BUSE YETER

2023

PREDICTING MIDDLE SCHOOL STUDENTS' MATHEMATICS ACHIEVEMENTS USING DECISION TREES

BUSE YETER

A THESIS SUBMITTED

FOR

THE DEGREE OF MASTER OF ARTS

IN

CURRICULUM AND INSTRUCTION

İHSAN DOĞRAMACI BILKENT UNIVERSITY

ANKARA

SEPTEMBER 2023

To my beloved family and friends...

Predicting Middle School Students’ Mathematics Achievement Using Decision Trees

The Graduate School of Education

of

İhsan Doğramacı Bilkent University

by

Buse Yeter

In Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in

Curriculum and Instruction

Ankara

September 2023

İHSAN DOĞRAMACI BILKENT UNIVERSITY

GRADUATE SCHOOL OF EDUCATION

Predicting Middle School Students' Mathematics Achievement Using Decision Trees

Buse Yeter

August 2023

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 Curriculum & Instruction.

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 Curriculum & Instruction.

Assoc. Prof. Dr. Jennie F. 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 Curriculum & Instruction.

Asst. Prof. Dr. Zerrin Toker, TED University (Examining Committee Member)

Approval of the Graduate School of Education

Prof. Dr. Orhan Arıkan (Director)

ABSTRACT

PREDICTING MIDDLE SCHOOL STUDENTS' MATHEMATICS ACHIEVEMENTS USING DECISION TREES

Buse Yeter

MA in Curriculum and Instruction

Advisor: Asst. Prof. Dr. İlker Kalender

September 2023

In today's world, it is crucial to raise individuals who can think critically and flexibly, see things from a variety of perspectives, and develop creative solutions to common challenges in many countries including Türkiye. One of the ways to attain this purpose is to provide qualified mathematics education. During this qualified education determining the extent to which students meet the program goals is an important and fundamental part of the process of educating the individuals. Education should be provided in accordance with the requirements of the age using a curriculum with qualified measurement, assessment, and evaluation practices. This study aims to predict middle school students' achievement at four grade levels in mathematