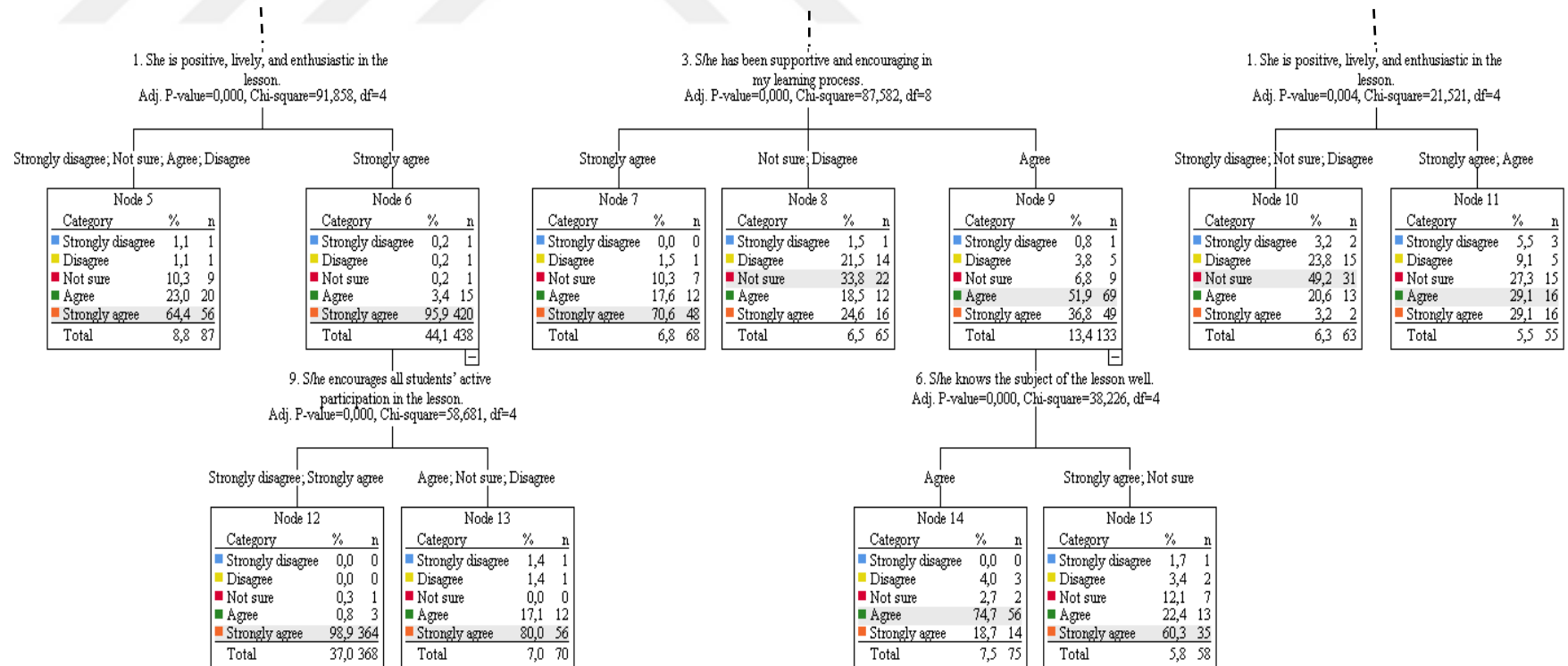


Figure 15 shows the nodes created under Node 2, Node 3, and Node 4. Under Node 2, Node 5 is a child node which shows that students who chose *strongly disagree* or *not sure* or *agree* or *disagree* for *positivity* question and *strongly agree* for *effectiveness* question had 64.4% ratio of answering *strongly agree* to the target question and a ratio of 23% to answer *agree*, which in total equals to 87.4% ratio of positive answers.

Node 12 is the first child node of Node 6. It presents that students, who chose *strongly agree* for *effectiveness* and *positivity* questions and *strongly agree* or interestingly *strongly disagree* for *encouraging active participation* question, opted for *strongly agree* for the target question with a ratio of 98.9 and *agree* with a ratio of 0.8, which in total 99.7% ratio to provide positive answers.

Node 13 is the second child node of Node 6. Node 13 shows that when the same group of students above chose *agree* or *not sure* or *disagree* for *encouraging active participation* question, they provided *strongly agree* for the target question with a ratio of 80% and *agree* with a ratio of 17.1%, which in total equals to 97.1% ratio of positive answers.

Node 7 is about the students who chose *agree* for *effectiveness* question and *strongly agree* for *being supportive* question. These students opted for *strongly agree* for the target question with a ratio of 70.6% and *agree* with a ratio of 17.6%. Combined, these scores indicate 88.2% ratio of positive answers.

Under Node 9, there are two child nodes, which are Node 14 and Node 15. Node 14 is about the students who replied *agree* for all *effectiveness* and *being supportive* and *subject knowledge* questions. These students provided *strongly agree* to the target question with a ratio of 18.7% and *agree* with a ratio of 74.7%. This is 93.4% chance of positive answers in total.

Node 15 is the second child node under Node 9. It demonstrates that when the same group of students mentioned above chose *strongly agree* or *not sure* for *subject knowledge* question, they chose *strongly agree* for the target question with a possibility of 60.3% and *agree* with a possibility of 22.4%, which in total 82.7% chance of positive answers.

To summarize, for TW instructors, the most powerful predictor of teaching quality is *effectiveness* question. Following *effectiveness* question, *positivity* and *being supportive* questions are the second most important predictors. In the third level, *encouraging active participation* and *subject knowledge* are the next most important predictors of teaching quality.

The Most Significant Predictors of Overall Teaching Performance for LS Instructors

Figure 16 presents the outline of the analysis for LS instructors. Including the root node, there are 17 nodes in total. There are five nodes in the first level, seven nodes in the second level, and four nodes in the third level. In the first level, the most significant predictor of teaching quality is *being supportive*. In the second level, only two predictors, which are *enabling students to evaluate different perspectives* and *encouragement to share views*, are found. In the third level, *effectiveness* and *preparedness* are the most important predictors.

This tree is divided into four sections. Section 1, 2, 3, and 4 are presented in Figure 17, Figure 18, Figure 19, and Figure 20 respectively.

Figure 16

Outline of the Decision Tree Produced for LS Instructors

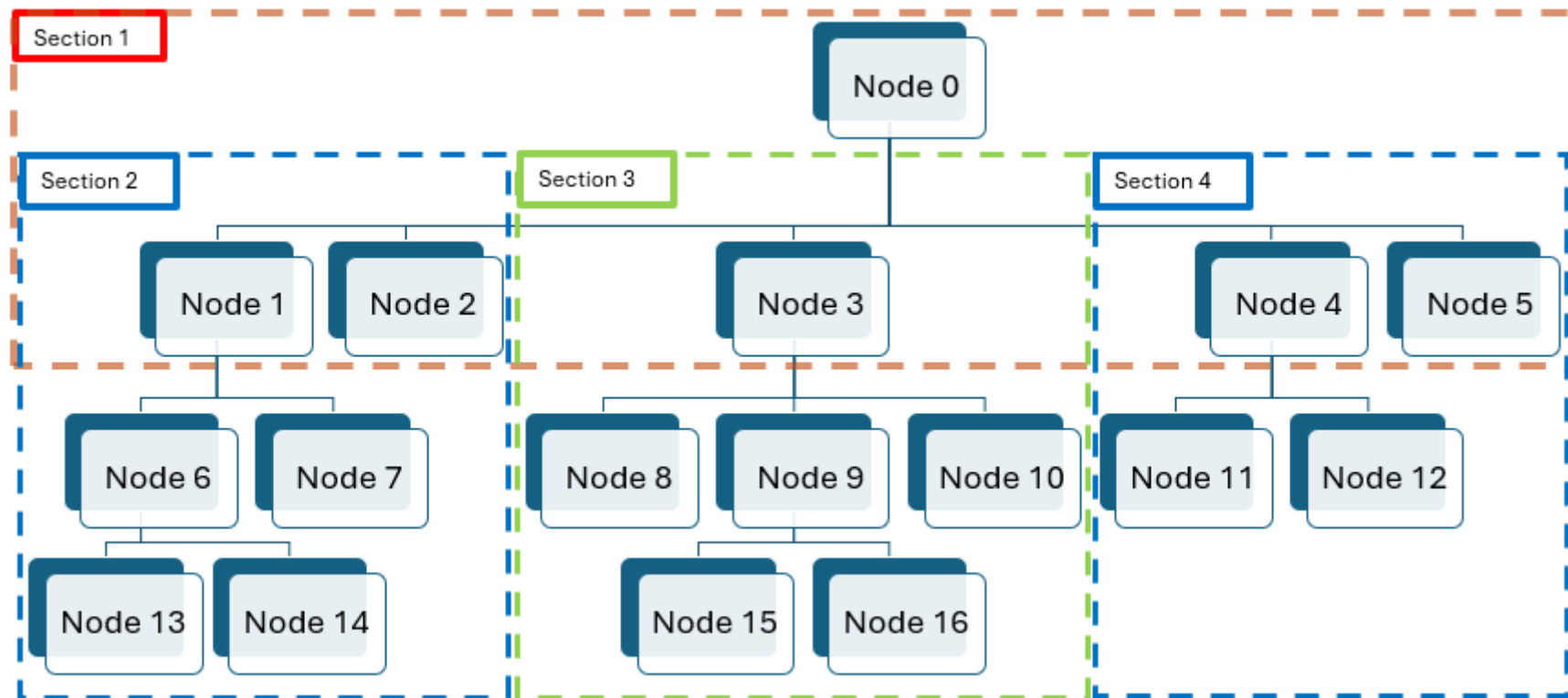
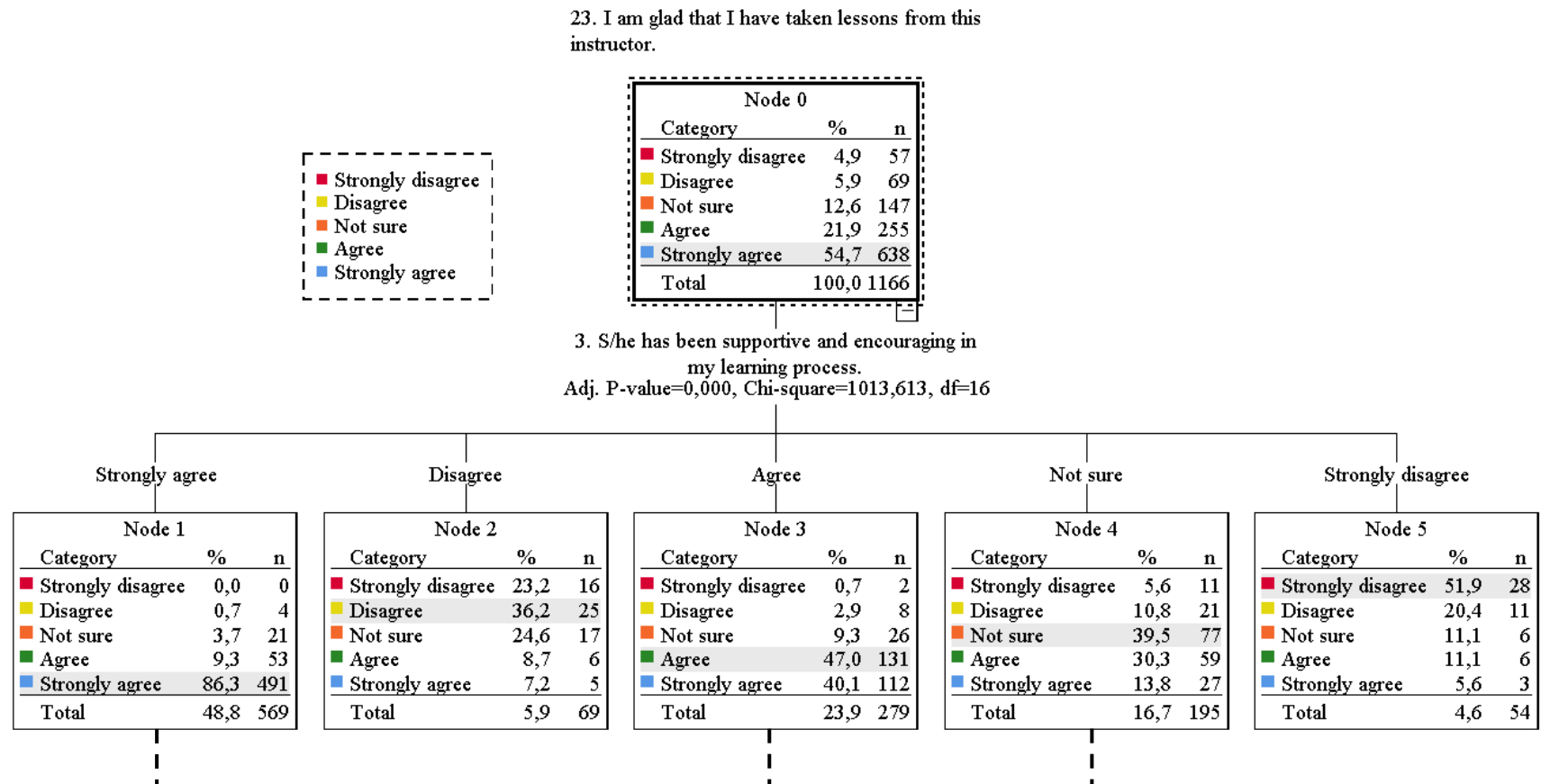


Figure 17

Section 1 of the Decision Tree for LS Instructors



In Figure 17, Node 0 shows the distribution of the students' answers for the target question for LS instructors. 54.7% (n = 638) of the students answered *strongly agree* for the target question while 21.9% (n = 255) of them answered *agree*. The most important predictor of the target question is *being supportive*, which is different from the ones found in the previous analyses. Node 1 presents the highest correlation with the target question in the first level. According to Node 1, students who replied *strongly agree* for *being supportive* had 86.3% ratio to answer *strongly agree* for the target question and a ratio of 9.3% to answer *agree*, which is in total a ratio of 95.6% to provide a positive answer.

The lowest percentages of positive answers are presented in Node 2 and Node 5. Node 2 shows that when students answered *disagree* for *being supportive*, they only had a ratio of 15.9% to provide positive answers to the target question. Similarly, Node 5 shows that students who chose *strongly disagree* for *being supportive* question only had a total of 16.7% ratio to provide positive answers for the target question.

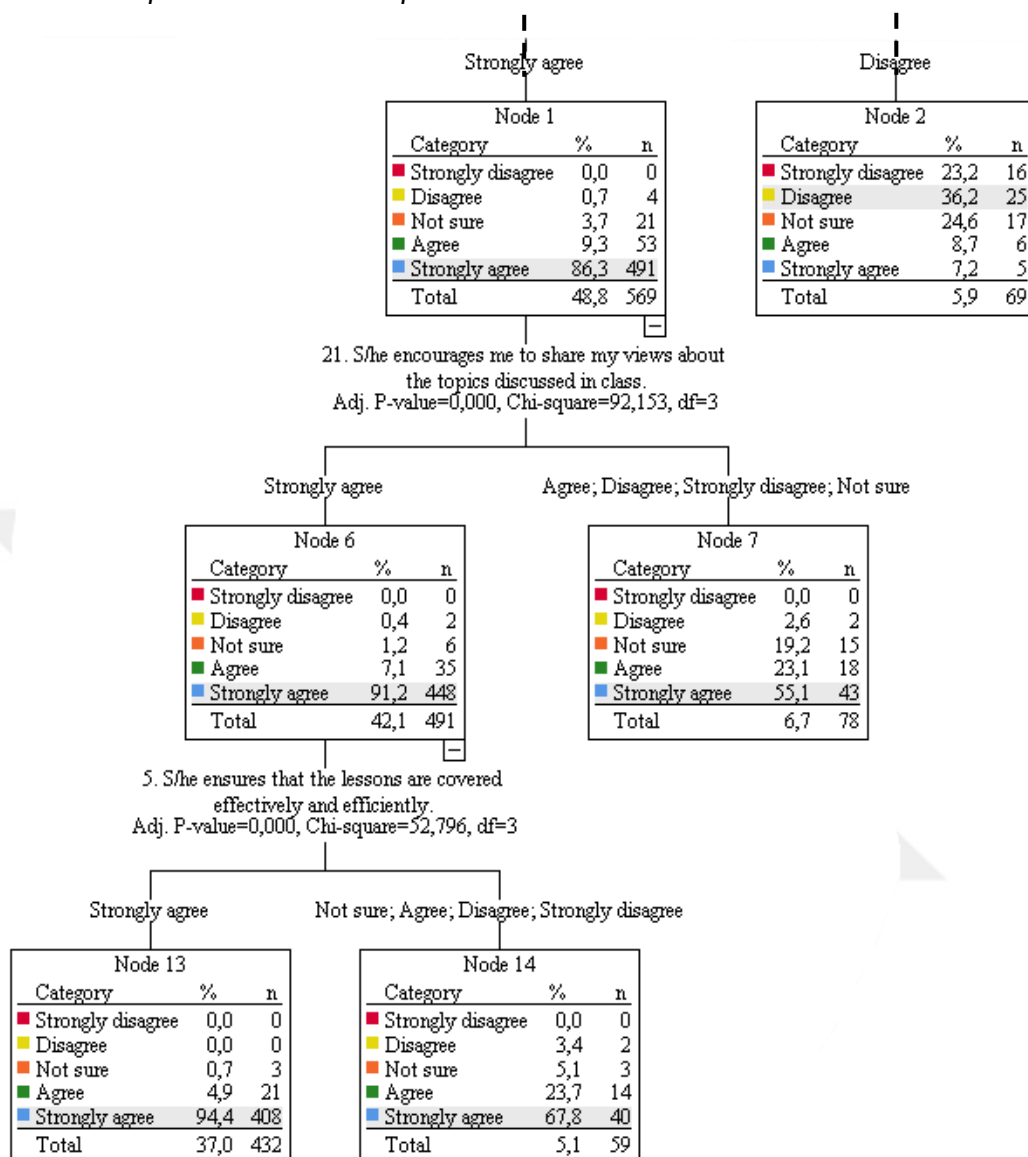
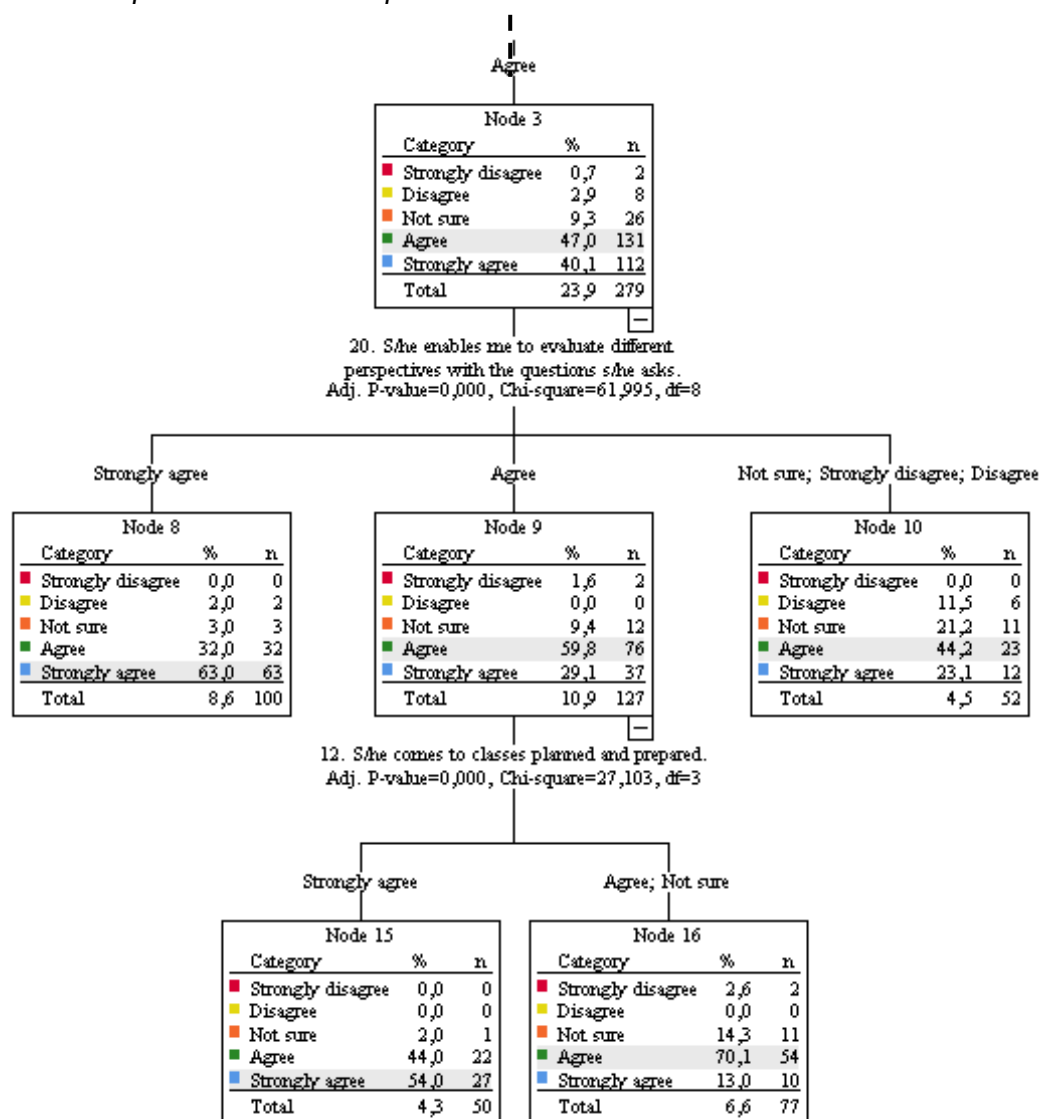
Figure 18*Section 2 of the Decision Tree for LS Instructors*

Figure 18 shows the second section of the decision tree for LS instructors. In the second level, Node 7 is a child node which produces higher percentages of positive answers than the root node itself. Node 7 is created based on *encouragement to share views*. Students who chose *strongly agree* for *being supportive* and any answer except *strongly agree* to *encouragement to share views* question still had 78.2% total ratio of providing positive answers to the target question.

Nodes 13 and 14 are created based on *effectiveness* question. Node 13 presents the idea that students who answered *strongly agree* to *being supportive*, *encouragement to share views*, and *effectiveness* questions had a ratio of 99.3% to provide positive answers to the target question. However, it is found that if the same group of students chose any answer except *strongly agree* for *effectiveness* question, their ratio to choose a positive answer to the target question dropped to 91.5%, as can be seen in Node 14.

Figure 19

Section 3 of the Decision Tree for LS Instructors



In Figure 19, nodes created under Node 3 are presented. Node 8 reveals that students who opted for *strongly agree* for *enabling students to evaluate different perspectives* question and *agree* for *being supportive* had a 63% ratio to choose *strongly agree* for the target question and a 32% ratio to choose *agree*, which means 95% positive answers.

Nodes 15 and 16 are created under Node 9. Node 15 shows that students, who replied *agree* to *being supportive* and *enabling students to evaluate different perspectives* questions and *strongly agree* for *preparedness* question, replied *strongly agree* to the target question with a ratio of 54% and *agree* with a ratio of 44%, which means 98% ratio of providing positive answers in total. Node 16 presents that when the same group of students chose *agree* or *not sure* for *preparedness* question, their ratio to provide positive answers to the target question dropped to 83.1% from 98%, which is a drastic change.

Figure 20

Section 4 of the Decision Tree for LS Instructors

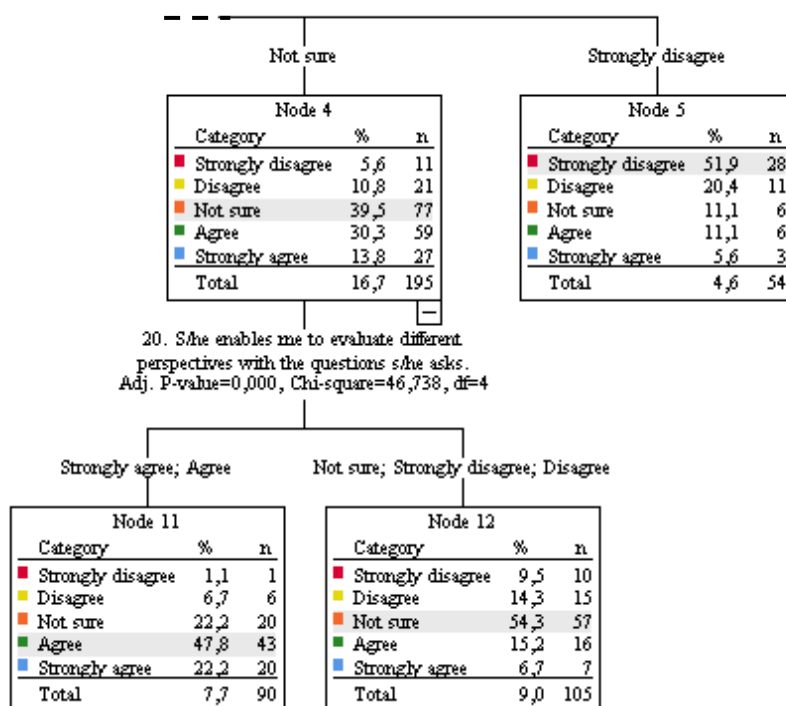


Figure 20 shows the last section of the decision tree for LS instructors.

However, as none of the nodes produce higher percentages of positive answers than the root node, they are not detailed here.

The Most Significant Predictors of Overall Teaching Performance Depending on Students' Proficiency Levels

In this section, the results of the analyses based on students' proficiency levels will be shared. These analyses include beginner level (AF), elementary level (A), intermediate level (B), and upper-intermediate (C) level instructors. Repeat A, repeat B and repeat C classes are not included in this part as their repeat status might affect the way they evaluate their instructors, and this situation might interfere with the results of this analysis. In the analyses for whole data and for different course types, the parent nodes were determined to be 100, and child nodes were 50 as the sample sizes were larger. However, in this part, the sample sizes are relatively smaller; therefore, child nodes are reduced to 25 so that a more detailed analysis can be conducted.

The Most Significant Predictors of Overall Teaching Performance for AF Level Instructors

Figure 21 presents the outline of the decision tree for AF level instructors. Except the root node, there are 10 nodes in total. The most significant predictor of teaching quality in the first level is *effectiveness*. There are four nodes created based on *effectiveness*. In the second level, *being supportive* and *feedback* are the second most significant predictors, and four more nodes are created based on them. In the third level, *encouragement to share views* is the next most significant predictor. This analysis is divided into two sections, and the first section is demonstrated in Figure 22.

Figure 21

Outline of the Decision Tree Produced for AF Level Instructors

Outline of the Decision Tree Produced for AF Level Instructors

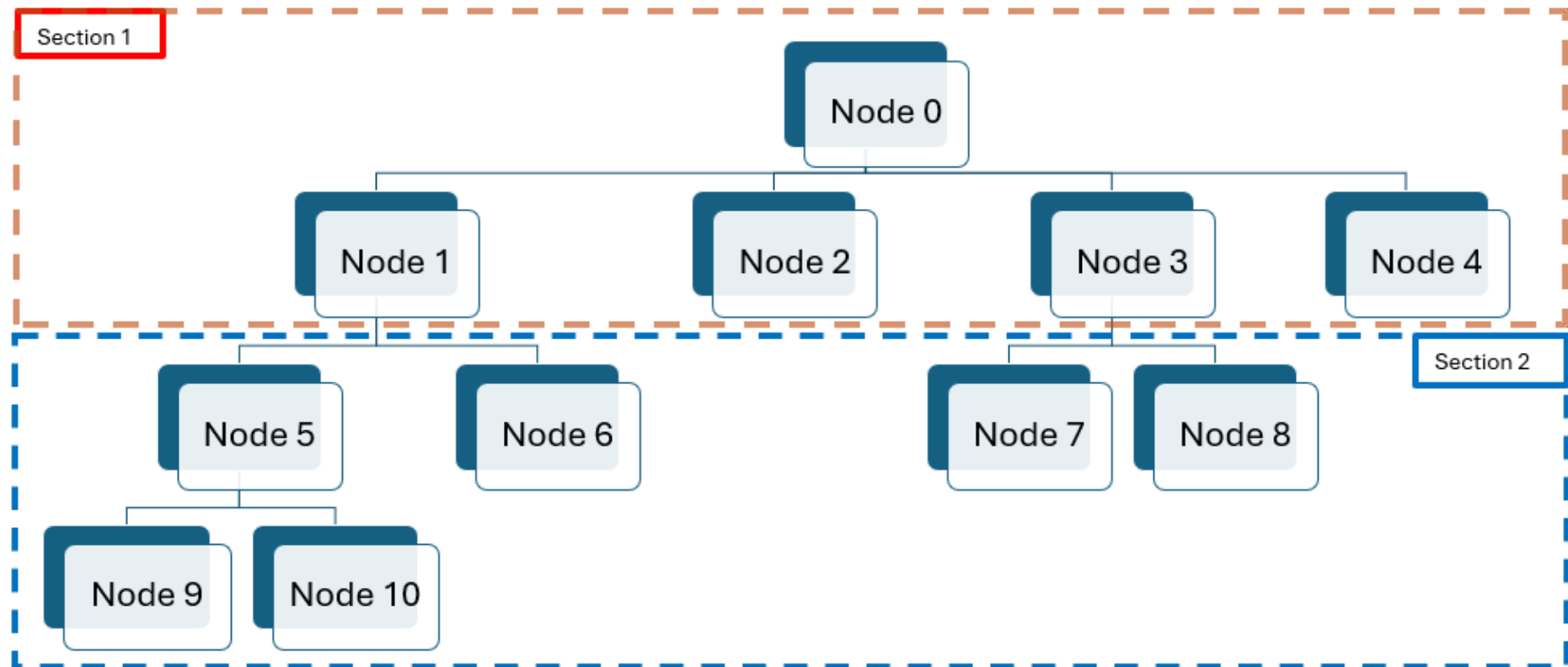


Figure 22

Section 1 of the Decision Tree for AF Level Instructors



Figure 22 shows the first section of the decision tree for AF level instructors. Node 0 exhibits that 61.8% (n = 286) of AF level students chose *strongly agree* for *glad to take lessons* question while only 17.5% (n = 81) of them chose *agree*. In the first level, Node 1 shows that students who answered *strongly agree* to *effectiveness* question had a total ratio of 99.2% to provide positive answers to the target question while this rate dropped to 77.7% when students answered *agree* to *effectiveness* question as can be seen in Node 3.

Node 1 and Node 3 produces additional nodes which are explained under Figure 23.



Figure 23

Section 2 of the Decision Tree for AF Level Instructors

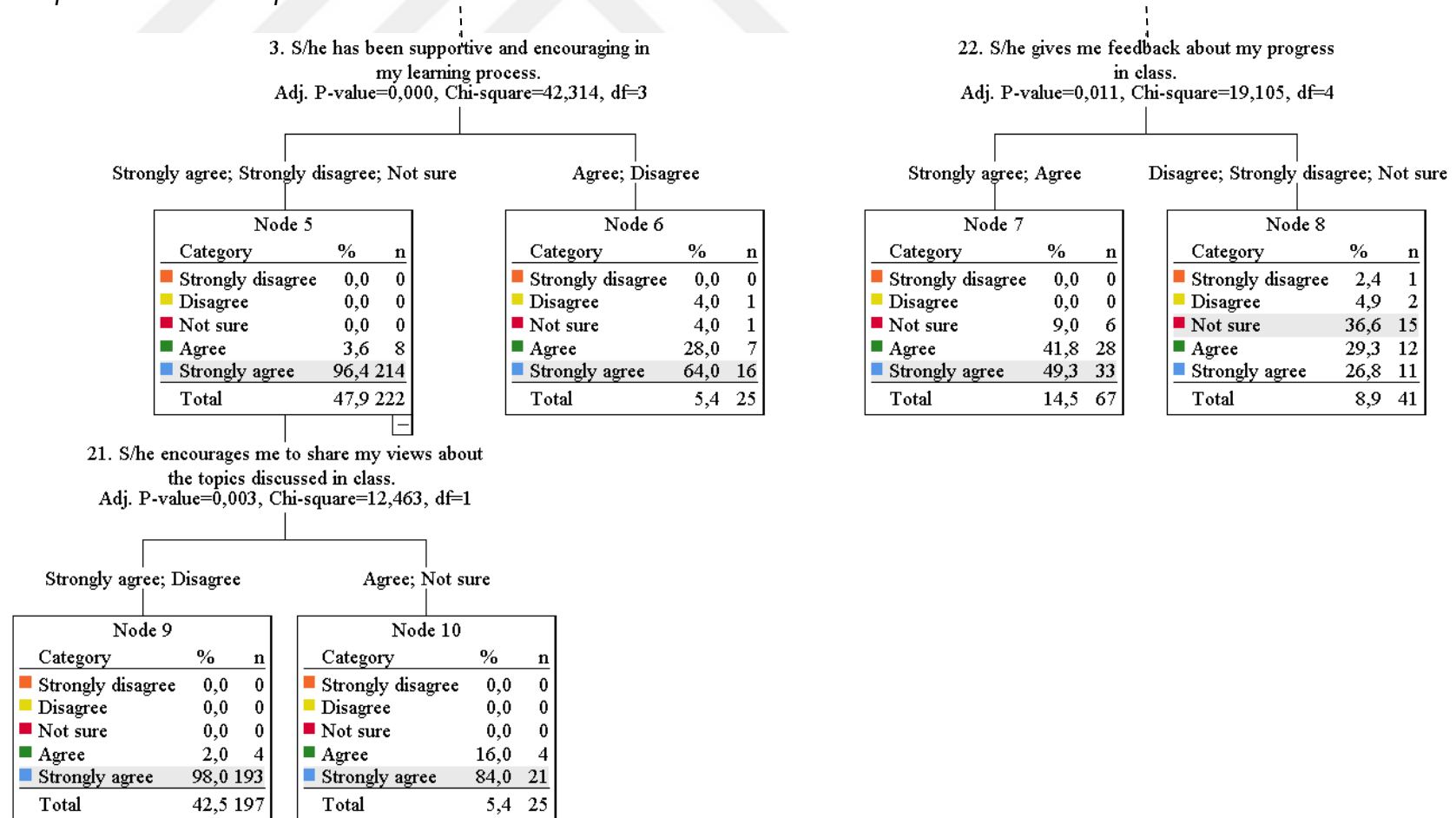


Figure 23 shows the second section of the decision tree for AF level instructors. Node 6 demonstrates that students who replied *strongly agree* for *effectiveness* question and *agree* or *disagree* for *being supportive* question still had a 64% ratio to reply *strongly agree* for the target question and a ratio of 28% to reply *agree*.

Node 5 produces two child nodes, which are Node 9 and Node 10. Node 9 shows that students who chose *strongly agree* for *effectiveness* and *strongly agree* or *strongly disagree* or *not sure* for *being supportive* and *strongly agree* or *disagree* for *encouragement to share views* chose *strongly agree* for the target question with a ratio of 98% and *agree* with a ratio of 2%, which is 100% chance of positive answers.

Similarly, Node 10 also shows that when the same group of students answered *agree* or *not sure* for *encouragement to share views* question, their ratio of answering *strongly agree* for the target question dropped to 84% from 98%, and their ratio of answering *agree* increased to 16% from 2%, which is in total 100% chance of positive answers again.

Moreover, Node 3 produces two child nodes, which are Node 7 and Node 8. Node 7 demonstrates that 49.3% of students, who chose *agree* for *effectiveness* and *strongly agree* or *agree* for *feedback*, chose *strongly agree* for the target question again. 41.8% of the same group chose *agree* for the target question.

Node 8 and Node 4 are not detailed here as they cannot produce higher percentages of positive answers than the root node.

The Most Significant Predictors of Overall Teaching Performance for A Level Instructors

Figure 24 shows the outline of the decision tree for A level instructors. Except the root node, there are 8 nodes in total. The most significant predictor of the target variable is *being supportive* in the first level, and there are four nodes created based on *being supportive*. In the second level, the most significant predictor is *encouragement to share views*, which produces two more nodes. In the third level, the most significant predictor is *variety of activities*, and two more nodes are created based on that.

Figure 25 demonstrates the first section of the decision tree for A level instructors. Node 0 shows that 62.2% of all A level students ($n = 240$) chose *strongly agree* and 17.6% of them ($n = 68$) chose *agree* for the target question.

Being supportive is the most significant predictor of teaching quality in the first level, which is a different result from those of previous analyses.

Node 1 presents that 99% of all students who answered *strongly agree* to *being supportive* question also gave a positive answer to the target question (94.3% *strongly agree*, 4.7% *agree*).

Node 2 shows that 85.1% of all students who chose *agree* for *being supportive* also gave a positive answer to the target question (40.7% *strongly agree*, 44.4% *agree*).

Node 3 and Node 4 cannot produce higher positive answer percentages than the root node; therefore, they are not detailed here.

Figure 24

Outline of the Decision Tree Produced for A Level Instructors

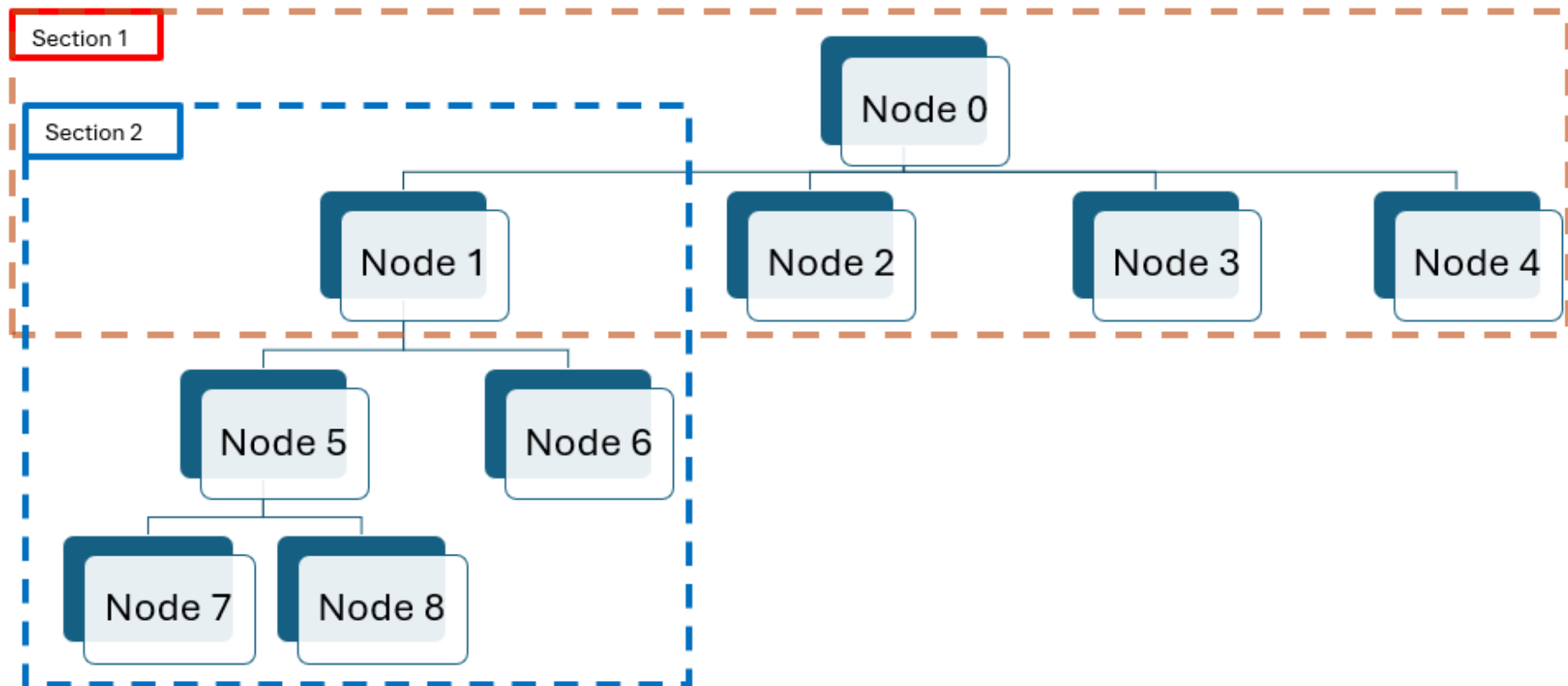


Figure 25

Section 1 of the Decision Tree for A Level Instructors

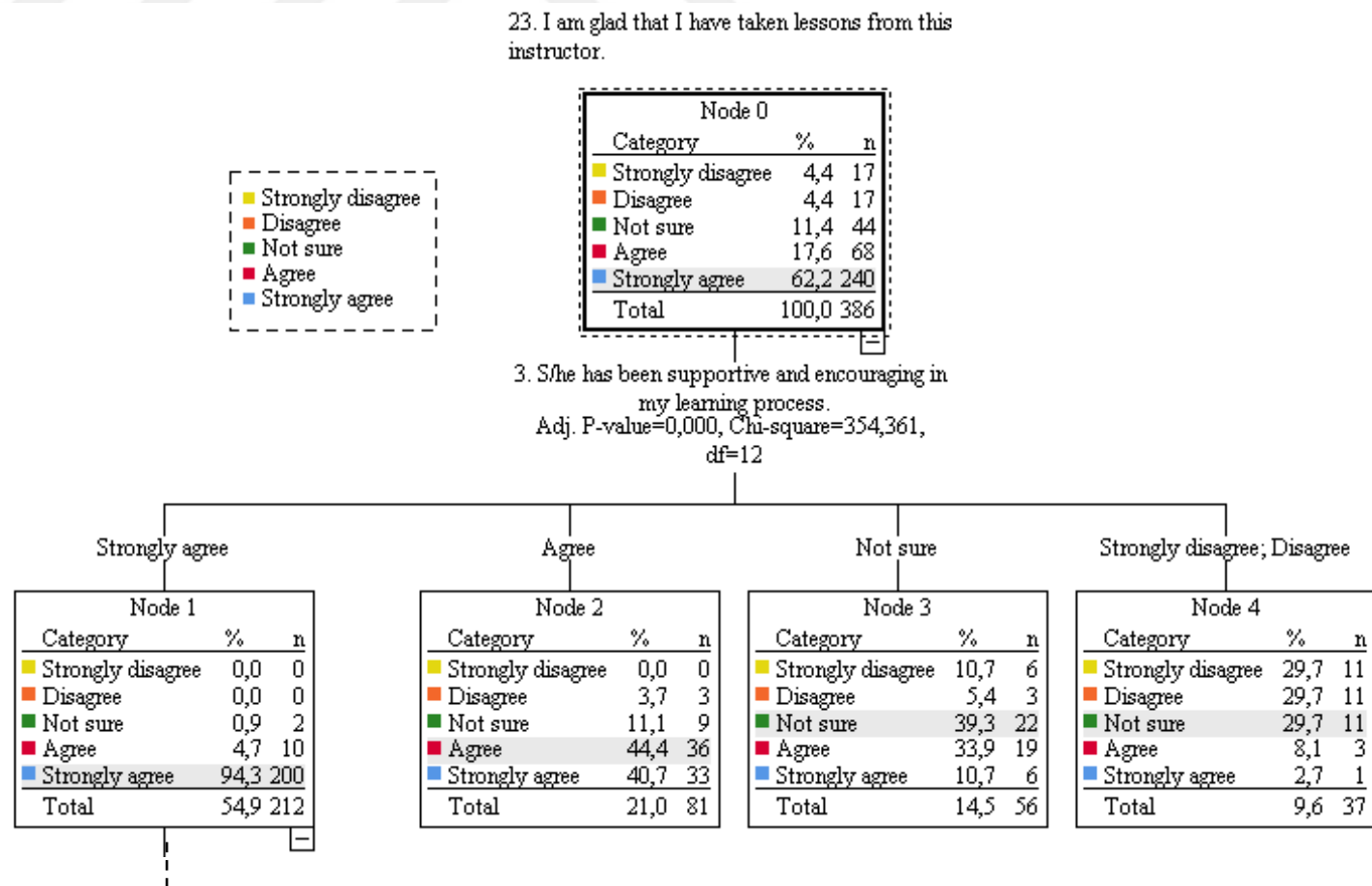


Figure 26

Section 2 of the Decision Tree for A Level Instructors

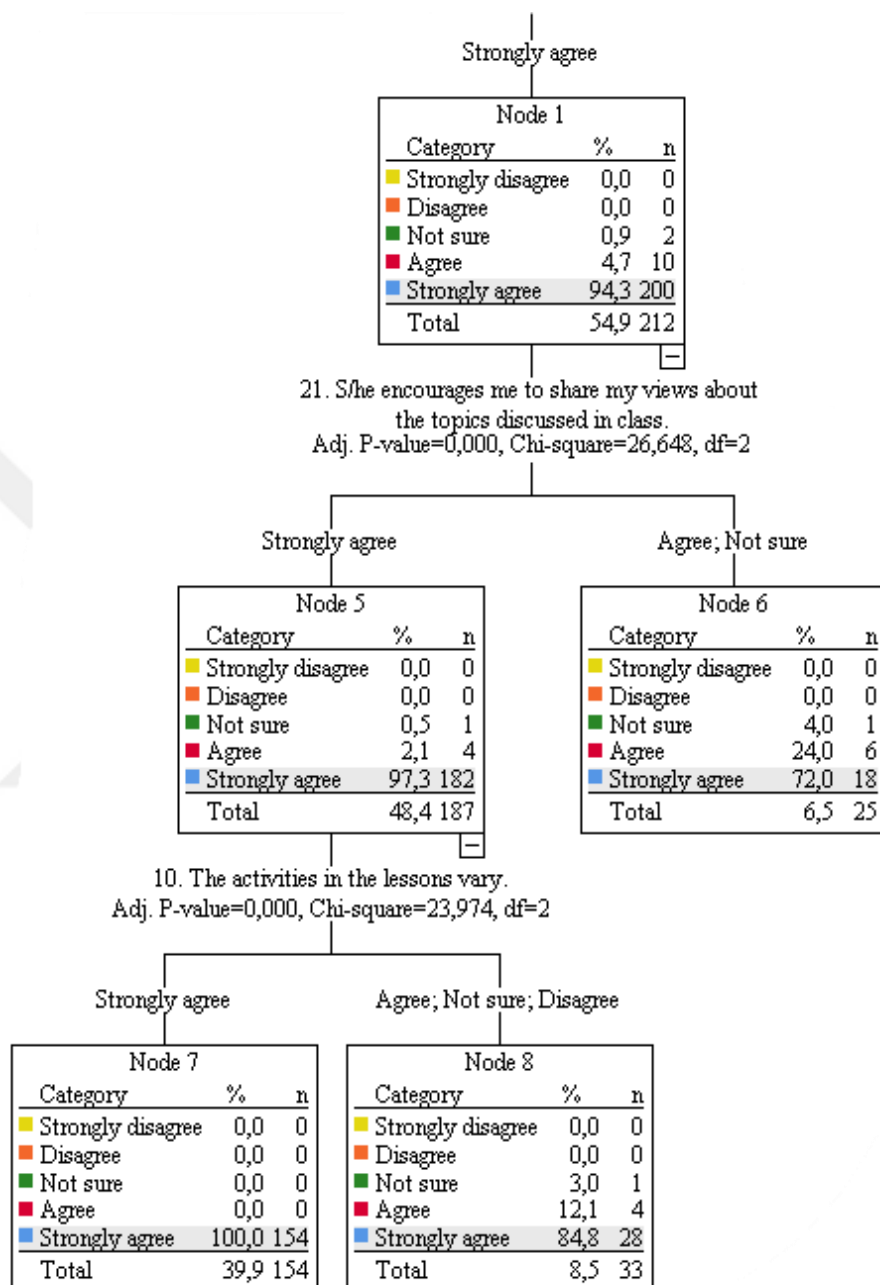


Figure 26 shows the second section of the decision tree for A level instructors. Node 1 produces two child nodes, which are Node 5 and Node 6.

Node 6 points out that students who replied *strongly agree* for *being supportive* and *agree* or *not sure* for *encouragement to share views* had a ratio of 72% to choose *strongly agree* and 24% to choose *agree* for the target question, which is 96% positive answers in total.

Node 5 produces two child nodes, which are Node 7 and Node 8. Node 7 shows that 100% of all the students who replied *strongly agree* for *being supportive* and *encouragement to share views* and *variety of activities* also replied *strongly agree* for the target question.

Node 8 demonstrates that when the same group answered *agree* or *not sure* or *disagree* for *variety of activities*, their ratio of answering *strongly agree* for the target question dropped to 84.8%, and their ratio of answering *agree* was 12.1%.

The Most Significant Predictors of Overall Teaching Performance for B Level Instructors

Figure 27 demonstrates the outline of the decision tree for B level instructors. Except the root node, 19 nodes are available in total. The most significant predictor of overall teaching quality is *effectiveness* in the first level. Then, in the second level, *feedback*, *encouragement to share views*, and *enabling students to evaluate different perspectives* are the most important predictors. In the third level, again *enabling students to evaluate different perspectives* and *recommending publications in English* are the most important predictors of overall teaching quality.

Figure 27

Outline of the Decision Tree Produced for B Level Instructors

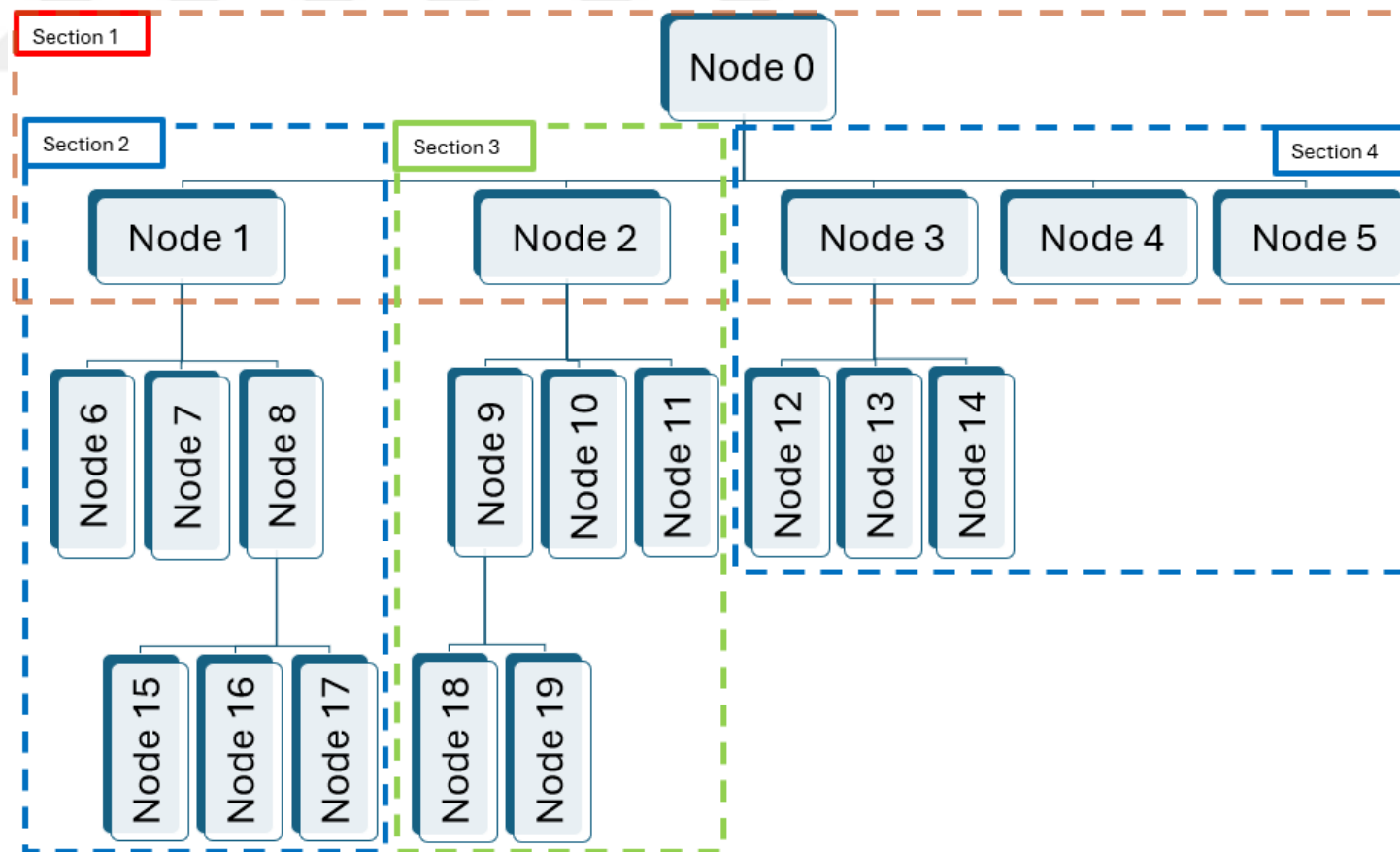


Figure 28

Section 1 of the Decision Tree for B Level Instructors



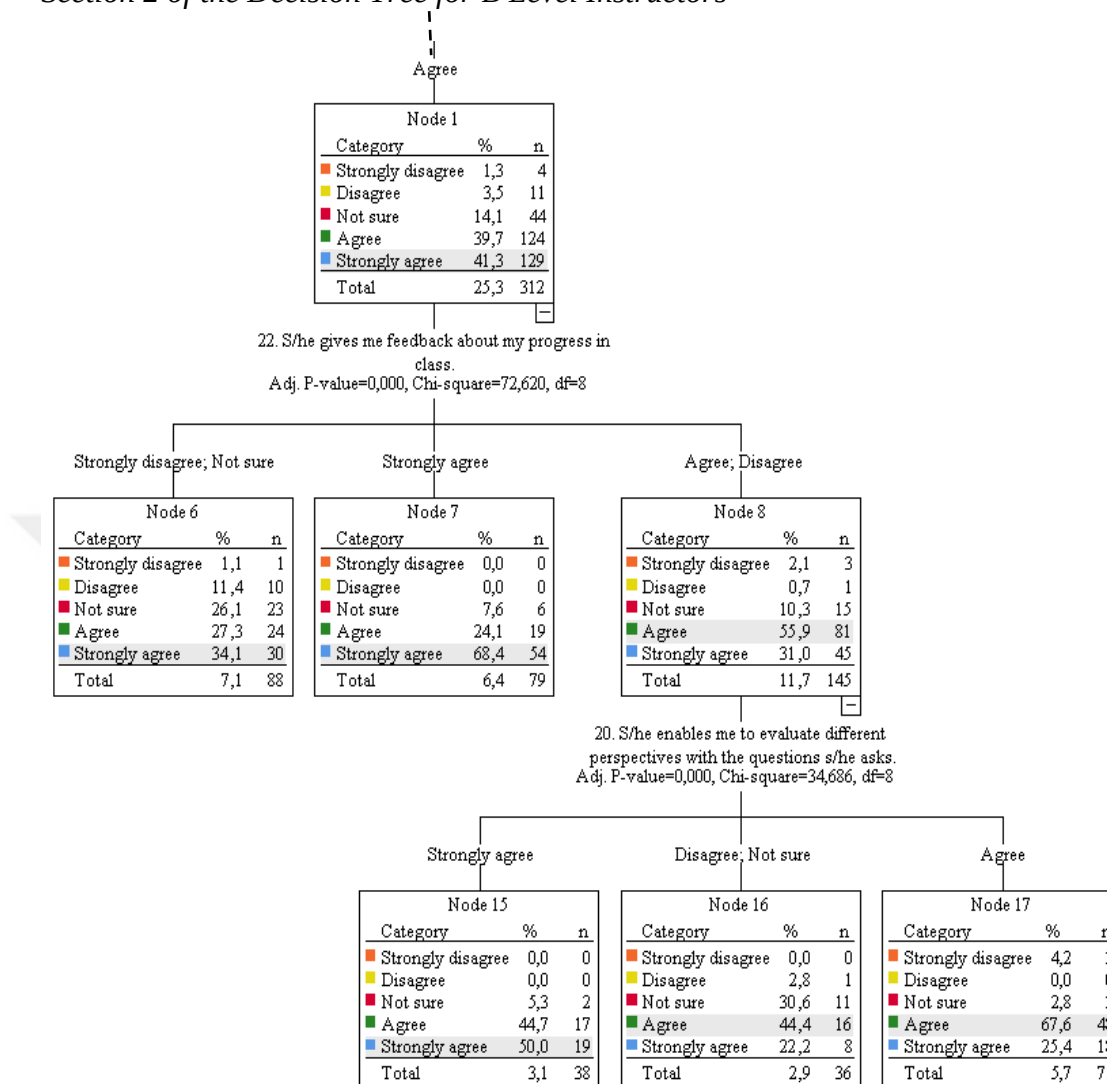
Figure 28 shows the first section of the decision tree for B level instructors. Node 0 demonstrates that of all B level students, 60.6% ($n = 749$) chose *strongly agree* for the target question and 18.9% ($n = 233$) chose *agree*.

The most significant predictor in the first level is *effectiveness*, based on which five nodes are created (Nodes 1, 2, 3, 4, and 5).

Node 1 shows that students who replied *agree* to *effectiveness* had a total positive answer ratio of 81% to the target question. Node 1 produces three child nodes which are Node 6, Node 7, and Node 8, but they are detailed under Figure 29.

Node 2 presents the fact that totally 96.8% of all students, who replied *strongly agree* to *effectiveness*, provided positive answers to the target question. Node 2 produces three child nodes, which are Node 9, Node 10, and Node 11. These nodes are detailed under Figure 30.

Nodes 3, 4, and 5 are not presented here since they cannot produce higher percentages of positive answers than the root node.

Figure 29*Section 2 of the Decision Tree for B Level Instructors*

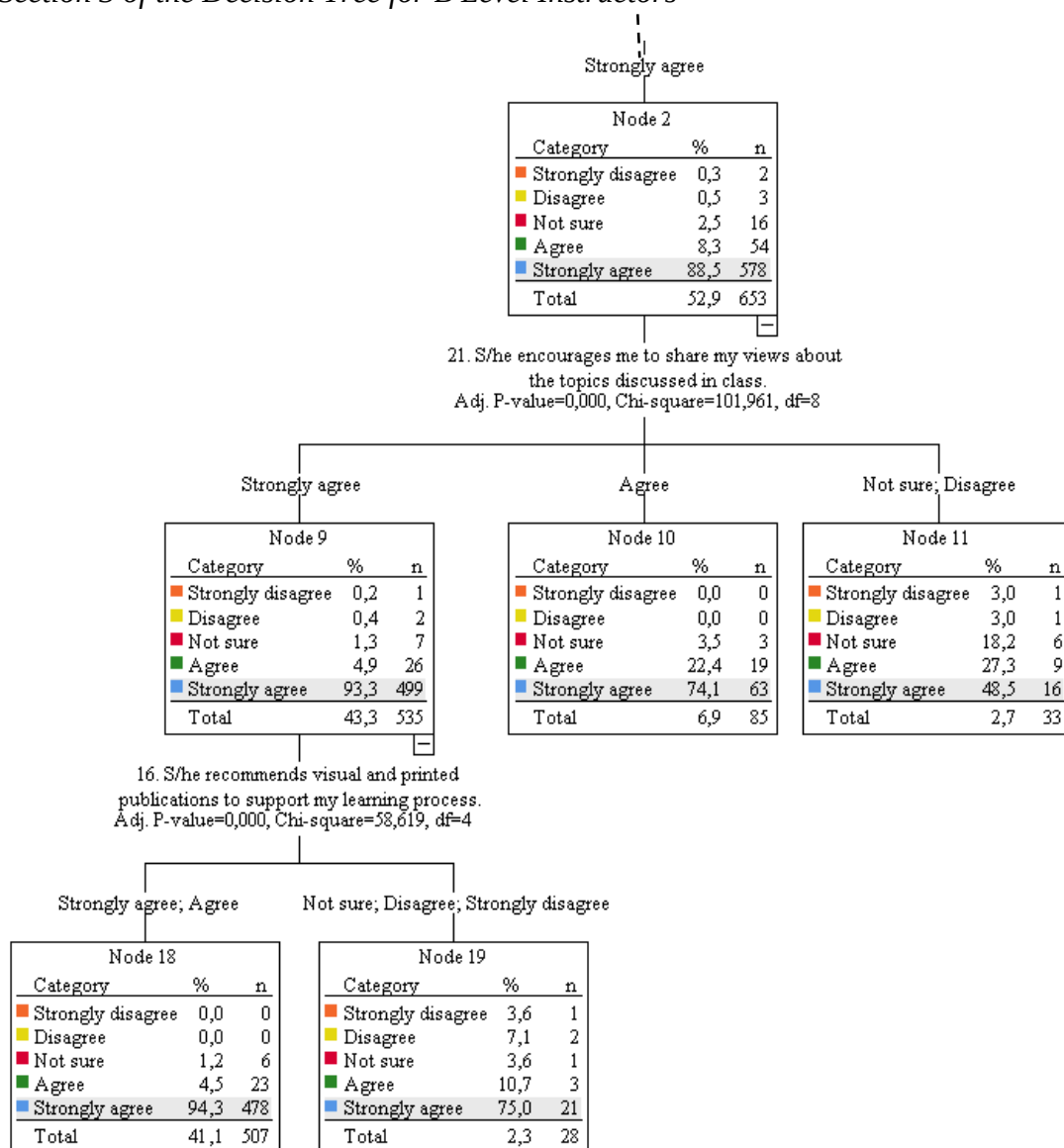
In Figure 29, Node 1 represents the group of students who answered *agree* to *effectiveness*. Node 1 produces Nodes 6, 7, and 8. Because Node 6 cannot produce higher ratio of positive answers than the root node, it is not detailed here.

Node 7 is a terminal node which shows that students, who answered *agree* to *effectiveness* and *strongly agree* to *feedback*, answered *strongly agree* with a ratio of 68.4% and *agree* with a ratio of 24.1% to the target question. This means 92.5% positive answer ratio.

Node 8 creates three child nodes, which are Nodes 15, 16, and 17. Node 15 shows that when students answered *agree* to *effectiveness* and *agree* or *disagree* to *feedback* and *strongly agree* to *enabling students to evaluate different perspectives*, they also answered *strongly agree* to the target question with a ratio of 50% and *agree* with a ratio of 44.7%. In total, it means 94.7% positive answers.

Node 17 shows that when the same group of students opted for *agree* for *enabling students to evaluate different perspectives*, they chose *strongly agree* for the target question with a ratio of 25.4% and *agree* with a ratio of 67.6%, which is 93% positive answers in total.

Node 16 is not detailed here because it cannot produce higher percentages of positive answers than the root node.

Figure 30*Section 3 of the Decision Tree for B Level Instructors*

In Figure 30, Node 2 represents the group of students who answered *strongly agree* for *effectiveness*. It creates three child nodes, which are Nodes 9, 10, and 11.

Node 9 also creates two more child nodes which are Nodes 18 and 19. Node 18 indicates that 94.3% of all the students who replied *strongly agree* to *effectiveness* and *encouragement to share views* questions and *strongly agree* or *agree* to *recommending publications in English* question also answered *strongly agree* to the target question. 4.5% of the same group answered *agree* to the target question. This is 98.8% positive answer ratio.

Node 19 shows that when the same group above changed their answers for *recommending publications in English* question and chose *not sure* or *disagree* or *strongly disagree*, their ratio of answering *strongly agree* to the target question dropped to 75% from 94.3%. Their ratio of answering *agree* was 10.7%. In total, it is 85.7% positive answer ratio.

Node 10 demonstrates that when students answered *strongly agree* to *effectiveness* and *agree* to *encouragement to share views*, their ratio of answering *strongly agree* to the target question was 74.1% and *agree* ratio was 22.4%. In total, there is 96.5% positive answer rate.

Figure 31

Section 4 of the Decision Tree for B Level Instructors

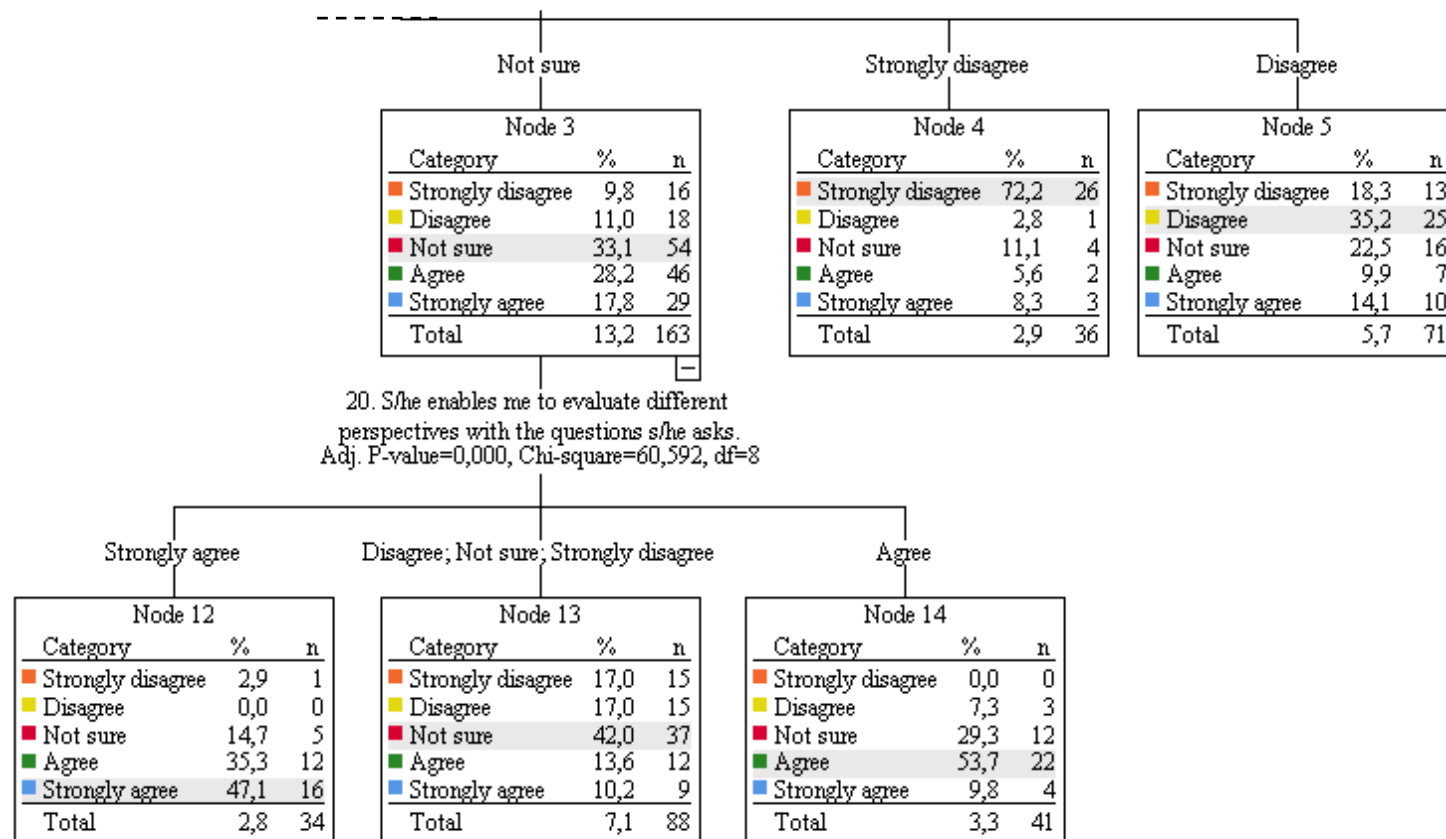


Figure 31 presents the fourth section of the decision tree for B level instructors. Node 3, Node 4, Node 5, Node 13, and Node 14 cannot produce higher positive answer percentages than the root node; therefore, they are not explained here.

Only Node 12 can produce higher positive answer ratio. Node 12 shows that 82.4% of all students, who replied *not sure* for *effectiveness* and *strongly agree* for *enabling students to evaluate different perspectives*, provided positive answers to the target question (47.1% *strongly agree*, 35.3% *agree*).

As a final remark, Node 4 produces the lowest positive answer ratio in B level analysis. Node 4 shows that only 13.9% of all students who chose *strongly disagree* for *effectiveness* gave positive answers to the target question (8.3% *strongly agree*, 5.6% *agree*).

The Most Significant Predictors of Overall Teaching Performance for C Level Instructors

Figure 32 shows the outline of the analysis created for C level instructors. Besides the root node, there are 18 nodes in total. In the first level, the most significant predictor of overall teaching quality is *being supportive*. Five nodes are created based on this variable. In the second level, *effectiveness* and *enabling students to evaluate different perspectives* are the two most important predictors. In the third level, *asking questions to encourage students to express opinions* and *variety of activities* are the final most significant predictors of the target variable.

Figure 33 demonstrates the first section of the decision tree for C level instructors.

Figure 32

Outline of the Decision Tree Produced for C Level Instructors

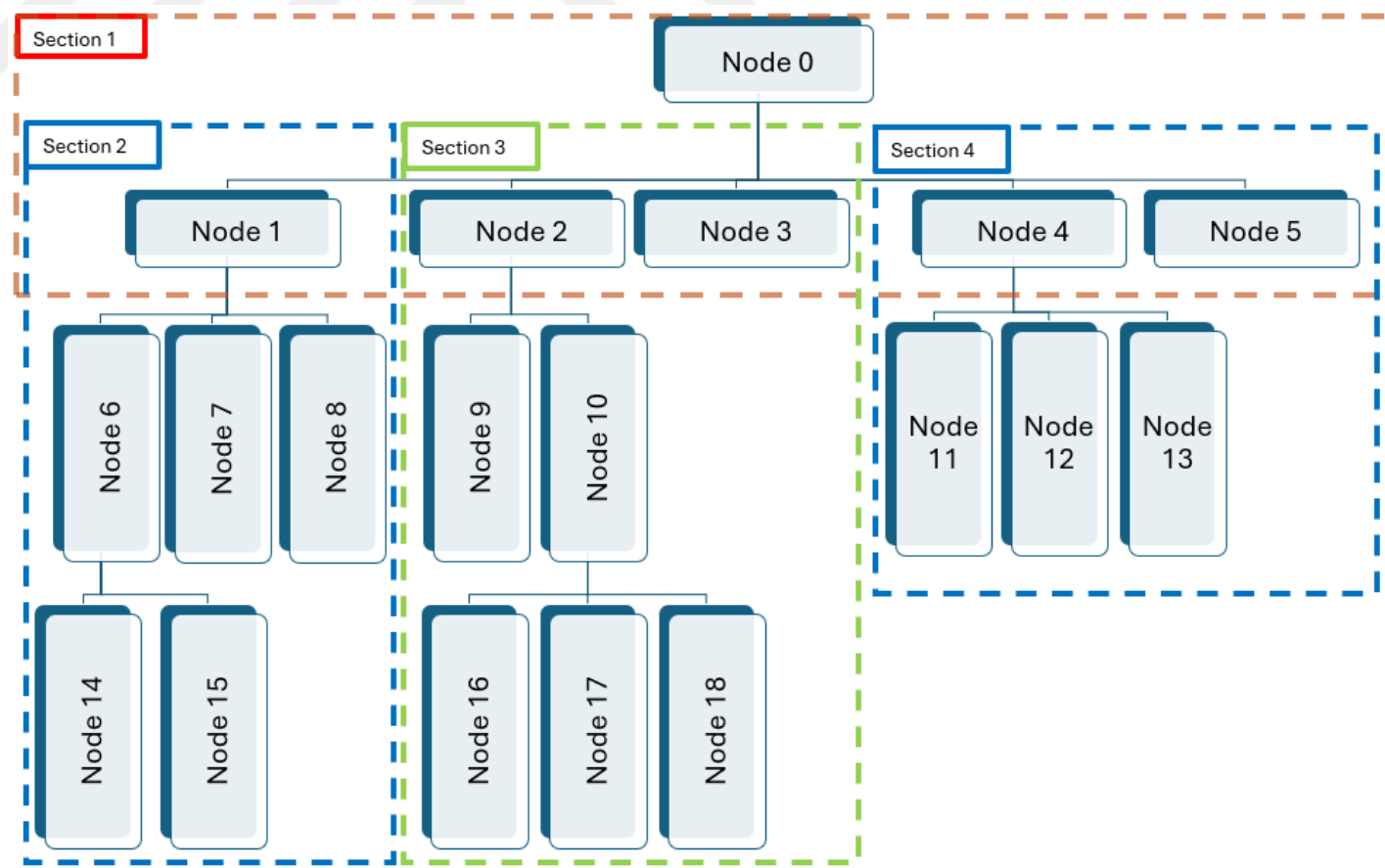


Figure 33

Section 1 of the Decision Tree for C Level Instructors



Figure 33 shows the first section of the decision tree for C level instructors. Node 0 shows that there are 1449 students in C level. 67.2% of them ($n = 974$) answered *strongly agree* for the target question, and 15.8% ($n = 229$) answered *agree*. In the first level, there are five nodes created based on *being supportive*. Only Node 2 can produce a higher ratio of positive answers than the root node. Node 2 shows that 92.4% ($n = 829$) of all students who replied *strongly agree* for *being supportive* also replied *strongly agree* for the target question as well. 5.6% of them ($n = 50$) answered *agree* for the target question. In short, there is 98% positive answer ratio.

Figure 34 demonstrates the second section of the decision tree for C level instructors. Node 1 produces Nodes 6, 7, and 8. Among these nodes, only Node 7, which is a terminal node, has a higher ratio of positive answers than the root node. Node 7 presents that students who answered *agree* for *being supportive* and *strongly agree* for *effectiveness* had a ratio of 73.9% to answer *strongly agree* for the target question and 20.3% ratio of answering *agree*.

Node 6 also produces two child nodes, which are Nodes 14 and 15. Node 14 indicates that 16.8% of all students, who replied *agree* for *being supportive* and *effectiveness* questions and *agree*, or *not sure*, or *disagree*, or *strongly disagree* for *asking questions to encourage students to express opinions*, replied *agree* for the target question. 68.8% of them responded *agree* to the target question. In total, there is a ratio of 85.6% for positive answers.

Node 15 indicates that when the same group of students changed their answers for *asking questions to encourage students to express opinions* and replied *strongly agree*, their positive answer ratio to the target question rose to 92.6% in total (72.2% *strongly agree*, 20.4% *agree*).

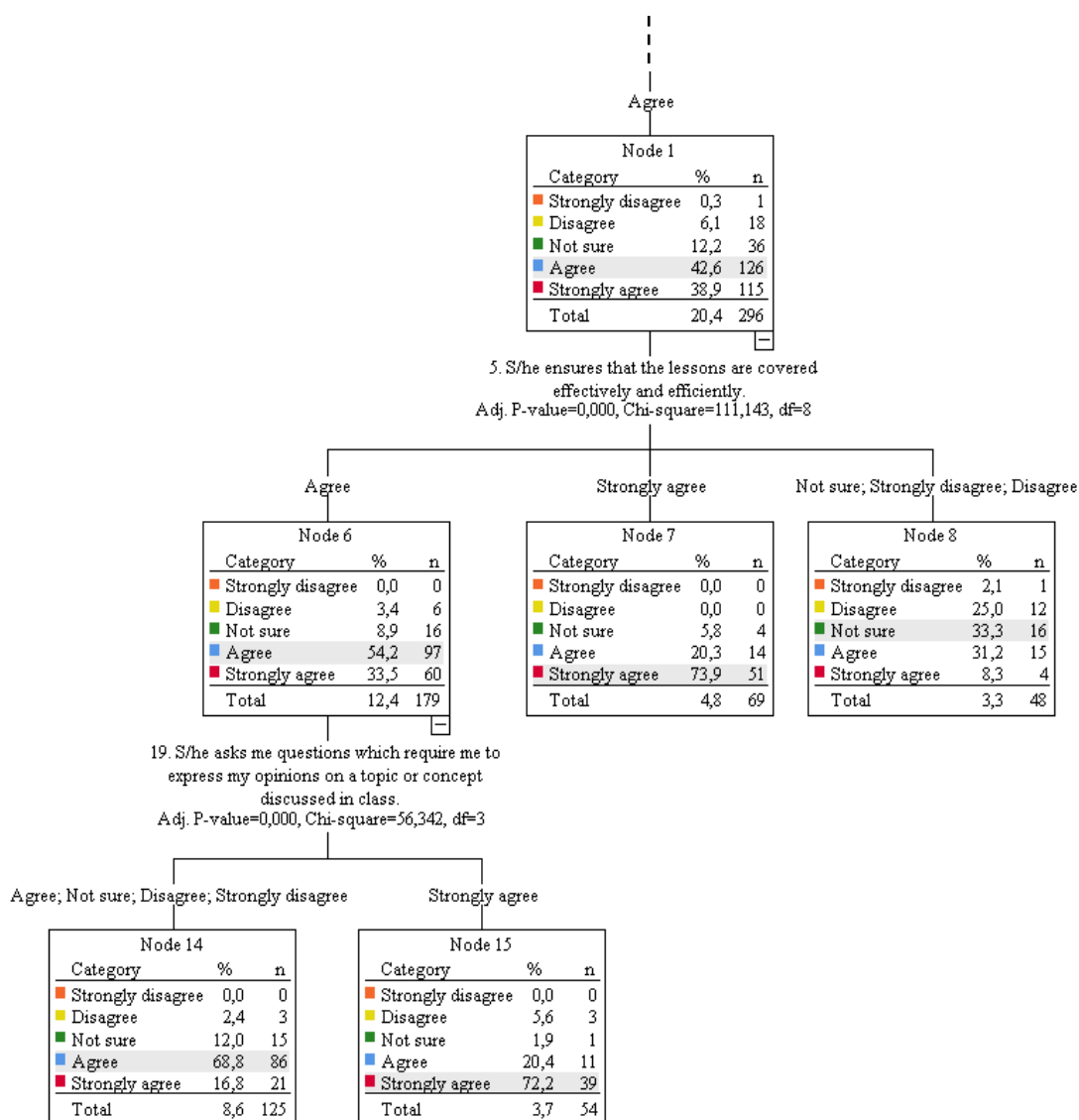
Figure 34*Section 2 of the Decision Tree for C Level Instructors*

Figure 35

Section 3 of the Decision Tree for C Level Instructors

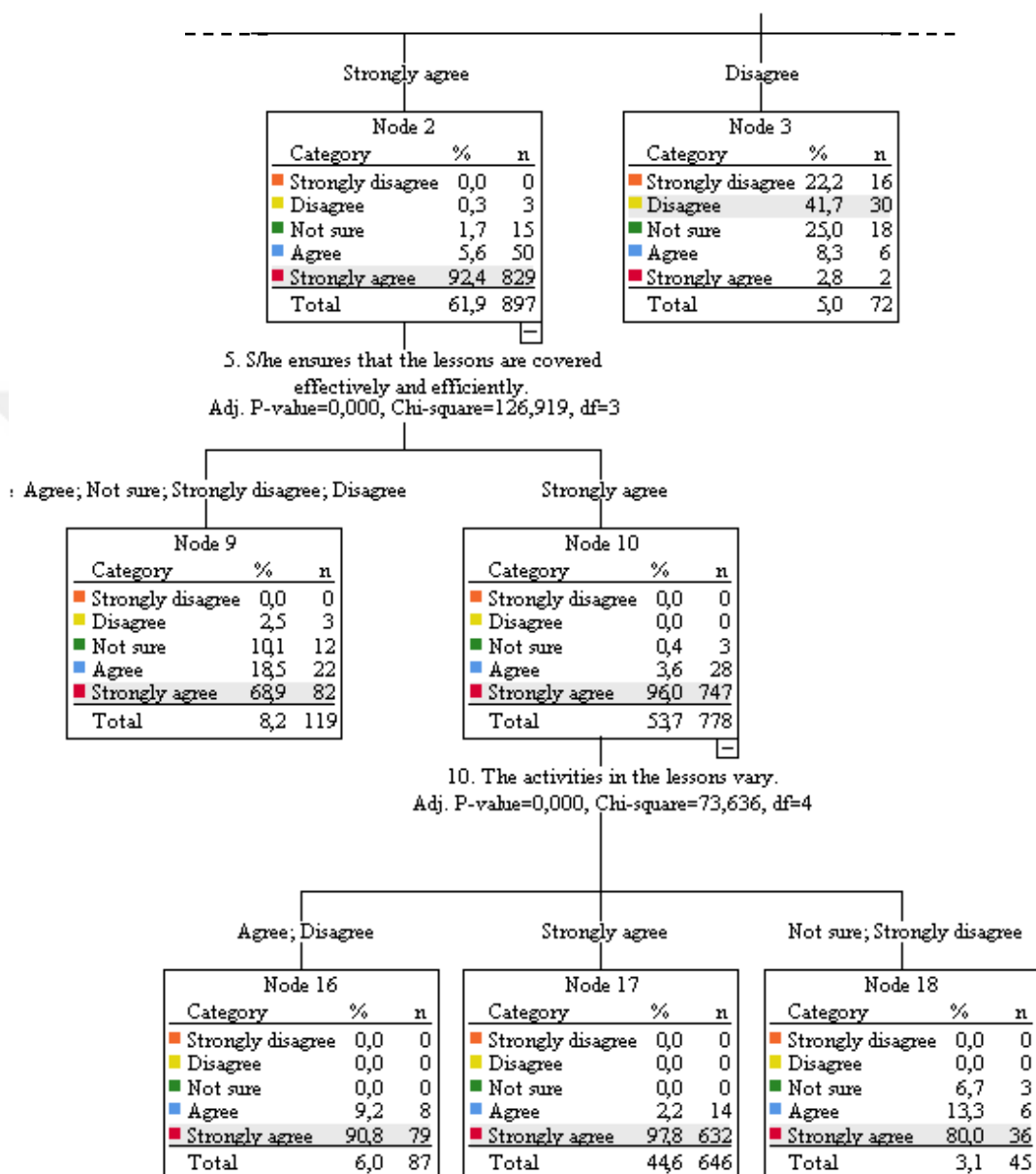


Figure 35 demonstrates the third section of the decision tree for C level instructors. Node 3 has a lower ratio of positive answers than the root node, so it is not detailed here. Node 2 produces two child nodes, which are Node 9 and Node 10.

Node 9 is a terminal node which shows that when students answered *strongly agree* for *being supportive* and *agree* or *not sure* or *strongly disagree* or *disagree* for *effectiveness*, they had 68.9% ratio to answer *strongly agree* for the target question while this ratio was 18.5% for *agree*. This means 87.4% of respondents provided positive answers to the target question.

Node 10 produces Nodes 16, 17, and 18. Node 16 is another terminal node which shows that 100% of all students who responded *strongly agree* for *being supportive* and *effectiveness* and *agree* or *disagree* for *variety of activities* provided positive answers to the target question (90.8% *strongly agree*, 9.2% *agree*).

Node 17 shows that when the same group of students answered *strongly agree* for *variety of activities*, their positive answer rate is 100% again for the target question; however, they answered *strongly agree* with a ratio of 97.8% and *agree* with a ratio of 2.2%.

Finally, Node 18 presents that when the same group mentioned above answered *not sure* or *strongly disagree* for *variety of activities*, their positive answer ratio for the target question dropped to 93.3% (80% *strongly agree*, 13.3% *agree*).

Figure 36

Section 4 of the Decision Tree for C Level Instructors

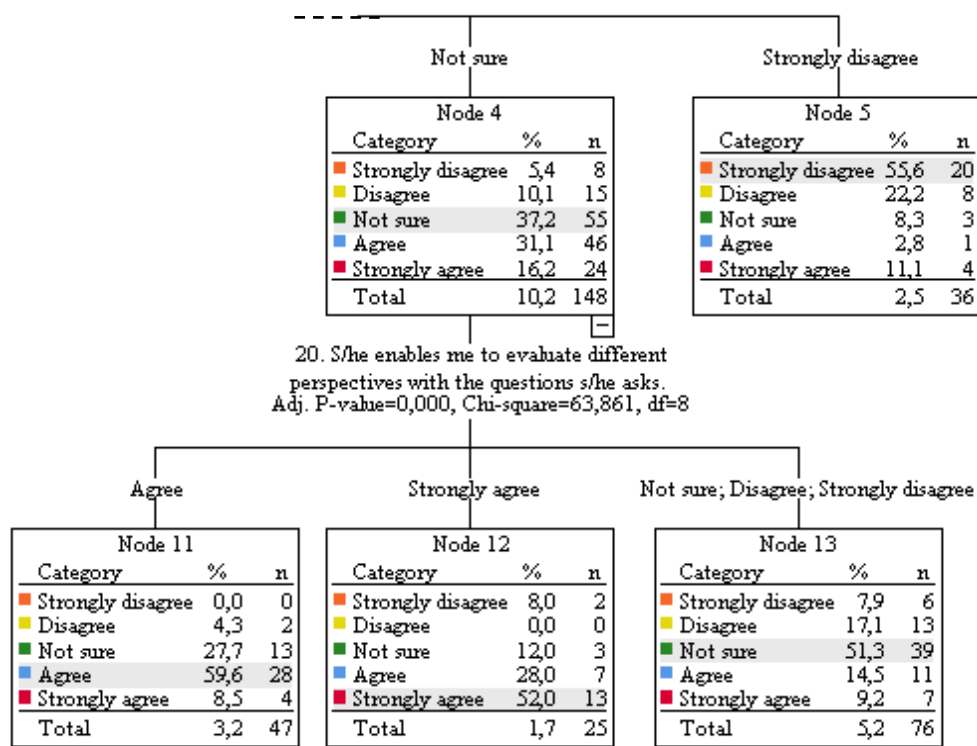


Figure 36 demonstrates the fourth section of the decision tree for C level instructors. In this section, none of the nodes can produce a higher ratio of positive answers than the root node. Therefore, they are not explained in detail here.

The Most Significant Predictors of Overall Teaching Performance Depending on Number of Trimesters Spent by Students in Preparatory School

This part presents the results of the analyses which are repeatedly conducted on the SET forms of instructors who worked in the first trimester, the second trimester, and the third trimester separately. The aim in this part is to discover whether the most significant predictors of teaching quality change depending on the number of trimesters spent by students in preparatory school.

The Most Significant Predictors of Overall Teaching Performance for the Instructors in the 1st Trimester

Figure 37 presents the outline of the decision tree produced for the instructors in the first trimester. Apart from Node 0, there are 23 nodes in three levels in depth. In the first level, the most significant predictor of overall teaching quality is *being supportive*. In the second level, *effectiveness*, *enabling students to evaluate different perspectives*, and *encouragement to share views* are the most significant predictors. Finally, *feedback* is the most important predictor in the third level.

Figure 38, Figure 39, and Figure 40 respectively present the first, second, and third sections of the decision tree created for the 1st trimester

Figure 37

Figure of the Decision Tree Produced for the Instructors in the 1st Trimester

Question 1

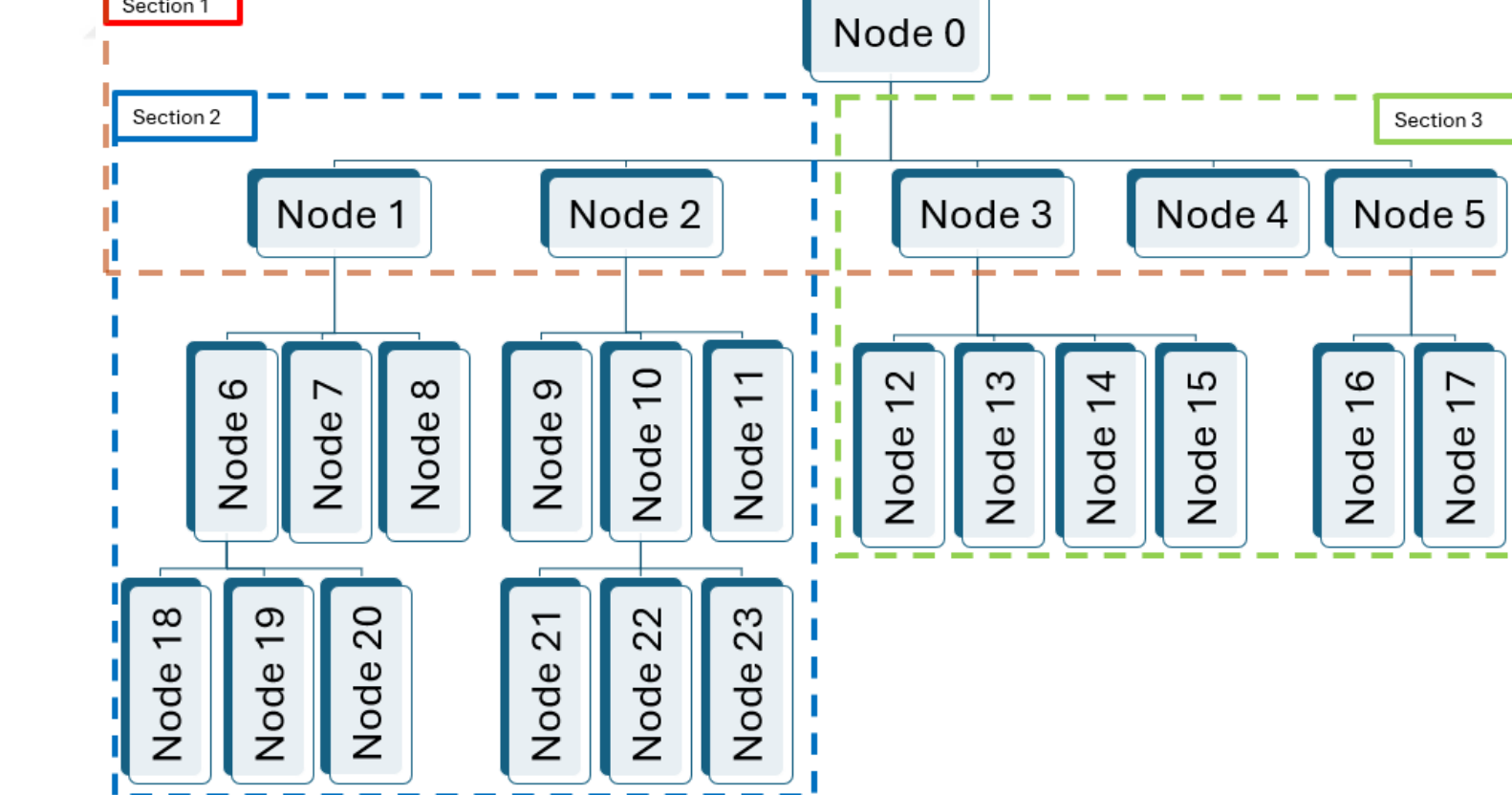


Figure 38

Section 1 of the Decision Tree for the Instructors in the 1st Trimester

23. I am glad that I have taken lessons from this instructor.

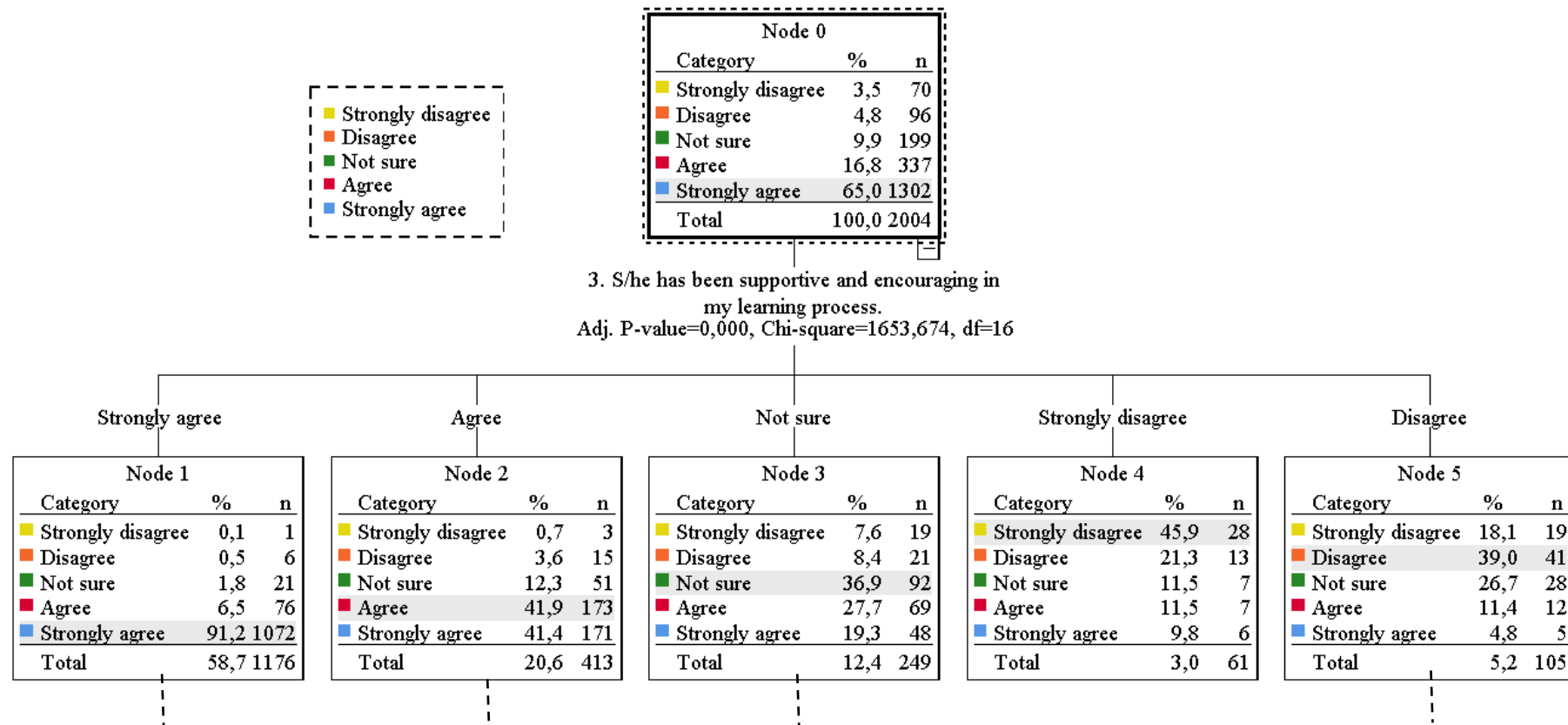


Figure 38 shows the first section of the decision tree for the instructors in the 1st trimester. Node 0 shows that there were 2004 students who participated in this SET survey in the 1st trimester. 65% of them ($n = 1302$) replied *strongly agree* for the target question while only 16.8% of them ($n = 337$) replied *agree*.

In the first level, only Node 1 and Node 2 can produce a higher ratio of positive answers than the root node. Node 1 indicates that 91.2% ($n = 1072$) of all students who answered *strongly agree* for *being supportive* also answered *strongly agree* for the target question. 6.5% ($n = 76$) of them answered *agree*. This means 97.7% positive answer ratio.

Node 2 shows that students who answered *agree* for *being supportive* provided positive answers to the target question with a total ratio of 83.3% (41.4% *strongly agree*, 41.9 *agree*).

Figure 39

Section 2 of the Decision Tree for the Instructors in the 1st Trimester

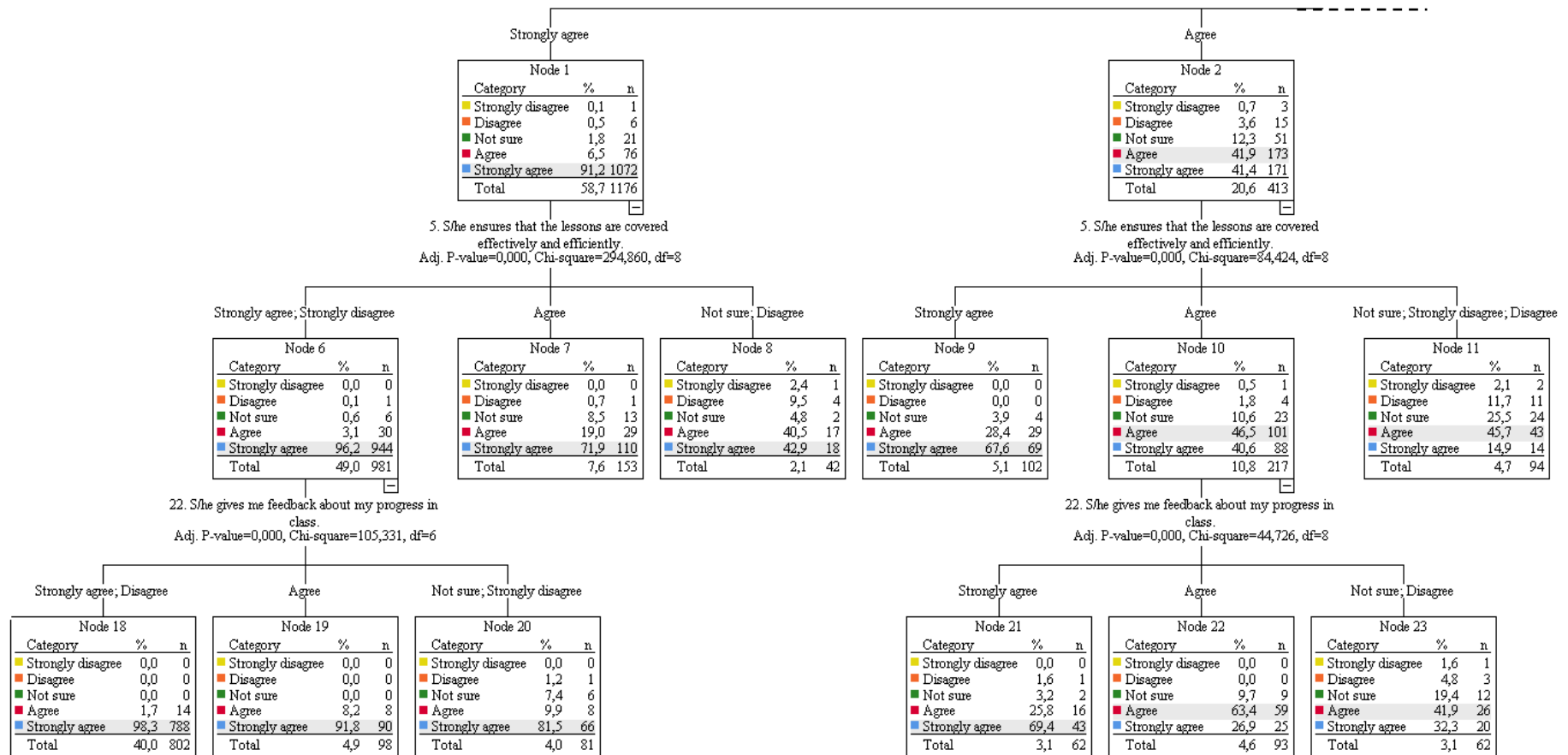


Figure 39 demonstrates the nodes created under Node 1 and Node 2. Under Node 1, there are three nodes, which are Node 6, Node 7, and Node 8.

Node 7 shows that when students responded *strongly agree* for *being supportive* and *agree* for *effectiveness*, their *strongly agree* ratio for the target question was 71.9% and *agree* ratio was 19%, which is 90.9% positive answer ratio.

Node 8 indicates that when the same group of students above responded *not sure* or *disagree* for *effectiveness*, their *strongly agree* ratio dropped to 42.9% and their *agree* ratio was 40.5%.

Node 6 has three child nodes, which are Node 18, Node 19, and Node 20. Node 18 indicates that when students replied *strongly agree* for *being supportive*, and *strongly agree* or *strongly disagree* for *effectiveness*, and *strongly agree* or *disagree* for *feedback*, they replied *strongly agree* to the target question with a ratio of 98.3%. Their ratio to reply *agree* was 1.7%. This means a ratio of 100% for positive answers to the target question.

Node 19 presents that when the same group of students responded *agree* for *feedback* question, their *strongly agree* ratio dropped to 91.8%, but their *agree* ratio rose to 8.2%. This is again 100% positive answer rate to the target question.

Node 20 shows that when the same group replied *not sure* or *strongly disagree* for the *feedback* question, they chose *strongly agree* for the target question with a ratio of 81.5% and *agree* with a ratio of 9.9%. This is 91.4% positive answer ratio.

Node 2 also produces three child nodes, which are Node 9, Node 10, and Node 11. Node 9 presents that students who chose *agree* for *being supportive* and *strongly agree* for *effectiveness* chose *strongly agree* for the target question with a ratio of 67.6% and *agree* with a ratio of 28.4%.

Node 10 produces three child nodes, which are Node 21, Node 22, and Node 23. Node 21 points out that students, who responded *agree* for *being supportive* and *effectiveness*, and *strongly agree* for *feedback*, also responded *strongly agree* for the target question with a ratio of 69.4% and *agree* with a ratio of 25.8%. This is 95.2% positive answer rate.

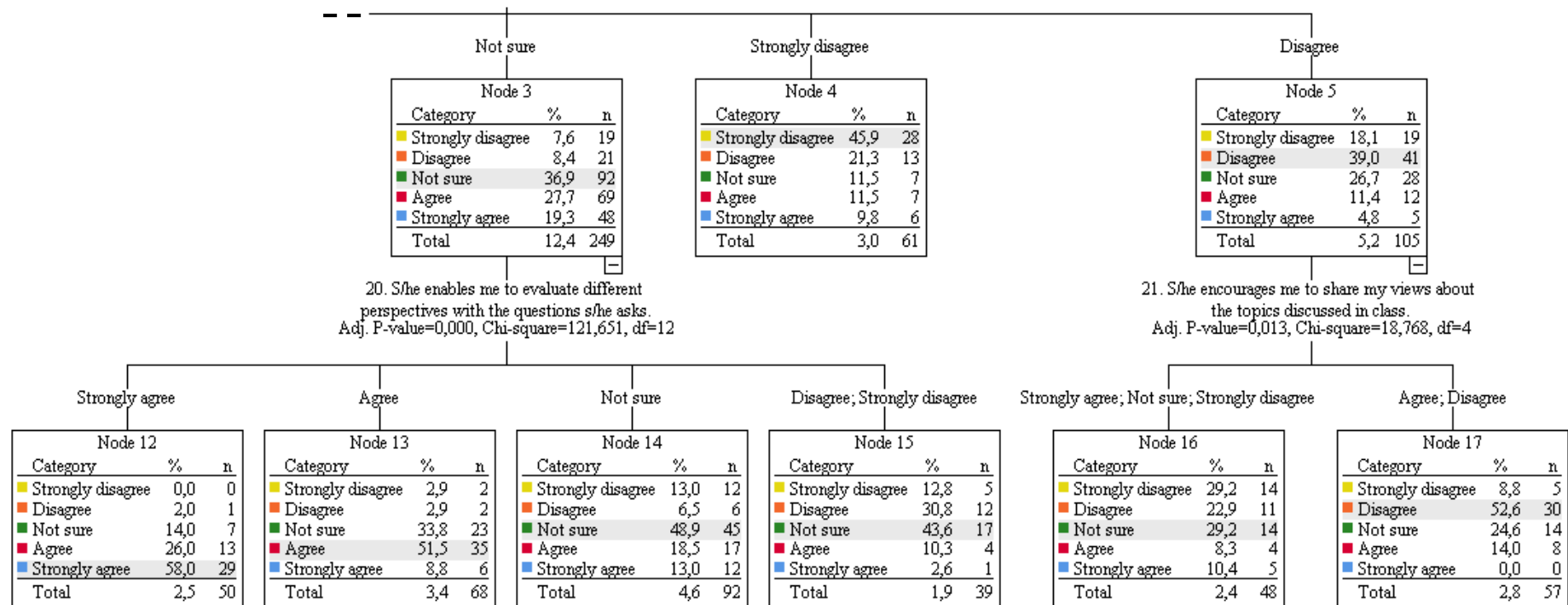
Node 22 shows that once the same group of students changed their answer for *feedback* and responded *agree*, then their *strongly agree* ratio for the target question dropped to 26.9%, which is a striking decrease. Their *agree* ratio was 63.4% for the target question. This means 90.3% positive answer rate.

Figure 40 shows the last section of the decision tree for the instructors in the 1st trimester. However, none of the nodes, except Node 12, can produce higher percentages of total positive answers than the root node. Therefore, only Node 12 is explained in this part.

Node 12 demonstrates that students who replied *not sure* for *being supportive* and *strongly agree* for *enabling students to evaluate different perspectives* also replied *strongly agree* for the target question with a ratio of 58% and *agree* with a ratio of 26%. This is 84% positive answer ratio.

Figure 40

Section 3 of the Decision Tree for the Instructors in the 1st Trimester



The Most Significant Predictors of Overall Teaching Performance for the Instructors in the 2nd Trimester

Figure 41 shows the outline of the decision tree produced for the instructors in the 2nd trimester. Except the root node, there are 19 nodes in three levels in depth. In the first level, the most significant predictor of overall teaching quality is effectiveness. In the second level, the most significant predictors are *encouragement to share views* and *enabling students to evaluate different perspectives*. In the third level, *guidance* and *feedback* are the most significant predictors.

Figure 42, Figure 43, and Figure 44 respectively demonstrate the first, second, and third sections of the decision tree for the instructors in the 2nd trimester.

Figure 41

Outline of the Decision Tree Produced for the Instructors in the 2nd Trimester

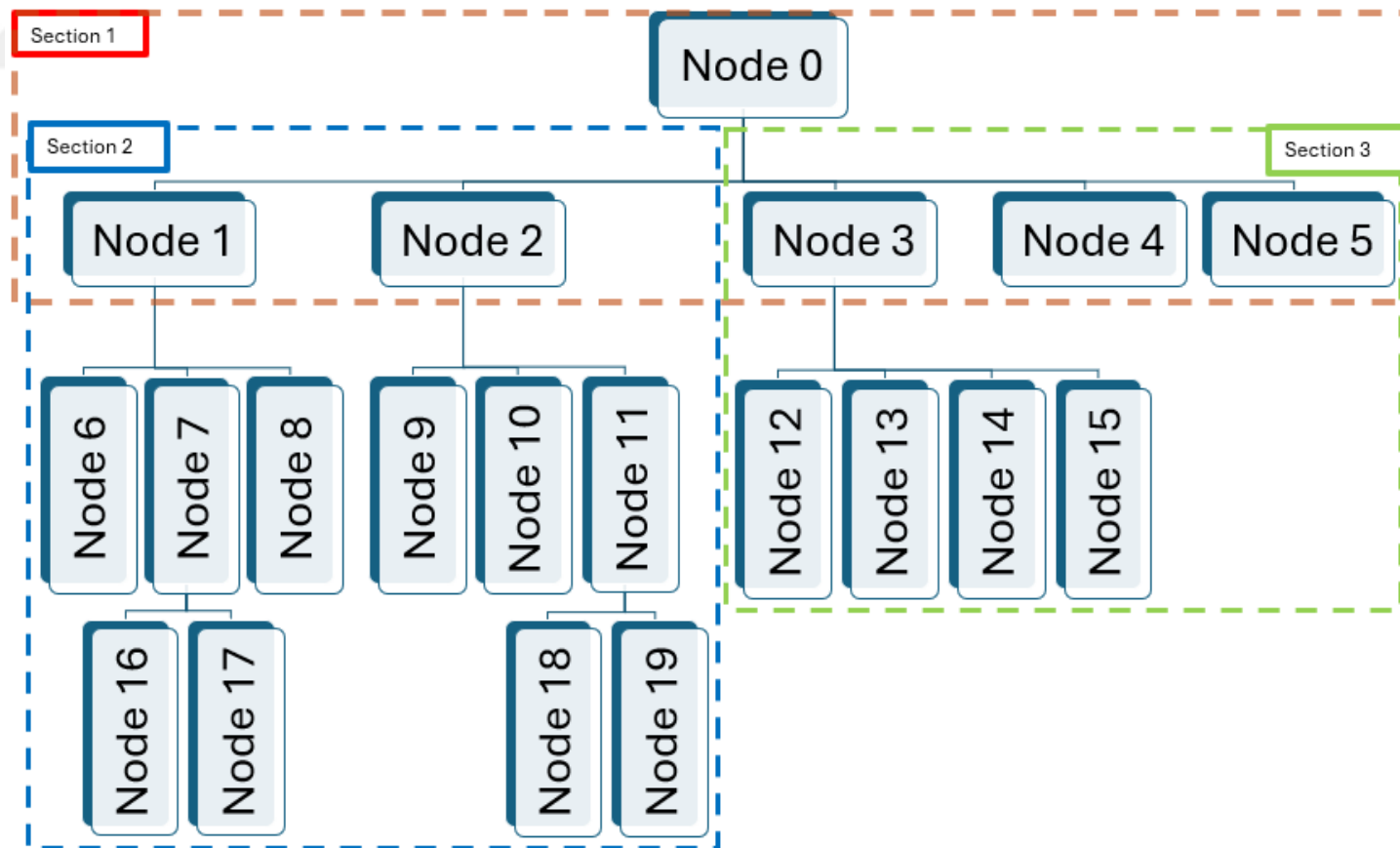


Figure 42

Section 1 of the Decision Tree for the Instructors in the 2nd Trimester

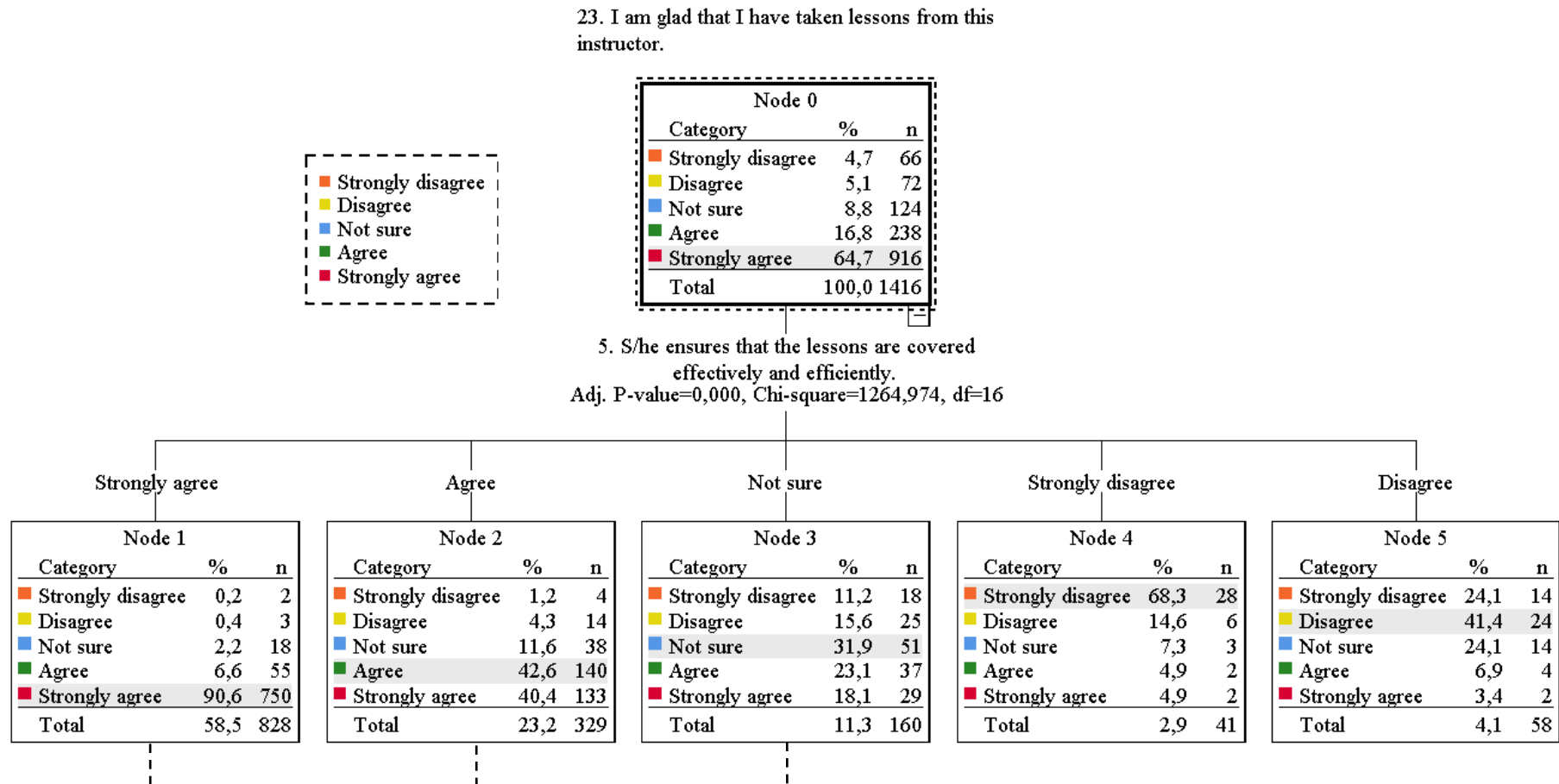


Figure 42 shows the first section of the decision tree for the instructors in the 2nd trimester. Node 0 presents that there were 1416 students who took this survey in the 2nd trimester. 64.7% (n = 916) of them answered *strongly agree* for the target question while 16.8 (n = 238) of them answered *agree*.

The most significant predictor of overall teaching quality is *effectiveness* in the first level. Nodes 1, 2, 3, 4, and 5 are created based on *effectiveness*. Only Node 1 and Node 2 can produce higher percentages of positive answers than the root node.

Node 1 shows that 97.2% of all students who answered *strongly agree* for *effectiveness* also provided positive answers for the target question (90.6% *strongly agree*, 6.6% *agree*).

Node 2 shows that when students answered *agree* for *effectiveness*, they answered *strongly agree* for the target question with a ratio of 40.4% and *agree* with a ratio of 42.6%. This is 83% positive answer rate.

Figure 43 demonstrates the second section of the tree.

Figure 43

Section 2 of the Decision Tree for the Instructors in the 2nd Trimester

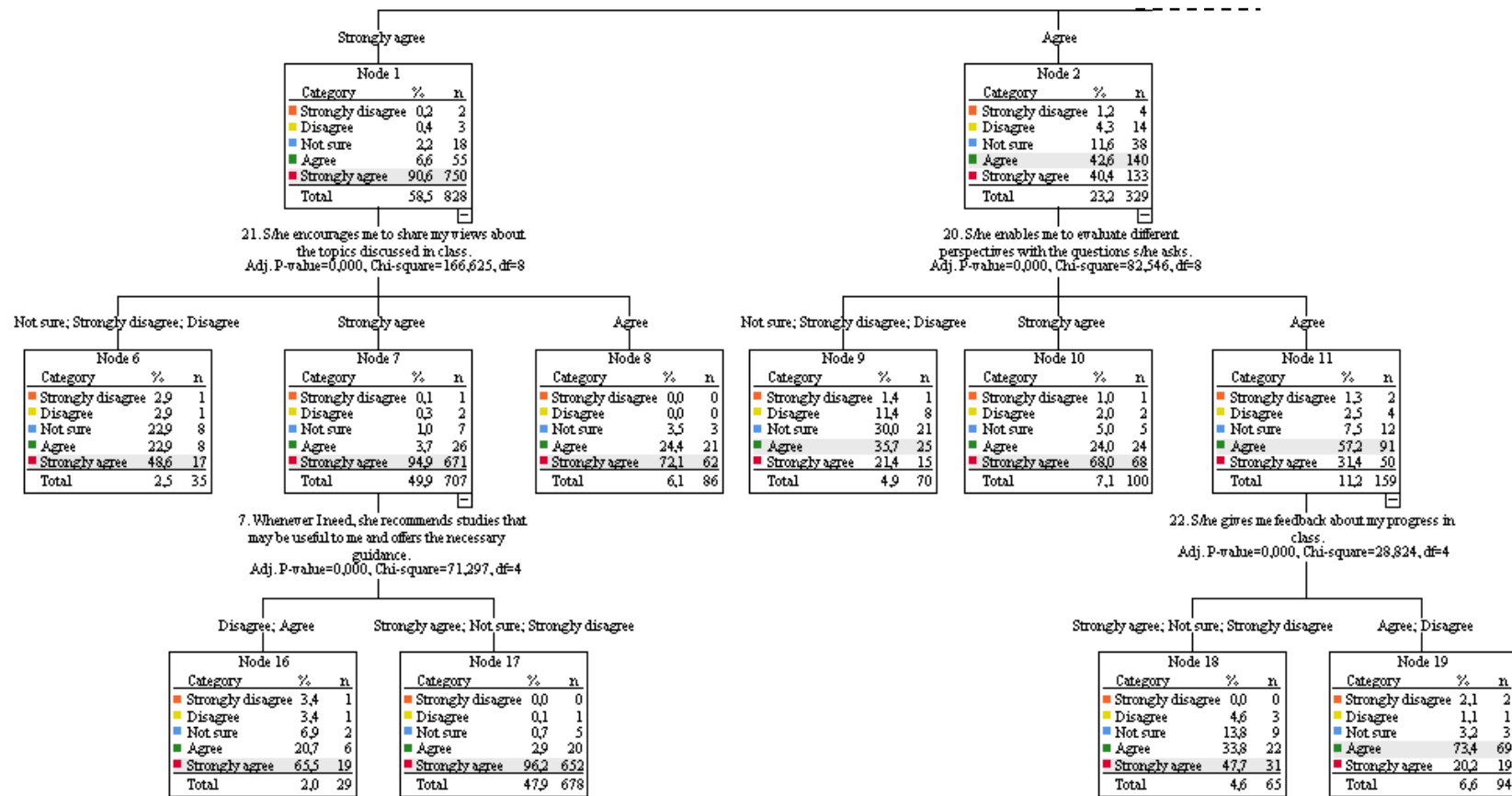


Figure 43 shows the second section of the decision tree for the instructors in the 2nd trimester. Node 1 produces three child nodes, which are Node 6, Node 7, and Node 8. Among them, only Node 8 is a terminal node which can produce a higher rate of positive answers than the root node. Node 8 shows that totally 96.5% of students who chose *strongly agree* for *effectiveness* and *agree* for *encouragement to share views* provided positive answers for the target question (72.1% *strongly agree*, 24.4% *agree*).

Node 7 also produces two child nodes, which are Node 16 and Node 17. Node 16 demonstrates that when students answered *strongly agree* for *effectiveness* and *encouragement to share views* but *disagree* or *agree* for *guidance*, they provided positive answers for the target question with a ratio of 86.2% (65.5% *strongly agree*, 20.7% *agree*).

Node 17 points out that when the same group of students changed their answers for *guidance* and chose *strongly agree*, *not sure*, or *strongly disagree*, their positive answer ratio for the target question rose to 99.1% (96.2% *strongly agree*, 2.9% *agree*).

Node 2 creates three child nodes, which are Node 9, Node 10, and Node 11. Node 9 has a lower ratio of positive answers than the root node. Node 10 shows that when students answered *agree* for *effectiveness* and *strongly agree* for *enabling students to evaluate different perspectives*, their positive answer ratio for the target question is 92% (68% *strongly agree*, 24% *agree*).

Node 11 produces two child nodes, which are Node 18 and Node 19. Node 18 demonstrates that 81.5% of all students who replied *agree* for *effectiveness* and *enabling students to evaluate different perspectives* and *strongly agree* or *not sure* or

strongly disagree for *feedback* also provided positive answers for the target question (47.7% *strongly agree*, 33.8% *agree*).

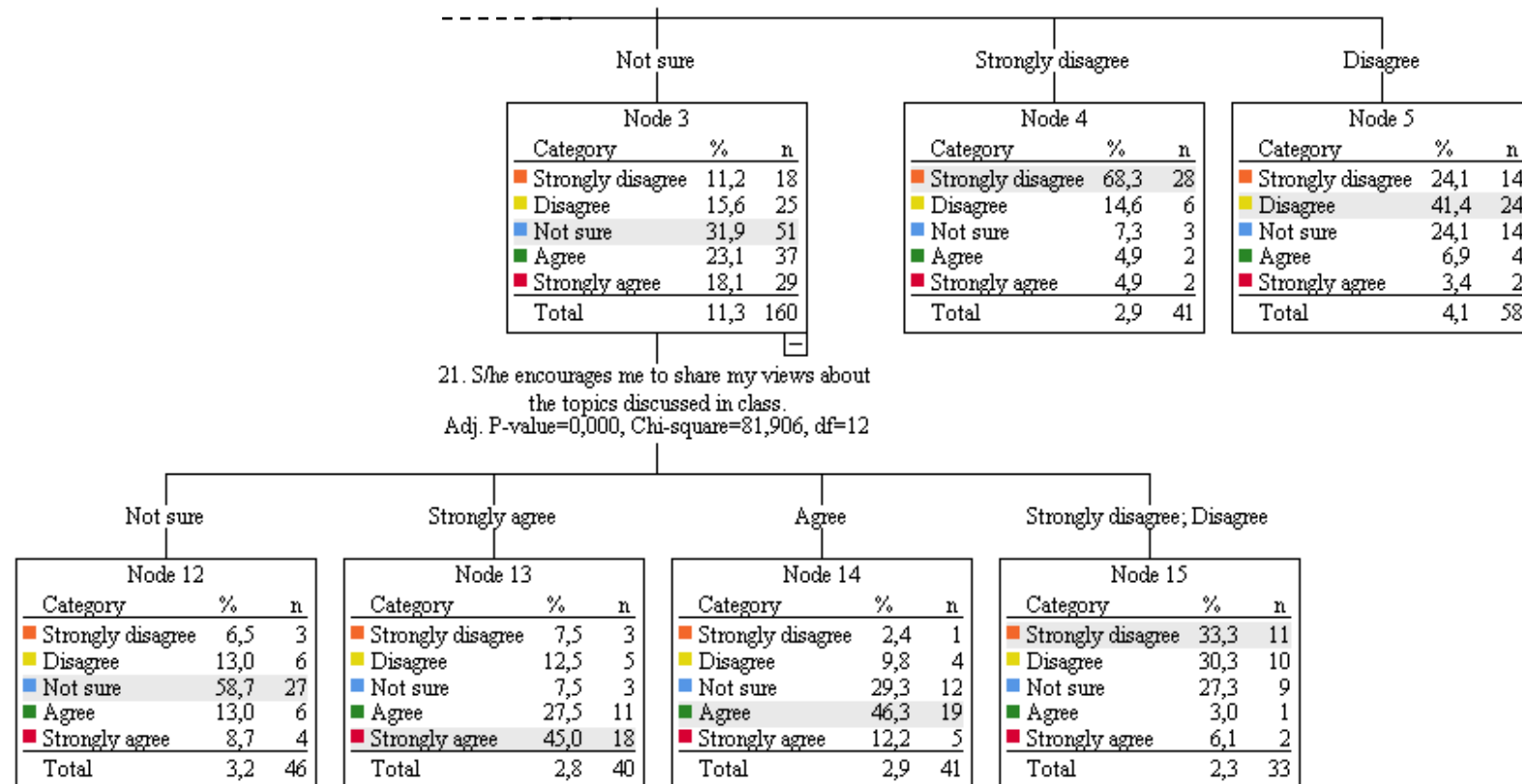
Node 19 shows that when the same group of students above altered their answers for *feedback* and chose *agree* or *disagree*, their positive answer ratio for the target question rose to 93.6% in total (20.2% *strongly agree*, 73.4% *agree*).

Figure 44 presents the last section of the decision tree for the 2nd trimester. However, as none of the nodes in this section can produce higher percentages of positive answers than the root node, they are not explained here.

On the other hand, it might be useful to mention the node which produces the lowest positive answer ratio when compared with the root node. Node 15 shows that only 9.1% of all students who answered *not sure* for *effectiveness* and *strongly disagree* or *disagree* for *encouragement to share views* provided positive answers for the target question (6.1% = *strongly agree*, 3% = *agree*).

Figure 44

Section 3 of the Decision Tree for the Instructors in the 2nd Trimester



The Most Significant Predictors of Overall Teaching Performance for the Instructors in the 3rd Trimester

Figure 45

Outline of the Decision Tree Produced for the Instructors in the 3rd Trimester

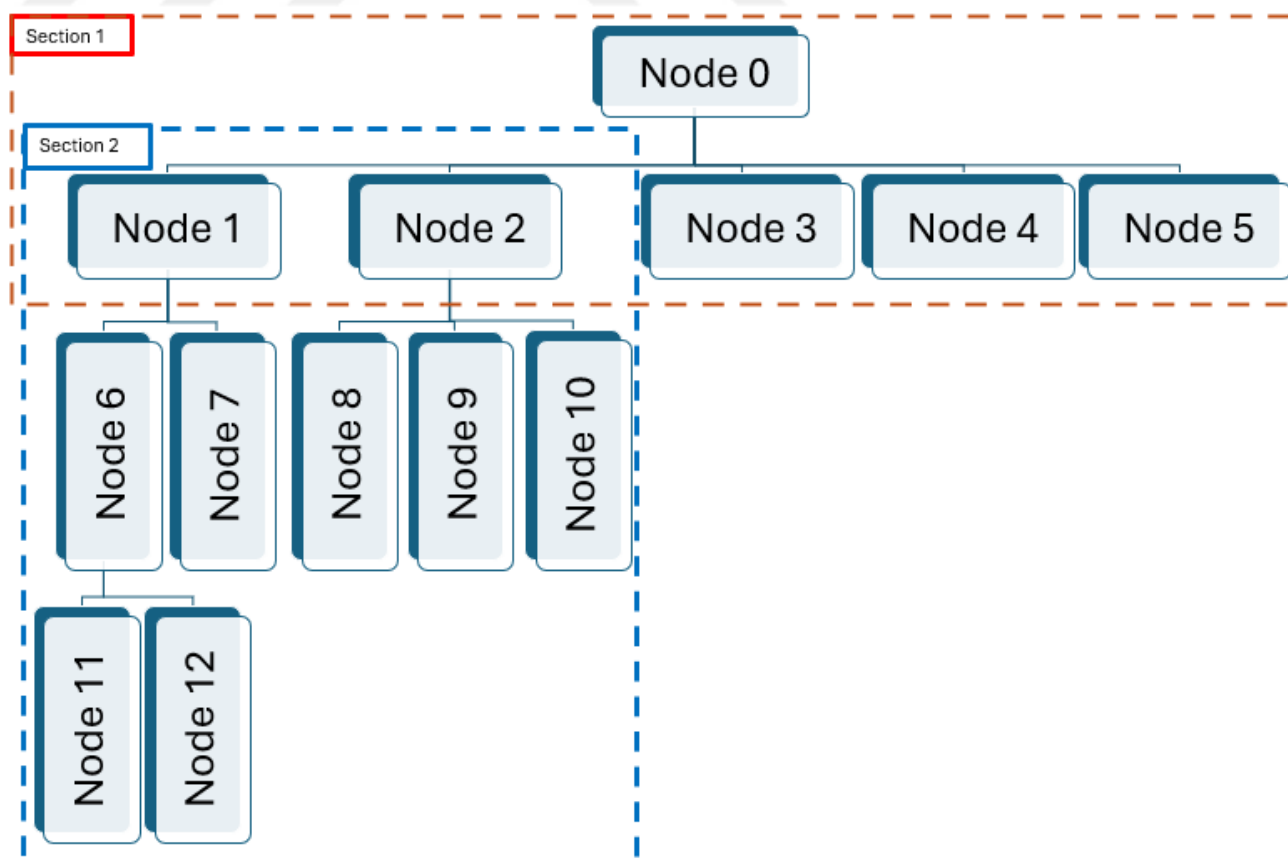


Figure 45 shows the outline of the decision tree created for the instructors in the 3rd trimester. Except the root node, 12 nodes are available in this analysis. In the first level, the most significant predictor of overall teaching quality is *effectiveness*. In the second level, *positivity* and *feedback* are the most important predictors. In the third level, *feedback* is the most important predictor again.

Figure 46 exhibits the first section of the decision tree for the 3rd trimester instructors. Node 0 shows that there were 861 students who took part in this SET survey. 62.3% (n = 536) of them answered *strongly agree* for the target question while 17.2% (n = 148) of them answered *agree*. The most significant predictor of teaching quality is *effectiveness* in the first level. Five nodes are created based on this predictor. They are Nodes 1, 2, 3, 4, and 5. However, only Node 1 and Node 2 can produce higher ratios of positive answers than the root node.

Node 1 shows that 97.1% of all students who replied *strongly agree* for *effectiveness* gave positive answers to the target question (91.4% *strongly agree*, 5.7% *agree*).

Node 2 demonstrates that when students replied *agree* for *effectiveness*, their positive answer ratio to the target question dropped to 80.4% (34.1% *strongly agree*, 46.3% *agree*).

On the other hand, Node 5 presents the lowest ratio of positive answers when compared with the root node in this analysis. It shows that when students answered *strongly disagree* for *effectiveness*, their total positive answer ratio to the target question was only 9.1%.

Figure 46

Section 1 of the Decision Tree for the Instructors in the 3rd Trimester

23. I am glad that I have taken lessons from this instructor.

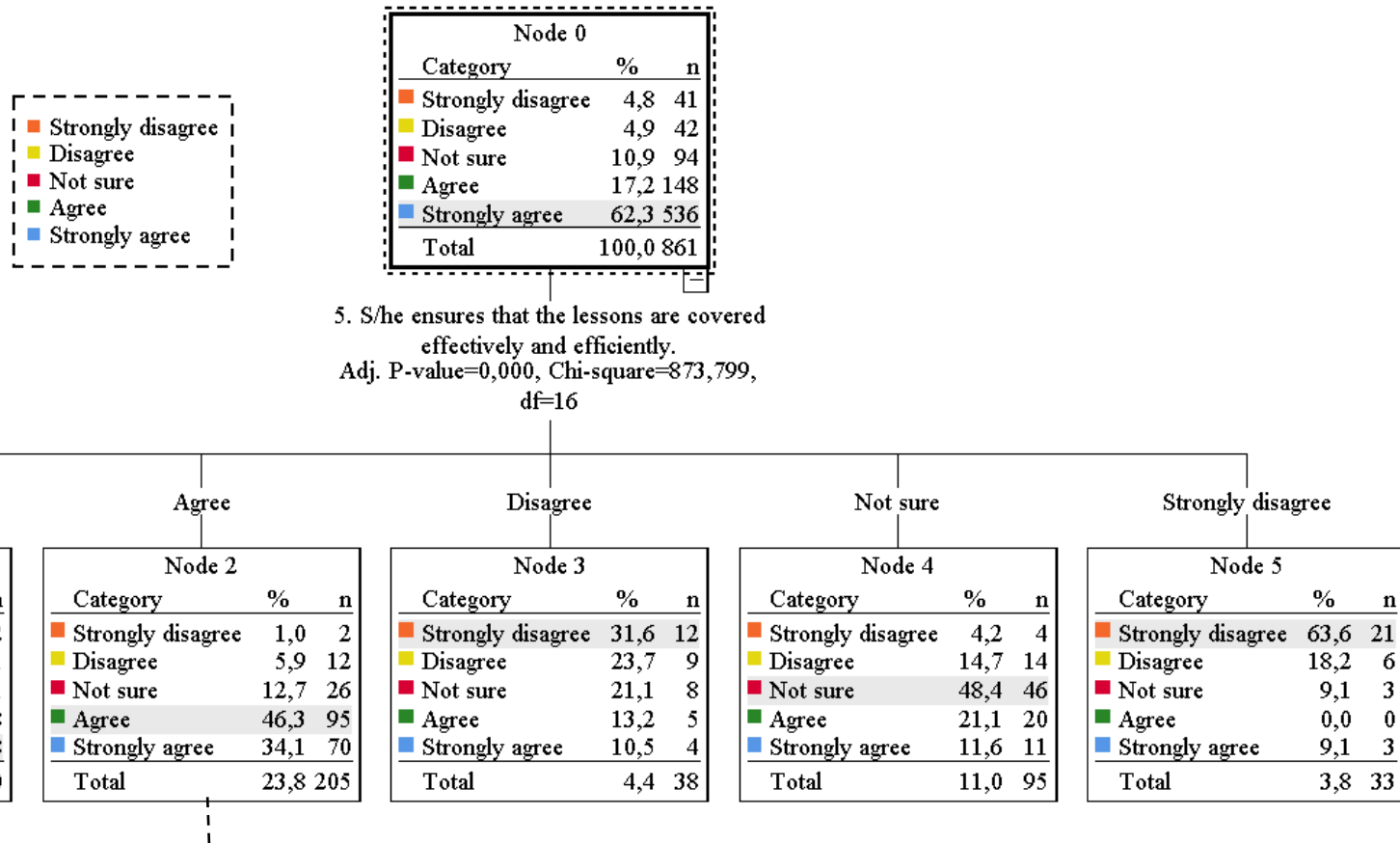


Figure 47

Section 2 of the Decision Tree for the Instructors in the 3rd Trimester

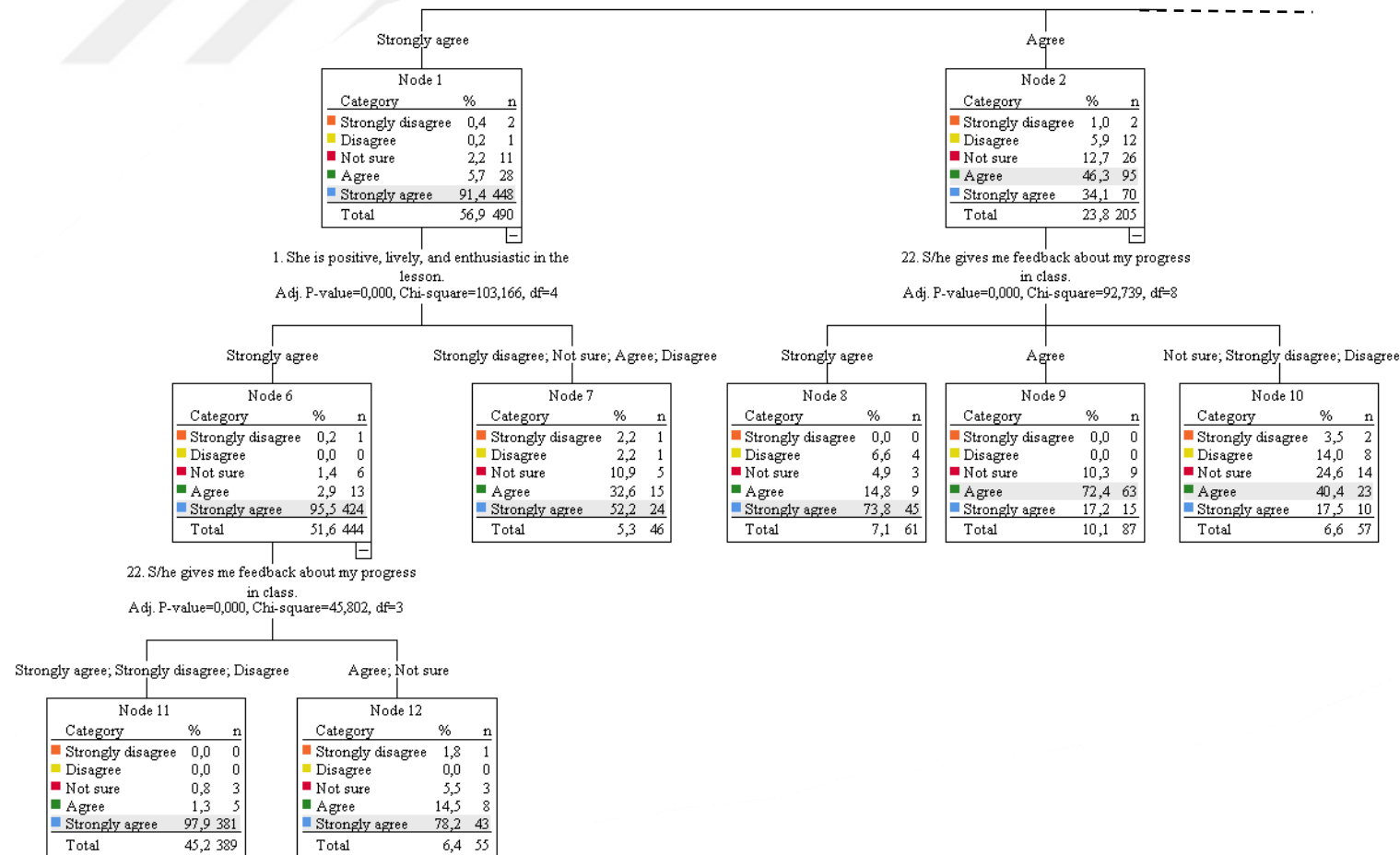


Figure 47 shows the second section of the decision tree for the 3rd trimester instructors. Node 1 produces two child nodes, which are Node 6 and Node 7.

Node 6 also produces two more child nodes, which are Node 11 and Node 12. Node 11 demonstrates that when students answered *strongly agree* for *effectiveness* and *positivity*, and *strongly agree* or *strongly disagree* or *disagree* for *feedback*, their positive answer ratio to the target question was 99.2% (97.9% *strongly agree*, 1.3% *agree*). Node 12 presents the idea that when the same group of students changed their answer for *feedback* and chose *agree* or *not sure*, their positive answer ratio to the target question was 92.7% in total (78.2% *strongly agree*, 14.5% *agree*).

Node 7 shows that when students replied *strongly agree* for *effectiveness* but *strongly disagree* or *not sure* or *agree* or *disagree* for *positivity*, their positive answer ratio to the target question was 84.8% in total (52.2% *strongly agree*, 32.6% *agree*).

Node 2 also produces three child nodes. They are Node 8, Node 9, and Node 10. Node 8 shows that 88.6% of all students who responded *agree* for *effectiveness* and *strongly agree* for *feedback* provided positive answers to the target question (73.8% *strongly agree*, 14.8% *agree*).

Node 9 demonstrates that when the same group of students above answered *agree* to *effectiveness*, their positive answer ratio to the target question was 89.6% (17.2% *strongly agree*, 72.4% *agree*).

Node 10 is not explained here since it cannot produce a higher ratio of positive answers than the root node.

Summary of Chapter 4

Chapter 4 can be summarized as follows:

- 11 different decision tree analyses were conducted under 4 parts in this study. The first analysis was conducted for the whole sample.

Then, three different analyses were conducted for MC, TW, and LS instructors. After that, four different analyses were conducted for AF, A, B, and C level instructors. Finally, three analyses were done for the 1st trimester, the 2nd trimester, and the 3rd trimester.

- Each decision tree analysis had three levels in depth with different numbers of nodes, which mainly depended on the number of participants.
- In all the analyses, only 13 questions out of 22 were found to be significant predictors of overall teaching quality. If ordered according to their significance level, they are *effectiveness, being supportive, enabling students to evaluate different perspectives, encouragement to share views, feedback, positivity, variety of activities, asking questions to encourage students to express opinions, subject knowledge, guidance, encouraging active participation, preparedness, and recommending publications in English.*

CHAPTER 5: DISCUSSION

Introduction

This study's goal was to discover the most significant predictors of overall teaching quality as measured by student evaluation of teaching forms, which were from a preparatory school of a non-profit private university. This chapter presents the discussion of the results of the study along with implications for practice and implications for further research. Finally, limitations of this study are presented.

Overview of the Study

To identify significant predictors of teaching quality based on student evaluation of teaching, a CHAID analysis, a decision tree analysis method, was first used on the whole sample on SPSS. Next, whether these predictors differ according to variables which are course type, students' proficiency level, and time spent by students at preparatory school was analyzed. The sample included 4281 SET forms that were filled in three trimesters by students who studied at this preparatory school in 2021-2022 academic year. However, each analysis, which was conducted for a different variable, included different numbers of participants because the number of students in each trimester, proficiency level, and course type was different.

The SET form was comprised of 23 Likert-type questions. The item "*I am glad that I have taken lessons from this instructor*" was chosen as the target question (the dependent variable) which showed students' overall satisfaction with teaching quality. Through decision tree analysis, it was possible to find out which questions in this SET survey could predict students' answers for the target question. Out of 22 questions in this SET survey, only 13 were found to be the predictors of student satisfaction.

However, these 13 questions varied in their level of significance.

Effectiveness and *being supportive* were the most significant predictors because only they appeared in the first levels of all the analyses. Then, *enabling students to evaluate different perspectives*, *encouragement to share views*, *feedback*, and *positivity* seem to share similar significance levels. Finally, *variety of activities*, *asking questions to encourage students to express opinions*, *subject knowledge*, *guidance*, *encouraging active participation*, *preparedness*, and *recommending publications in English* appeared in the third levels of the decision trees once or twice, which show that they are important, but their predicting power is less than the previous ones.

This study helps understand which characteristics of instructors make them quality instructors according to students. Furthermore, this study provides insights into how students' characteristics (time spent by students at preparatory school) and course characteristics (course grade and course type) affect the way students evaluate their instructors.

Discussion of Major Findings

For this study, 11 different decision trees were created using SPSS Version 26. First, an analysis was conducted for the whole sample. Then, three analyses were conducted for main course (MC), TOEFL & writing (TW), and listening & speaking (LS) instructors. Later, four more analyses were conducted for beginner (AF), elementary (A), intermediate (B), and upper-intermediate (C) level instructors. Finally, three analyses were done for the instructors teaching in the first, second, and third trimesters. Therefore, discussion section is divided into five parts. First, the findings from all 11 decision trees are discussed. Then, the findings of the whole

sample are discussed. After that, the comparisons of whole sample with different course types, proficiency levels, and trimesters are given.

Overall Findings from 11 Decision Trees

Table 5, which is given to make findings more visual, shows the most significant predictors of overall teaching quality obtained from all 11 analyses.

Table 5

The Most Significant Predictors of Overall Teaching Quality Obtained from all the 11 Analyses According to the Levels of the Trees

Question	Level 1	Level 2	Level 3
Q5 (Effectiveness)	7	2	1
Q3 (Being supportive)	4	3	1
Q20 (Enabling students to evaluate different perspectives)	0	7	2
Q21 (Encouragement to share views)	0	6	2
Q22 (Feedback)	0	3	4
Q1 (Positivity)	0	3	1
Q10 (Variety of activities)	0	0	2
Q19 (Asking questions to encourage students to express opinions)	0	0	2
Q6 (Subject knowledge)	0	0	1
Q7 (Guidance)	0	0	1
Q9 (Encouraging active participation)	0	0	1
Q12 (Preparedness)	0	0	1
Q16 (Recommending publications in English)	0	0	1

The findings of all 11 analyses indicate that there are certain characteristics of instructors that can help increase student satisfaction. Out of 22 questions used as

independent variables, only 13 were defined as predictors in the decision trees of all the analyses. Table 5 above shows the frequencies of each question in each level of all the decision trees created in this study. The questions are ordered according to how frequently they appeared in each level subsequently. Level 1 indicated the highest importance. *Effectiveness* is the most significant item because it appeared in the first level seven times; and then, *being supportive* is the second most significant item in all 11 CHAID analyses because it appeared four times in the first level. There is no other item that appeared in the first level in all the decision trees. Therefore, it can be assumed that students think that quality teaching means an instructor is effective and supportive above everything else, which are discussed below.

Being Effective

An instructor's being effective and efficient, which appeared seven times in the first level, twice in the second level and once in the third level, is not a surprising finding because all the students would like their instructors to be effective and efficient. Although being effective and efficient is open to interpretation, students might still have a common perception about these traits. When literature is reviewed, it can be seen that effective instructors are associated with many different characteristics by students, so it can be asserted that effective and efficient teaching results from the combination of many factors such as encouraging students for active participation, creating a safe place for students' self-expression, and guidance (Barrantes, 2022; Wei & Hui, 2019). Moreover, Barnett (2019) asserted that for an instructor to be effective and efficient, he/she should be aware of and cater for the individual learning needs of the students, which is not always easy. Passos (2009) stated that effective teaching is not a permanent characteristic of an instructor;

instead, it is the result of constantly changing relationships between instructors, students, and the other factors around them.

Therefore, it can be claimed that being an effective instructor is the most significant predictor of teaching quality because it is a trait which requires students to evaluate teaching quality from a broader perspective. This question does not only focus on one aspect of instructors. Thus, it should not be surprising that it can highly predict students' overall satisfaction with the instructor. For instance, in the decision tree for the whole sample (see Figure 5), it can be seen that students' ratio of providing a positive answer (agree and strongly agree) to the target question was significantly increased by *effectiveness* (from 81.2% to 97.7%). A similar increase can be observed in the analysis of AF level instructors. Students' positive answer ratio for the target question rose from 79.3% to 99.2% thanks to *effectiveness*. On the other hand, Schneider (2013) also claimed that effective teaching is difficult to describe because of its complex nature. The reason is that each student might have a slightly different perception about effectiveness and efficiency. For instance, some students like more strict instructors, so their effective and efficient teaching perception might be different from that of students who like instructors with more flexible teaching styles. Additionally, some students like explicit style of teaching where everything is explained in detail by the instructor. Those students like direct rules to be given to them. However, some students might like implicit style of teaching where they discover and analyze the topic that they are learning with the help of the instructor. Those students do not like lecture style lessons. Therefore, it should not be forgotten that students' preferences play a role in their decisions as to whether an instructor is effective and efficient or not. That means two students who

provide similar ratings for an instructor might have different notions of effectiveness in their minds.

Being Supportive

The second most significant predictor of students' overall satisfaction with the instruction is *being supportive*, which was the only predictor in the first level after *effectiveness*. It was seen four times in the first level, three times in the second level, and once in the third level. Its significant predictive power can be observed in A level instructors' SET forms (see Figure 25). Thanks to *being supportive*, students' positive answer ratio showed increase from 79.8% to 99%. This result is in line with other researchers' studies. For example, Cho et al. (2018) conducted a study to discover students' perceptions of quality teachers. Among other findings, an instructor's providing support for students was an influential factor for students to be satisfied with the lesson. León et al. (2015) also found similar results which indicate an instructor's being supportive is an important trait that makes his/her lessons satisfactory. Good et al. (2009) also found that supportive instructors facilitate learning better because students know that they can get help whenever they need from their instructors. The fact that *being supportive* is the second most significant predictor of students' overall satisfaction with instructors show how much an instructor's care for his/her students makes a difference in students' lives. There are instructors or teachers who are very knowledgeable in their areas, but their lack of support or indifference towards students makes students feel unimportant in their classes. This feeling discourages students from showing more effort in these instructors' classes. However, if students feel that they are important for an instructor and if they can get all the support they need from that instructor, they will be more willing to learn in that class. In short, among other characteristics of instructors,

being supportive is a trait which helps students decide whether they are content with an instructor or not. However, it should be noted that when students answered *not sure* to effectiveness in the whole sample analysis (see Figure 8), none of the answer types to *being supportive* could increase positive answer ratios to the target question, which again shows that an instructor's effectiveness is a more significant predictor than his/her being supportive for student satisfaction. If an instructor cannot conduct lessons effectively, this instructor's being supportive do not improve students' satisfaction with him/her.

Enabling Students to Evaluate Different Perspectives

Apart from *effectiveness* and *being supportive*, there is no other significant predictor in the first level in any of the decision trees. However, *enabling students to evaluate different perspectives* appeared in the second level seven times, and in the third level twice. This item's predictive power can be seen in MC instructors' SET forms (see Figure 11 and Figure 12) because while students' positive answer ratio to the target question was 88.9% with the help of *effectiveness* item, this ratio rose to about 94% with the help of *enabling students to evaluate different perspectives* item. This is an important finding which signals the importance of critical thinking skills. Another interesting fact is that this question was not available in the previous versions of this SET form. In the previous versions, there was a question which asked students whether instructors helped them think critically through different activities. However, some students could not grasp the meaning of critical thinking at that time; therefore, the question was rephrased, and it became "*S/he enables me to evaluate different perspectives with the questions s/he asks*", but its importance did not change because Karaca (2022) conducted a different analysis on the previous version of the same SET form used in this study, and she also found that the item

about critical thinking was associated with overall teaching quality. Similarly, Merrill (2018) also emphasized the importance of critical thinking, and he stated that when learners actively interact with each other during classes, they start to evaluate their peers' ideas and they start to reflect on their own ideas, which facilitate deeper learning. Therefore, *enabling students to evaluate different perspectives* is a sign of effective teaching. Likewise, Tsai and Lin (2012) found in their study that intellectual stimulation is correlated with students' satisfaction, and intellectual stimulation means presenting students with different perspectives for problem solving. Students' preference to develop their critical thinking skills might stem from its emphasis as a 21st century skill. In many schools, curriculum is now based on critical thinking because education is supposed to have transformative power; therefore, students might be affected by this policy. Moreover, rather than being the passive receivers of information, students might like being active and thinking about problems and their solutions by considering different perspectives. This might give them a sense of achievement. For this reason, *enabling students to evaluate different perspectives* might be one of the most powerful predictors of teaching quality for students in all the analyses.

Encouragement to Share Views

Another significant predictor in the list is *encouragement to share views*, which is relatively less important than the predictors discussed above. The predictive power of this item can be seen in the analysis of whole sample (see Figure 7). While the positive answer ratio to the target question was about 82%, *encouragement to share views* increased it to about 93%, which is a drastic increase. This finding is in line with the findings of Raymond (2008). She conducted a study to find out whether students' and instructors' perceptions of teaching quality matched or not. The study

showed that instructors' encouragement to make sure that students show active participation in discussions is an important trait which shows teaching quality according to both instructors and students. Delaney et al. (2010) published a research paper which was about students' perceptions of teaching quality in higher education. They found out that students like engaging instructors who encourage them to share their views and to take active roles in class, similar to what is found in this current study. This proves that students like being given chances to express themselves. This shows that students do not like being the passive receivers of the information; instead, they want to be actively involved. Moreover, some students might feel insecure about expressing their opinions, so they might wait for the instructors to encourage them or to give them a chance to do that. Therefore, these reasons might explain why *encouragement to share views* is a powerful predictor of effective teaching.

Feedback

The next most powerful predictor is "*S/he gives me feedback about my progress in class*", which appeared in the second level three times and in the third level four times. Although this question is shortened as *feedback* in the analysis part, it may not be about all the feedback types students receive in the preparatory school. The reason is that some students tend to ignore the written feedback that they get about their essays or written works unless they have to for revision. As the question focuses on the feedback about students' progress in class, this might be an indicator of students' wish to hear some good things about their performance in class. Raymond (2008) also found similar results. She discovered that providing frequent student feedback is deemed important by both students and instructors. Hattie (2008) also claimed that providing constant feedback is an important tool for effective

learning. As students get the chance to learn about their progress through feedback, they might feel more motivated to learn. In B level analysis, *feedback* seems to increase positive answer ratio from about 80% to about 93% (see Figure 29).

Positivity

Positivity is also a powerful predictor of overall teaching quality appearing three times in the second level of all trees and one time in the third level. *Positivity* could increase positive answer ratios from 97.4% to 99.3% as can be seen in the decision tree of TW instructors in Figure 14 and Figure 15. This finding is in line with Whitaker's (2020) claims. He asserted in his book that an important indicator of effective instructors is positive atmosphere in their classes. Even if instructors do not like all their students, it is important for them to create a positive environment in class to foster effective teaching. If students feel comfortable and safe in a classroom environment, it will be easier for them to focus on lessons, and they will be more willing to show active participation. Moreover, Ibad and Sharjeel (2021) found in their study that students think that positivity is a trait of effective instructors.

Variety of Activities & Asking Questions to Encourage Students to Express Opinions

The next equally significant predictors of overall teaching quality are *variety of activities* and *asking questions to encourage students to express opinions*, both of which appeared twice in the third level. As the new generation is born in an era of technology, they are used to being exposed to many different technological tools and they can find a game or an application or a website that can entertain them whenever possible. However, this situation makes them prone to being bored easily in classrooms. They always seek for a new activity that can attract their attention. Therefore, students expect instructors to bring variety of activities in class which can

make them actively engaged. Students start to lose their interest in classes where there are only dull and repetitive activities. This notion is also supported by some studies. For instance, Delaney (2010) emphasized that using variety of techniques to accommodate students' different needs is a trait which shows an instructor's ability to teach in an effective manner. As students learn in different ways and for different reasons, finding the suitable activities which can cater these needs improves the effectiveness of teaching. Moreover, in language classes, although students want to express their opinions, they might feel insecure about that for several reasons such as lacking necessary English skills to express themselves, being shy, and feeling scared of making mistakes (Li & Heydarnejad, 2024). According to their study, students who received encouragement from their instructors to express opinions in class felt more confident to show active participation. Moreover, Good et al. (2009) also claimed that carefully planned questions that can encourage students to show active participation are an asset of efficacious teaching. Therefore, it can be understood that students expect successful instructors to use purposeful questions which can encourage them to express their opinions, and in the absence of such incentives, students tend to stay quiet and reserved, which hinders the effectiveness of the lesson.

Subject Knowledge, Guidance, Encouraging Active Participation, Preparedness, Recommending Publications

The following predictors which are *subject knowledge, guidance, encouraging active participation, preparedness, and recommending publications* appeared in the third level once, so they can be claimed to have equal significance levels in this study. First of all, an instructor's subject knowledge is of utmost importance to be able to conduct the lessons effectively. Without good knowledge of

subject, an instructor cannot help his/her students achieve their goals. This finding is paralleled with the findings of Mohammed and Pandhiani (2017), Liu et al. (2016), and Lumpkin and Multon (2013). Secondly, it is not also surprising to find out *guidance* is deemed important by the students. The reason is that preparatory school students try to learn a new language and culture, which are not always easy to discover on their own. There are thousands of resources and techniques available to learn languages, and finding which one is the best for the needs of students is a challenging task. Therefore, instructors should be familiar with the techniques and resources that can best cater for students, and they should guide students accordingly. Moreover, the preparatory school where this study took place uses TOEFL ITP as a proficiency exam, and the resources are limited for this exam. The questions of TOEFL ITP are also different from many other language exams' questions. Hence, students might expect their instructors in this institution to help them understand the basics of TOEFL ITP as well. Cakici (2017) also claimed that instructors should not be people who only teach subjects in class, but they should be a guide or an advisor who help students become autonomous learners. They should recommend students extra studies which can help students improve outside the school, too. Chung and Wei (2020) asserted that language teaching should not merely focus on linguistic features; instead, teachers should be the guide for students to make them understand how and where these linguistic features are used in daily life by providing students with numerous opportunities. Thirdly, *encouraging active participation* is similar to *asking questions to encourage students to express opinions* in terms of making students actively participate in classes. However, *encouraging active participation* seems to be a broader term. Active participation in language classes is important because undoubtedly language is the most important medium of communication, and

if students do not actively participate during lessons, this means they do not communicate with each other, which is against the rationale behind learning a language. Hu (2020) also emphasized the importance of encouraging active participation in class through the use of variety of activities. Freeman et al. (2014) found that active participation in STEM classes improves students' success rates. Similarly, Hummel (2014) also mentioned the importance of active participation in language classes as languages cannot be learned as mechanical rules without having meaningful communication. Furthermore, it is not surprising that *preparedness* is one of the predictors because to manage the classroom well and to convey the subject matter effectively, instructors should be prepared. The classroom activities should carefully be planned to facilitate learning in the best possible way. Whitaker (2020) also declared that effective teachers do nothing randomly. They always have a plan in their minds and act accordingly to run classroom activities smoothly. Latif and Miles (2013) discovered the same finding in their study, which was about students' perceptions of effective teaching. They found that students regard an instructor's preparedness as an important characteristic for teaching quality. Finally, *recommending publications in English* is a significant predictor because in language teaching world, there are numerous resources for learners to do practice outside class. Nevertheless, finding the appropriate ones for the level and needs of students might be a daunting task. Hence, instructors should help learners find the right option for them. Renandya and Jacobs (2016) claimed that extensive reading and listening activities, which are done outside class environment, to improve language skills are important for students because students are exposed to many different language functions thanks to these activities. Therefore, instructors should guide students for the quality publications to enhance their learning.

After discussing all the predictors which were found in all the analyses, it might be useful to revisit the framework created by Wang and Williamson (2022) to understand the factors affecting student satisfaction. Their framework shows the relationship between different factors and SET scores. There are six qualities which are reported to affect student satisfaction in this framework. They are teaching qualities observable by students, teaching qualities unobservable by students, student qualities, student biases, course characteristics, and course grade. This study focuses only on teaching qualities observable by students, course characteristics, and student qualities. Out of all 13 predictors found in all the analyses, only efficiency, classroom discussion (appearing as enabling students to evaluate different perspectives, encouragement to share views, and asking questions to encourage students to express opinions in this study), feedback, positivity, innovative techniques (appearing as variety of activities in this study), time available outside of classroom (appearing as guidance in this study), and preparedness are available in the framework of Wang and Williamson (2022). Although these seven predictors are available in Wang and Williamson's framework, the other six predictors that were found in this study are not mentioned at all. Moreover, this study also aimed to discover whether course type, course level, and time that students spend at school affect the way students evaluate their instructors. Wang and Williamson's (2022) framework also includes course type and course level as the important factors affecting students' satisfaction with the instructors. However, there is no reference to the time spend by students at school in that framework. Therefore, it can be claimed that although Wang and Williamson included many significant factors that affect students' satisfaction in their framework, there are still other factors which are

important yet not mentioned by them. Hence, it might be useful for other researchers if the scope of Wang and Williamson's (2022) framework is extended.

Discussion of the Findings for the Whole Sample

As can be seen in Figure 4, there are 37 nodes except the root node in three levels in the decision tree for the whole sample. Table 6 demonstrates the predictors which appeared in each level of this decision tree.

Table 6

Predictors Appearing in Each Level of the Decision Tree for the Whole Sample

Level	Predictors
1	Effectiveness
2	Being supportive, encouragement to share views, enabling students to evaluate different perspectives
3	Feedback, positivity, enabling students to evaluate different perspectives, being supportive, asking questions to encourage students to express opinions

As all these predictors have already been discussed, they are not discussed again here. However, some terminal nodes demonstrate intriguing results, so it might be important to mention them here.

First of all, as can be seen in Figure 6, Node 21 shows that students who chose *strongly agree* for *effectiveness* and *being supportive* and *strongly agree* or *disagree* for *feedback* have 99.4% ratio to provide positive answers to the target question. This is surprising to see because it seems that these students might answer *strongly agree* or *disagree* for *feedback*, but this situation does not change anything

about their answer to the target question. This seems contradictory, but the reason might stem from the numbers of students who chose *strongly agree* or *disagree* for *feedback*. When descriptives of the whole sample is created on SPSS, it can be seen that there are hundreds of students who chose *strongly agree* while there are only a few students who opted *disagree* for *feedback*. SPSS tends to classify even contradictory groups together when the numbers of the participants are so low that they do not make a statistically significant difference to the results. Therefore, it should be noted that nodes which exhibited intriguing results by combining contrasting answers together resulted from SPSS' working algorithm. There are numerous nodes as such in all the analyses, so they are not repetitively discussed here as most of them have similar explanations.

Discussion of the Findings for Different Course Types and Their Comparison with the Whole Sample

In this part, the predictors found in the analyses of MC, TW, and LS instructors' SET results are shared along with the predictors found for the whole sample (see Table 7). This helps understand how each analysis reveals certain similar and different predictors.

Table 7

The Comparison of Predictors for the Whole Sample, MC, TW, and LS Instructors

Level	Whole Sample	MC	TW	LS
1	Effectiveness	Effectiveness	Effectiveness	Being supportive
2	Being supportive, encouragement to share views, enabling students	Positivity, enabling students to evaluate different	Positivity, being	Encouragement to share views, enabling

Level	Whole Sample	MC	TW	LS
	to evaluate different perspectives	perspectives	supportive	students to evaluate different perspectives
3	Feedback, positivity, enabling students to evaluate different perspectives, being supportive, asking questions to encourage students to express opinions	Encouragement to share views	Encouraging active participation, subject knowledge	Effectiveness, preparedness

In Table 7, in the first level, although whole sample, MC, and TW instructors' analyses include effectiveness as the most significant predictor, being supportive seems more significant for LS instructors according to the students. This situation might stem from the fact that in LS courses, students are supposed to speak relatively more than in other classes, and if students feel shy, they may need support from their instructors to actively engage with the activities. Zumbrunn et al. (2014) also claimed that supportive instructors enhance their students' engagement levels during classes. In the second level of the trees, all the predictors are visible at least twice, which indicates a common perception of quality teaching in the second level. When these predictors are checked, it can be seen that getting support from instructors to be actively engaged with the lessons and to express opinions is an

important factor for students to be satisfied with the instructors. Moreover, positivity does not seem to be an important predictor for the whole sample and LS instructors. Although positivity is visible in the third level of the whole sample, it does not appear for LS instructors at any level, which is interesting. An explanation might be the fact that LS instructors are mostly Americans in the preparatory school, and because of cultural differences, being positive might be perceived differently for them. Students may not be able to fully comprehend positivity of instructors from a different country. Furthermore, encouragement to share views is available in the second level of whole sample and LS instructors, and it is available in the third level of MC instructors. However, it never appears for TW instructors. The reason might be the fact that TW lessons mostly include TOEFL ITP test techniques and writing strategies, which means students do not have a lot of opportunities to share their views about different matters. In the third level of these decision trees, differences can be observed. First of all, feedback and asking questions to encourage students to express opinions are available in the whole sample analysis, but they are not visible in any of MC, TW, and LS instructors' analyses. This might mean that their predictive power is so low that when the sample sizes get smaller in MC, TW, and LS instructors' analyses, they cannot predict students' satisfaction levels. However, encouragement to share views is deemed significant for MC instructors while encouraging active participation is important for TW instructor. As both of these items evaluate an instructor's ability to promote active participation, they are similar in nature. On the other hand, subject knowledge only appears in the analysis of TW instructors, which might be a hint for those instructors' need to be very knowledgeable in TOEFL ITP strategies as this exam has very specific types of questions which require expertise in the field. Another interesting finding is that

preparedness is deemed important only for LS instructors. This situation may also imply that for effective listening and speaking lessons, instructors should be prepared carefully to make sure that students benefit from the lessons. Students might find MC and TW lessons more mechanic as they mainly include skills like grammar, vocabulary, reading and writing. However, students might expect their LS instructors to bring more enjoyable and engaging speaking and listening activities which require more preparation. Burroughs et al. (2019) also emphasized the importance of teacher preparedness for instructional quality. Finally, when root nodes of all these analyses are checked, it can be seen that overall satisfaction level for the whole sample was 81.2% while it was 76.6% for LS instructors, 78.4% for TW instructors, and 89.3% for MC instructors. This means MC instructors receive significantly higher rates of student satisfaction when compared to TW and LS instructors. This might happen because MC instructors spend 15 hours a week with a class while TW and LS instructors spend only 5 hours a week. This might cause instructors to form better rapport with their MC students, which in turn results in higher student satisfaction levels. Sweeney et al. (2009) found that teacher rapport influences SET results. Khandelwal (2009) also emphasized the importance of instructor rapport to be considered as an effective instructor. Beran and Violato (2005) also claimed that when students and instructors spend more time together in class, they have increased rapport, which in turn leads to better student ratings.

Discussion of the Findings for Different Proficiency Levels and Their Comparison with the Whole Sample

In this part, the results of the analyses for AF, A, B, and C level instructors are shared along with the results of the whole sample to be able to compare them more easily. As can be seen in Table 8, in the first level while effectiveness is

deemed important for whole sample, AF and B level instructors, being supportive is a more significant predictor for A and C level instructors. Students might expect C level instructors to be supportive because these students take the proficiency exam at the end of the trimester, and that is why they are stressed. Therefore, an instructor's being supportive might decrease students' level of anxiety. This notion aligns with the findings of Lei et al. (2018). They found that increased instructor support is associated with lower anxiety level of students. However, it is not very clear why A level students deem being supportive more than effectiveness. On the other hand, being supportive is regarded as a significant predictor in the second level only in the whole sample and AF level analysis, so this means being supportive is available in all analyses at some level except B levels, which is an interesting finding. B level students apparently value other traits of instructors more than their being supportive. In the second level, being supportive, feedback, encouragement to share views, and enabling students to evaluate different perspectives are the most significant predictors in all proficiency levels. These findings are similar to the findings of the analysis for the whole sample. It can be understood that students value opportunities where they can express themselves with the support of instructors and use their critical thinking skills while interacting with their peers. They also like receiving feedback as to their progress. These findings are in line with Mercer and Dörnyei's (2020) statements. They claimed that interpersonal communication improves students' motivation to learn, so it is natural for students to want to be actively involved in classroom activities. Ripoll-Núñez et al. (2018) found that students in social sciences departments valued critical thinking skills and having good interpersonal relationships more than engineering students. Similarly, preparatory schools where English is mainly taught for an extended period of time require

students to have interpersonal relationships and to critically think about different matters while learning English; therefore, it is reasonable that preparatory school students in this study value encouragement to share views and being able to evaluate different perspectives, as well. In the third level, different from the levels one and two, variety of activities seems to be a new predictor only in A and C levels. As discussed before, variety of activities increases teaching effectiveness. Hu (2020) claimed that using diverse techniques and activities to teach is an essential part of quality teaching since they can provide a more engaging environment for students with different needs. Another interesting finding is that recommending publications in English appears only in B levels. The reason might be the level of the students. B level students are intermediate level, and this is a level where students can partially understand authentic texts or sources on their own. This situation might explain why they are interested in publications in English. For AF and A level students, it is too hard to engage with these types of materials, and perhaps, C level students are so worried about passing TOEFL ITP exam that they do not spare time to enjoy publications in English for their personal development, and apparently publications in English are not valued significantly in the whole sample. Finally, when satisfaction levels of students in each proficiency level are compared, it can be seen that it is 79.3% for AF level instructors, 79.8% for A level instructors, 79.5% for B level instructors, 83% for C level instructors while it is 81.2% for whole sample. Although differences are not substantial, C level instructors seem to obtain higher ratings from students. This finding is not in line with the findings of Yu et al. (2022) and Kalender (2011). They found that when grade level of a course increases, SET results for instructors decrease. However, in this study, C level instructors, who are teaching the highest proficiency level students, receive higher rates of satisfaction

from students. Nevertheless, this situation may stem from the fact that when students are motivated to learn, they tend to rate their instructors higher (Jones et al., 2022). It can be claimed that C level students are more proficient in English than the students in other levels, and that is why they are generally more motivated to learn English. They enjoy engaging in activities in class. This situation might explain why C level instructors are rated higher for satisfaction.

Table 8

The Comparison of the Predictors for the Whole Sample, AF, A, B, and C Level

Instructors

Level	Whole Sample	AF	A	B	C
1	Effectiveness	Effectiveness	Being supportive	Effectiveness	Being supportive
2	Being supportive, encouragement to share views, enabling students to evaluate different perspectives	Being supportive, feedback	Encouragement to share views	Feedback, encouragement to share views, enabling students to evaluate different perspectives	Effectiveness, enabling students to evaluate different perspectives
3	Feedback, positivity, enabling students to evaluate different perspectives, being supportive, asking questions to encourage students	Encouragement to share views	Variety of activities	Enabling students to evaluate different perspectives, recommending publications in	Asking questions to encourage students to express opinions, variety of

Level	Whole Sample	AF	A	B	C
	to express opinions			English	activities

Discussion of the Findings for Different Trimesters and Their Comparison with the Whole Sample

The following table demonstrates the predictors which appeared in the analyses of the SET results of the instructors teaching in the 1st trimester, the 2nd trimester, and the 3rd trimester along with the analysis of the whole sample.

Table 9

The Comparison of the Predictors for the Whole Sample, the 1st Trimester, the 2nd Trimester, and the 3rd Trimester Instructors

Level	Whole Sample	1 st Trimester	2 nd Trimester	3 rd Trimester
1	Effectiveness	Being supportive	Effectiveness	Effectiveness
2	Being supportive, encouragement to share views, enabling students to evaluate different perspectives	Effectiveness, enabling students to evaluate different perspectives, encouragement to share views	Encouragement to share views, enabling students to evaluate different perspectives	Positivity, feedback
3	Feedback, positivity, enabling students to evaluate different perspectives, being	Feedback,	Guidance, feedback	Feedback

Level	Whole Sample	1 st Trimester	2 nd Trimester	3 rd Trimester
	supportive, asking questions to encourage students to express opinions			

In the first level, being supportive appears only once, and it is in the 1st trimester. This might be caused by the fact that students start preparatory school in the 1st trimester, and everything is new and challenging for them. They do not know anybody at school, they do not know the rules and regulations, and they are not familiar with how the lessons are covered in preparatory school. In such a challenging situation, it is not surprising that students expect their instructors to be supportive, which can ensure a smooth transition into university life. Wilson et al. (2020) also revealed that faculty support causes students to have a positive attitude towards their experiences at university. When it comes to the second level, the 3rd trimester stands out from others because feedback and positivity appear only in the 3rd trimester. They are not available in the whole sample and the other trimesters. This might result from the fact that most of the students in the 3rd trimester are supposed to take TOEFL ITP as the proficiency exam, and they are extremely anxious about it because if they fail in this exam, they become repeat students, which means they cannot start studying in their departments. In this highly stressful environment, an instructor's positivity can have a beneficial effect on students' lives. Additionally, being supportive is available in the second level only in the whole sample. It hints the idea that even though students do not consider being supportive as the most significant predictor in the whole sample, it is still quite important for

them. However, this item does not appear in any level of the 2nd or 3rd trimesters. This may result from the idea that students are used to the mechanics of preparatory school in the 2nd and 3rd trimesters unlike they are in the 1st trimester, which might make being supportive less important for them. Moreover, feedback is available twice in both the second and third levels of the analysis for the 3rd trimester instructors. This shows the importance of feedback for students, and it is highly likely that these students have a desire to learn about their performance in class, which probably provides them with relief about their possibility to pass TOEFL ITP exam. Furthermore, feedback is available again in the third levels of all the decision trees including the whole sample, which indicates its importance. Hattie and Timperley (2007) also highlighted the importance of feedback saying it has a substantial impact on learning. In the third level, apart from feedback, the only different predictor is guidance, which appears in the 2nd trimester. Kalay (2017) also found that students highly regard instructors who allocate time to help and guide students whenever they need. Finally, the ratio of the satisfaction level of the students for the 1st trimester is 81.8% while it is 81.5% for the 2nd trimester, and 79.5% for the 3rd trimester. The same ratio is 81.2% for the whole sample. It seems that overall satisfaction levels of students start to decrease after the 1st trimester. Beran and Violato (2005) asserted that when students are presented with SET forms for the first time, they are willing to provide feedback. However, after taking the same SET form several times, they start to feel bored of them; therefore, they provide lower scores for their instructors. This claim might be the reason why student satisfaction with instructors starts to decline after the 1st trimester.

Summary

To summarize, effectiveness and being supportive are the two most powerful predictors of student satisfaction in all analyses. Following them, enabling students to evaluate different perspectives, encouragement to share views, feedback, and positivity are highly valued by students. Although their predictive power is relatively small, variety of activities, asking questions to encourage students to express opinions, subject knowledge, guidance, encouraging active participation, preparedness, and recommending publications in English are also contributing factors to student satisfaction.

Moreover, students' overall satisfaction level changes depending on course type, students' proficiency level, and the time spend by students at preparatory school. It was discovered that students' satisfaction is highest with MC instructors while LS instructors receive the lowest scores. When it comes to the proficiency level of students, C level (upper-intermediate) students have the highest level of satisfaction with instructors while students in other levels have lower levels of satisfaction. Additionally, students' level of satisfaction with instructors decreases in time because students in the first trimester have the highest level of satisfaction although students in the third trimester have the lowest satisfaction level.

Finally, it was found that instructors' qualities which can predict student satisfaction show difference depending on course type, proficiency level, and time students spend at preparatory school. For instance, whereas being supportive is the most significant predictor of student satisfaction in LS courses, effectiveness is deemed more important by students in MC and TW courses. There are many differences as such, so please see discussion of the findings for more information.

Implications for Practice

This study might be useful for curriculum planners and material developers to improve student satisfaction within preparatory schools at higher education institutions. The reason is that through 11 decision tree analyses, the powerful predictors of student satisfaction were discovered, and curriculum planners might take these into consideration while planning the yearly programs. For instance, encouraging students to participate actively in class and asking questions which direct students for critical thinking were found to be very significant predictors in this study; hence, material developers might design activities which are highly engaging for students. Moreover, while deciding on the books to be used in the preparatory school, curriculum planners might choose the ones with more interactive activities which promote critical thinking.

Additionally, this study is useful for instructors, as well. Being supportive and positive, or providing regular feedback and guidance for students is deemed valuable by students in this study, so instructors should take these into consideration when they are teaching so that they can enhance their teaching quality.

Finally, administrators can focus on these predictors while they are planning in-service training programs for instructors. Professional development units can arrange seminars, talks, or workshops to improve instructors in those areas.

Implications for Further Research

This study aimed to find the significant predictors of student satisfaction with the instructors. The most significant predictors are found to be effectiveness and being supportive. However, effectiveness is a term which might slightly change depending on students' preferences. To gain a better insight into what *effectiveness* and *efficiency* exactly mean for students, a more detailed study should be conducted

to explore students' perspectives on this matter. If students share their opinions on this matter, instructors can improve their teaching techniques accordingly.

Moreover, as student evaluation forms were anonymous in this study, analyses were conducted only on course type, proficiency level of students, and the time spent in the preparatory school. However, if demographics of these students were collected, this study could include more detailed analyses regarding student characteristics such as department and gender. Therefore, a more comprehensive study might be conducted including student demographics in the future.

Furthermore, survey questions should not be double-barreled, which means they should ask only one thing in each question. However, there were some double-barreled questions in the SET survey that was used in this research. Hence, another study can be conducted to detect these double-barreled questions and to improve the validity of this survey.

Limitations

This research is conducted in only one higher education institution. The SET form used in this institution is not used in any other institution, which can limit the extent to which the results can be generalized. However, most SET forms which are currently used all over the world measure similar course and teaching characteristics. Therefore, the results still shed some light on the predictors of student satisfaction with instructors.

Additionally, the curriculum is updated every year to improve teaching quality in the preparatory school where this study is conducted, and new instructors join the teaching staff at the beginning of every academic year. However, these changes might affect the results of the SET form used in the target institution. Therefore, to eliminate or minimize this problem as much as possible, only one

academic year's SET results were utilized in this study. This ensures that the teaching staff did not change, and the curriculum was the same all year round when the data were collected.

Furthermore, this study utilizes the data which were self-reported by students. This is a limitation because students might have had certain biases while evaluating their instructors, which can slightly affect the reliability of the results. Furthermore, according to the grading leniency hypothesis, SET results might be manipulated by using different means. However, as the number of student evaluation forms was very high, those biasing factors are not expected to affect the overall results significantly.

Moreover, the number of nodes in each decision tree analysis is directly related with sample size. As the whole sample size was large, the number of nodes and the number of predictors of student satisfaction are higher than the ones in the other analyses in this study. Therefore, if student numbers had been different, different results could have been obtained. This is a limitation, but the findings have shown that there are still common predictors which are repeatedly found in different analyses including different student numbers. Therefore, this study still provides a highly accurate picture of what the predictors of student satisfaction are in a preparatory school at a university.

What is more, when the SET forms were administered, there were some students who were absent because students tend to use their absenteeism at the end of trimesters. Therefore, although there are missing values in the data set, there is no clear pattern for these missing responses, which makes the missing values random.

In addition, in this study independence of observation is violated because each student filled out 3 different forms in one trimester for each of his/her instructors. Moreover, A, AF, and B level students who successfully completed the

1st trimester, continued studying in the preparatory school in the 2nd and 3rd trimesters. However, as the number of forms that were filled is remarkably high, this violation is negligible.

Furthermore, the SET form that was used in this study was created by the measurement and evaluation office of the preparatory school where this study took place. The researcher did not have the chance to revise or change some impure questions such as question number 1 (She is positive, lively, and enthusiastic in the lesson). This question evaluates three different qualities of the instructors, and it may be misleading for students. However, as the researcher did not have any authority on the preparation or administration of this SET form, it was impossible for her to improve these questions.

Finally, these findings were discovered in an English language preparatory school. Although they can be used by other preparatory schools, it might be difficult to generalize them for departmental courses in universities.

REFERENCES

- Adams, S., Bekker, S., Fan, Y., Gordon, T., Shepherd, L. J., Slavich, E., & Waters, D. (2021). Gender bias in student evaluations of teaching: 'Punish[ing] those who fail to do their gender right'. *Higher Education*, 83, 787-807. <https://doi.org/10.1007/s10734-021-00704-9>
- Al-Maamari, F. (2015). Response rate and teaching effectiveness in institutional student evaluation of teaching: A multiple linear regression study. *Higher Education Studies*, 5(6), 9-20. <https://doi.org/10.5539/hes.v5n6p9>
- Al-Issa, A., & Sulieman, H. (2007). Student evaluations of teaching: Perceptions and biasing factors. *Quality Assurance in Education*, 15(3), 302-317. <https://doi.org/10.1108/09684880710773183>
- Arnold, I. J., & Versluis, I. (2019). The influence of cultural values and nationality on student evaluation of teaching. *International Journal of Educational Research*, 98, 13-24. <https://doi.org/10.1016/j.ijer.2019.08.009>
- Arreola, R. A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems* (3rd ed.). Anker Pub. Co.
- Barnett, C. W., Matthews, H. W., & Jackson, R. A. (2003). A comparison between student ratings and faculty self-ratings of instructional effectiveness. *American Journal of Pharmaceutical Education*, 67(4), 884-889. <https://doi.org/10.5688/aj6704117>
- Barnett, K. (2019). *Characteristics of high quality teachers: A qualitative phenomenological study* (Publication No. 3541) [Doctoral dissertation, East Tennessee State University]. ETSU Campus Repository. <https://dc.etsu.edu/etd/3541/>

- Barrantes, L. (2022). A humanistic perspective for the mixed-proficiency language class. *Mextesol Journal*, 46(2), 1-7. <https://doi.org/10.61871/mj.v46n2-10>
- Barth, M. M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business*, 84(1), 40-46. <https://doi.org/10.3200/joeb.84.1.40-46>
- Basow, S. A., Codos, S., & Martin, J. L. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, 47(2), 352-363.
- Benton, S. L., & Cashin, W. E. (2014). Student ratings of instruction in college and university courses. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (pp. 279-326). Springer. <https://doi.org/10.1007/978-94-017-8005-6>
- Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment & Evaluation in Higher Education*, 30(6), 593-601. <https://doi.org/10.1080/02602930500260688>
- Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research & Development*, 18(1), 57-75. <https://doi.org/10.1080/0729436990180105>
- Briceno, C. E. (2020). *The relationship between student grades and student evaluations of professors and its impact on faculty practices* (Publication No. 28964387) [Doctoral dissertation, Saint Peter's University]. ProQuest Dissertations & Theses Global.
- Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the grading leniency story to the edge. The influence of student, teacher, and course characteristics

on student evaluations of teaching in higher education. *Educational Assessment, Evaluation and Accountability*, 23(4), 289-306. <https://doi.org/10.1007/s11092-011-9126-2>

Burroughs, N., Gardner, J., Lee, Y., Guo, S., Touitou, I., Jansen, K., & Schmidt, W. (2019). *Teaching for excellence and equity: Analyzing teacher characteristics, behaviors and student outcomes with TIMSS*. Springer Open. <https://doi.org/10.1007/978-3-030-16151-4>

Brown, G., & Atkins, M. (2002). *Effective teaching in higher education*. Routledge.

Cakici, D. (2017). An investigation of learner autonomy in Turkish EFL context. *International Journal of Higher Education*, 6(2), 89-99. <https://doi.org/10.5430/ijhe.v6n2p89>

Cattell, R. B. (1931). The assessment of teaching ability: A survey of professional opinion on the qualities of a good teacher. *British Journal of Educational Psychology*, 1(1), 48-72. <https://doi.org/10.1111/j.2044-8279.1931.tb03222.x>

Cho, J. I., Hamash, K., Otani, K., Reimer, N., & Deng, Y. (2018). Influential factors of student evaluations of teaching in a nursing program. *Teaching and Learning in Nursing*, 13(2), 86-94. <https://doi.org/10.1016/j.teln.2018.01.002>

Chung, J. H., & Wei, X. (2020). Teach effectively or guide wisely? Discussing the application of wisdom approach to language teaching in Thai higher education. *International Journal of Language Education*, 4(3), 322-333. <https://doi.org/10.26858/ijole.v4i3.15097>

Clayson, D. (2022). The student evaluation of teaching and likability: What the evaluations actually measure. *Assessment & Evaluation in Higher Education*, 47(2), 313-326. <https://doi.org/10.1080/02602938.2021.1909702>

- Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? *Journal of Marketing Education*, 31(1), 16-30.
<https://doi.org/10.1177/0273475308324086>
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research*, 51(3), 281-309. <https://doi.org/10.2307/1170209>
- Cox, S. R., Rickard, M. K., & Lowery, C. M. (2022). The student evaluation of teaching: Let's be honest – who is telling the truth? *Marketing Education Review*, 32(1), 82-93. <https://doi.org/10.1080/10528008.2021.1922924>
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education.
- Crumbley, D. L., & Reichelt, K. J. (2009). Teaching effectiveness, impression management, and dysfunctional behavior Student evaluation of teaching control data. *Quality Assurance in Education*, 17(4), 377-392. <https://doi.org/10.1108/09684880910992340>
- Cui, C. (2021). *Failed accountability and student evaluations of teaching in higher education: An experimental study* (Publication No. 29133385) [Doctoral dissertation, The University of Toledo]. ProQuest Dissertations & Theses Global.
- Çimen, O. (2010). *Methods of lecturers' instruction performance evaluation in higher education institutions: The evaluation of lecturer and course evaluation questionnaires* (278029) [Master's thesis]. YÖK Ulusal Tez Merkezi.

- Darmanin, M. (2020). *Democratic citizenship education: Guidelines and recommendations for implementation in higher education* (ED614458). European Students' Union. <https://eric.ed.gov/?id=ED614458>
- Daumiller, M., Janke, S., Hein, J., Rinas, R., Dickhäuser, O., & Dresel, M. (2022). Teaching quality in higher education: Agreement between teacher self-reports and student evaluations. *European Journal of Psychological Assessment*, 39(3), 176-181. <https://doi.org/10.31234/osf.io/9yw7f>
- De Ville, B., & Neville, P. (2013). *Decision trees for analytics using SAS enterprise miner*. SAS Institute.
- Delaney, J., Johnson, A., Johnson, T., & Treslan, D. (2010). *Students' perceptions of effective teaching in higher education*. Memorial University Distance Education and Learning Technologies. <https://research.library.mun.ca/8370/>
- Feistauer, D., & Richter, T. (2018). Validity of students' evaluations of teaching: Biasing effects of likability and prior subject interest. *Studies in Educational Evaluation*, 59, 168-178. <https://doi.org/10.1016/j.stueduc.2018.07.009>
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *National Academy of Sciences*, 111(23), 8410-8415. <https://doi.org/10.1073/pnas.1319030111>
- Gelber, S. M. (2020). *Grading the college: A history of evaluating teaching and learning*. Johns Hopkins University Press.
- Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis* (S283B050051). National Comprehensive Center for Teacher Quality. <https://gtlcenter.org/sites/default/files/docs/LinkBetweenTQandStudentOutcomes.pdf>

- Good, T. L., Wiley, C. R. H., & Florez, I. R. (2009). Effective teaching: an emerging synthesis. In L. J. Saha & A. G. Dworking (Eds.), *International handbook of research on teachers and teaching* (pp. 803-816). Springer. https://doi.org/10.1007/978-0-387-73317-3_51
- Haris, S., Jamil, B., Haris, M., Deebea, F., Khan, M. J., & Khan, I. Z. (2022). Factors affecting students' perception towards faculty evaluation of teaching at Nowshera Medical College, Pakistan. *The Professional Medical Journal*, 29(02), 258-264. <https://doi.org/10.29309/tpmj/2022.29.02.6407>
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hattie, J. A. C. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. <https://doi.org/10.4324/9780203887332>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hoefer, P., Yurkiewicz, J., & Byrne, J. C. (2012). The association between students' evaluation of teaching and grades. *Decision Sciences Journal of Innovative Education*, 10(3), 447-459. <https://doi.org/10.1111/j.1540-4609.2012.00345.x>
- Hu, C. C. (2020). Understanding college students' perceptions of effective teaching. *International Journal of Teaching and Learning in Higher Education*, 32(2), 318-328.
- Hummel, K. M. (2014). *Introducing second language acquisition: Perspectives and practices*. John Wiley & Sons.
- Ibad, F., & Sharjeel, M. Y. (2021). Evaluation of the professional competency of university faculty members as perceived by higher education students.

Journal of Education and Educational Development, 8(1), 141-159.

<https://doi.org/10.22555/joeed.v8i1.372>

Jones, B. D., Miyazaki, Y., Li, M., & Biscotte, S. (2022). Motivational climate predicts student evaluations of teaching: Relationships between students' course perceptions, ease of course, and evaluations of teaching. *AERA Open*, 8(1), 1-17. <https://doi.org/10.1177/23328584211073167>

Joye, S. W., & Wilson, J. H. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning*, 15(4), 126-138. <https://doi.org/10.14434/josotl.v15i4.13466>

Kahn, P., & Anderson, L. (2019). *Developing your teaching: Towards excellence* (2nd ed.). Routledge.

Kalay, D. (2017). What makes an effective EFL teacher? Investigations on student and teacher perspectives. *Turkish Studies*, 12(25), 421-444. <https://doi.org/10.7827/TurkishStudies.12187>

Kalender, İ. (2011). Contaminating factors in university students' evaluation of instructors. *Education and Science*, 36(162), 56-65.

Kalender, İ. (2013). Profiling instructional effectiveness to reveal its relationship to learning. *The Asia-Pacific Education Researcher*, 23(3), 717-726. <https://doi.org/10.1007/s40299-013-0145-2>

Karaca, R. C. (2022). *Dimensions of student evaluation of teaching forms at university level* (Publication No. 30158309) [Master's thesis, İhsan Doğramacı Bilkent University]. ProQuest Dissertations & Theses Global.

Karakaya, I., Sata, M., Corbaci, E. C., & Cetin, B. (2018). An investigation of the factors that affect high school students' attitudes towards social media by CHAID analysis. *Eurasian Journal of Educational Research*, 76, 21-40.

- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29(2), 119-127.
<https://doi.org/10.2307/2986296>
- Khandelwal, K. A. (2009). Effective teaching behaviors in the college classroom: A critical incident technique from students' perspective. *International Journal of Teaching and Learning in Higher Education*, 21(3), 299-309.
- Kirk, R. E. (2008). *Statistics: An introduction* (5th ed.). Thomson Wadsworth.
- Ko, J., Sammons, P., & Bakkum, L. (2013). *Effective teaching: A review of research and evidence* (ED546794). ERIC. <https://eric.ed.gov/?id=ED546794>
- Kulik, J.A. (2001), Student ratings: Validity, utility, and controversy. *New Directions for Institutional Research*, 2001(109), 9-25. <https://doi.org/10.1002/ir.1>
- Kurniati, E., & Cahyono, B. Y. (2018). Students' belief on using certain learning strategies in improving their speaking ability. *People: International Journal of Social Sciences*, 4(2), 1572-1579.
<https://doi.org/10.20319/pijss.2018.42.15721579>
- Latif, E., & Miles, S. (2013). Students' perception of effective teaching. *Journal of Economics and Economic Education Research*, 14(1), 121-129.
- Lei, H., Cui, Y., & Chiu, M. M. (2018). The relationship between teacher support and students' academic emotions: A meta-analysis. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.02288>
- León, J., Núñez, J. L., & Liew, J. (2015). Self-determination and STEM education: Effects of autonomy, motivation, and self-regulated learning on high school math achievement. *Learning and Individual Differences*, 43, 156-163.
<https://doi.org/10.1016/j.lindif.2015.08.017>

- Li, L., & Heydarnejad, T. (2024). May I come in? A probe into the contributions of self-esteem, teacher support, and critical thinking to anxiety and shyness in language classes. *BMC Psychology*, 12(7), 1-14.
<https://doi.org/10.1186/s40359-023-01501-y>
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106.
<https://doi.org/10.1016/j.stueduc.2016.12.004>
- Liu, S., Keeley, J., & Buskist, W. (2016). Chinese college students' perceptions of excellent teachers across three disciplines: psychology, chemical engineering, and education. *Teaching of Psychology*, 43(1), 70-74.
<https://doi.org/10.1177/0098628315620888>
- Lumpkin, A., & Multon, K. D. (2013). Perceptions of teaching effectiveness. *The Educational Forum*, 77(3), 288-299.
<https://doi.org/10.1080/00131725.2013.792907>
- Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52(1), 77-95.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-384). Springer.
https://doi.org/10.1007/1-4020-5742-3_9

- McPherson, M. A. (2006). Determinants of how students evaluate teachers. *The Journal of Economic Education*, 37(1), 3-20.
<https://doi.org/10.3200/jece.37.1.3-20>
- Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press. <https://doi.org/10.1017/9781009024563>
- Merrill, M. D. (2018). Using the first principles of instruction to make instruction effective, efficient, and engaging. In R. E. West (Ed.), *Foundations of learning and instructional design technology* (pp. 265-273). EdTech Books. <https://doi.org/10.59668/3>
- Mohammed, T. A., & Pandhiani, S. M. (2017). Analysis of factors affecting student evaluation of teaching effectiveness in Saudi higher education: The case of Jubail University College. *American Journal of Educational Research*, 5(5), 464-475. <https://doi.org/10.12691/education-5-5-2>
- Morgan, D. A., Sneed, J., & Swinney, L. (2003). Are student evaluations a valid measure of teaching effectiveness: Perceptions of accounting faculty members and administrators. *Management Research News*, 26(7), 17-32.
<https://doi.org/10.1108/01409170310783565>
- Nadler, M. K., & Nadler, L. B. (2000). Out of class communication between faculty and students: A faculty perspective. *Communication Studies*, 51(2), 176-188.
<https://doi.org/10.1080/10510970009388517>
- Okpala, C. O., & Ellis, R. (2005). The perceptions of college students on teacher quality: A focus on teacher qualifications. *Education*, 126(2), 374-383.
<https://eric.ed.gov/?id=EJ765687>

- Ory, J. C., Braskamp, L. A., & Pieper, D. M. (1980). Congruency of student evaluative information collected by three methods. *Journal of Educational Psychology*, 72(2), 181-185. <https://doi.org/10.1037//0022-0663.72.2.181>
- Önder, E., & Uyar, Ş. (2017). CHAID analysis to determine socioeconomic variables that explain students' academic success. *Universal Journal of Educational Research*, 5(4), 608-619. <https://doi.org/10.13189/ujer.2017.050410>
- Özgüngör, S., & Duru, E. (2015). Course and instructor characteristics distinguishing highest and lowest student ratings of instructors. *Eurasian Journal of Educational Research*, (61), 118-136. <https://doi.org/10.14689/ejer.2015.61.7>
- Park, E., & Dooris, J. (2020). Predicting student evaluations of teaching using decision tree analysis. *Assessment & Evaluation in Higher Education*, 45(5), 776-793. <https://doi.org/10.1080/02602938.2019.1697798>
- Passos, A. F. (2009). *A comparative analysis of teacher competence and its effect on pupil performance in upper primary schools in Mozambique and other SACMEQ countries* [Doctoral dissertation, University of Pretoria]. University of Pretoria Campus Repository. <https://repository.up.ac.za/bitstream/handle/2263/28160/00front.pdf?sequence=1&isAllowed=y>
- Patrick, J., & Smart, R. M. (1998). An empirical evaluation of teacher effectiveness: The emergence of three critical factors. *Assessment & Evaluation in Higher Education*, 23(2), 165-178. <https://doi.org/10.1080/0260293980230205>
- Polio, C., & Friedman, D. (2016). *Understanding, evaluating, and conducting second language writing research*. Routledge. <https://doi.org/10.4324/9781315747293>

- Pounder, J. S. (2007). Is student evaluation of teaching worthwhile? *Quality Assurance in Education*, 15(2), 178-191.
<https://doi.org/10.1108/09684880710748938>
- Pourjamal Ghouyjagh, H., Sotoudehnama, E., & Faghih, E. (2018). On the impact of teaching experience on EFL instructors' self-assessment of their instructional effectiveness. *Iranian Journal of English for Academic Purposes*, 6(1), 47-70.
- Raymond, S. M. (2008). *Effective and ineffective university teaching from the students' and faculty's perspectives: Matched or mismatched expectations?* [Doctoral dissertation, University of Exeter]. University of Exeter Campus Repository.
<https://ore.exeter.ac.uk/repository/handle/10036/40767?show=full>
- Remedios, R., & Lieberman, D. A. (2006). I liked your course because you taught me well: The influence of grades, workload, expectations, and goals on students' evaluations of teaching. *British Educational Research Journal*, 34(1), 91-115. <https://doi.org/10.1080/01411920701492043>
- Remmers, H. H., & Brandenburg, G. C. (1927). Experimental data on the Purdue rating scale for instruction. *Educational Administration and Supervision*, 13, 519-527.
- Renandya, W. A., & Jacobs, G. M. (2016). Extensive reading and listening in the L2 classroom. In W. A. Renandya & P. Handoyo (Eds.), *English language teaching today* (pp. 97-110). Routledge.
- Ripoll-Núñez, K. J., Mojica-Ospina, I. E., Torres-Riveros, A. C., & Castellanos-Tous, M. S. (2018). Teachers' and students' perceptions of

- excellence in teaching in Colombia. *New Directions for Teaching and Learning*, 2018(156), 57-65. <https://doi.org/10.1002/tl.20317>
- Schneider, G. (2013). Student evaluations, grade inflation and pluralistic teaching: Moving from customer satisfaction to student learning and critical thinking. *Forum for Social Economics*, 42(1), 122-135. <https://doi.org/10.1080/07360932.2013.771128>
- Sayavedra, M. (2014). Teacher evaluation. *ORTESOL Journal*, 31, 1-9.
- Seldin, P. (1993). How colleges evaluate professors: 1983 versus 1993. *AAHE Bulletin*, 12, 6-8.
- Semerena, R. E. (2021). The promise of education: The future. In H. V. Land, A. Corcoran, & D. C. Iancu (Eds.), *The promise of higher education: Essays in honour of 70 years of IAU* (pp. 321-323). Springer. <https://doi.org/10.1007/978-3-030-67245-4>
- Shah, M., & Sid Nair, C. (2012). The changing nature of teaching and unit evaluations in Australian universities. *Quality Assurance in Education*, 20(3), 274-288.
- Shevlin, M., Banyard, P., Davies, M. N.O., & Griffiths, M. D. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, 25(4), 397-405. <https://doi.org/10.1080/713611436>
- Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, 22(3), 199-213. <https://doi.org/10.1177/0273475300223004>
- Smele, S., Quinlan, A., & Lacroix, E. (2021). Engendering inequities: Precariously employed academic women's experiences of student evaluations of

teaching. *Gender and Education*, 33(8), 966-982.

<https://doi.org/10.1080/09540253.2021.1884194>

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642. <https://doi.org/10.3102/0034654313496870>

Spooren, P., & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: Will better teachers receive higher ratings and will better students give higher ratings? *Educational Studies*, 32(2), 201-214. <https://doi.org/10.1080/03055690600631101>

Stehle, S., & Spinath, B. (2014). Intended course objectives and perception of teaching effectiveness. *Psychology Learning & Teaching*, 13(3), 205-217. <https://doi.org/10.2304/plat.2014.13.3.205>

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, 42(4), 276-294. <https://doi.org/10.1080/01973533.2020.1756817>

Su, F., & Wood, M. (2012). What makes a good university lecturer? Students' perceptions of teaching excellence. *Journal of Applied Research in Higher Education*, 4(2), 142-155. <https://doi.org/10.1108/17581181211273110>

Subtirelu, N. C. (2015). "She does have an accent but...": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society*, 44(1), 35-62. <https://doi.org/10.1017/s0047404514000736>

Sweeney, A. D., Morrison, M. D., Jarratt, D., & Heffernan, T. (2009). Modeling the constructs contributing to the effectiveness of marketing lecturers. *Journal of*

Marketing Education, 31(3), 190-202.

<https://doi.org/10.1177/0273475309345198>

Taye, M. T., Muthanna, A., Sang, G., & Alduais, A. (2022). Understanding effective teaching beliefs of instructors and students: A qualitative Study at an ethiopian University. *Athens Journal of Education*, 9(4), 679-698.

<https://doi.org/10.30958/aje.9-4-8>

Tsai, K. C., & Lin, K. (2012). Rethink student evaluation of teaching. *World Journal of Education*, 2(2), 17-22. <https://doi.org/10.5430/wje.v2n2p17>

Uttl, B. (2021). Lessons learned from research on student evaluation of teaching in higher education. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools* (pp. 237-259). Springer.

<https://doi.org/10.1007/978-3-030-75150-0>

Wang, G., & Williamson, A. (2022). Course evaluation scores: Valid measures for teaching effectiveness or rewards for lenient grading? *Teaching in Higher Education*, 27(3), 297-318. <https://doi.org/10.1080/13562517.2020.1722992>

Wei, W., & Hui, S. K. (2019). Evaluating teacher performance in language learning classes: The gap between students and department administrators. *Teachers and Teaching*, 25(4), 486-500.

<https://doi.org/10.1080/13540602.2019.1625763>

Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating teaching in higher education. *The Journal of Economic Education*, 40(3), 227-261.

<https://doi.org/10.3200/JECE.40.3.227-261>

Whitaker, T. (2020). *What great teachers do differently: Nineteen things that matter most* (3rd ed.). Routledge.

- Wilson, D. M., Summers, L., & Wright, J. (2020). Faculty support and student engagement in undergraduate engineering. *Journal of Research in Innovative Teaching & Learning*, 13(1), 83-101. <https://doi.org/10.1108/jrit-02-2020-0011>
- Wilson, J. H., Beyer, D., & Monteiro, H. (2014). Professor age affects student ratings: Halo effect for younger teachers. *College Teaching*, 62, 20-24. <https://doi.org/10.1080/87567555.2013.825574>
- Yılmaz, E. (2018). *Differences in perceptions of effective teaching by faculty member's gender and student characteristics in a Turkish public university* [Master's thesis, Middle East Technical University]. METU Campus Repository. <https://open.metu.edu.tr/handle/11511/27616>
- You, Z., Si, Y. W., Zhang, D., Zeng, X., Leung, S. C., & Li, T. (2015). A decision-making framework for precision marketing. *Expert Systems with Applications*, 42(7), 3357-3367. <https://doi.org/10.1016/j.eswa.2014.12.022>
- Yu, Y., Lin, Y., Qi, J., & Yan, H. (2022). Biases in student evaluations of teaching: An exploratory analysis to a Chinese case study. *Innovations in Education and Teaching International*, 1-12. <https://doi.org/10.1080/14703297.2022.2145985>
- Yüner, B. (2023). Examination of the variables explaining school effectiveness by CHAID analysis. *Journal of Theoretical Educational Science*, 16(4), 808-820. <https://doi.org/10.30831/akukeg.1294397>
- Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education*, 37(2), 227-235. <https://doi.org/10.1080/02602938.2010.523819>

Zhao, L., Xu, P., Chen, Y., & Yan, S. (2022). A literature review of the research on students' evaluation of teaching in higher education. *Frontiers in Psychology*, 13, 1-9. <https://doi.org/10.3389/fpsyg.2022.1004487>

Zumbrunn, S., McKim, C., Buhs, E., & Hawley, L. R. (2014). Support, belonging, motivation, and engagement in the college classroom: A mixed method study. *Instructional Science*, 42(5), 661-684. <https://doi.org/10.1007/s11251-014-9310-0>



Appendix A

The Turkish Version of the Questions of the SET Form Used in This Study

1. Derslerde pozitif, canlı ve heveslidir.
2. Kendisi ile rahatlıkla iletişim (e-mail, vb.) kurulabilir.
3. Öğrenme sürecimde destekleyici ve teşvik edici olmuştur.
4. Sınıfa hâkimdir.
5. Derslerin etkili ve verimli bir biçimde yürütülmesini sağlar.
6. Derslerin içeriğine hâkimdir.
7. İhtiyaç duyduğumda bana yararlı olabilecek çalışmalar tavsiye eder ve gerekli yönlendirmeleri yapar.
8. Belirtilen ofis saatlerinde kendisine ulaşılabilir.
9. Tüm öğrencilerin derse aktif katılımını teşvik eder.
10. Derslerde yapılan etkinlikler çeşitlilik gösterir.
11. Derslere zamanında girip çıkar.
12. Derslere planlı ve hazırlıklı gelir.
13. Ders süresini sonuna kadar etkili kullanır.
14. Dersi İngilizce işlemeye özen gösterir (TOEFL dersleri hariç).
15. Görsel araçları (tahta, projeksiyon, internet, resim, video, e-book, ders kitabı vb.) etkin kullanır.
16. Öğrenme sürecimi destekleyecek görsel ve basılı yayınları (film-dizi, belgesel, gazete, dergi, kitap, online kaynaklar vb.) tavsiye eder.
17. Program dışı etkinliklere (English Friends, Foodie Club vb.) katılımı teşvik eder.
18. Sınıf kurallarını (ders giriş-çıkış saatleri, cep telefonu kullanımı, ders materyalleri vb.) titizlikle uygular.
19. Derste geçen bir konu ya da kavram ile ilgili görüşlerimi belirtmemi gerektiren sorular sorar.
20. Sorduğu sorular ile farklı düşünceleri değerlendirmemi sağlar.
21. Derste bahsedilen konularla ilgili yorum yapmamı teşvik eder.
22. Ders içindeki performansım ile ilgili bana geribildirim verir.
23. Bu öğretim elemanından ders aldığım için memnunum.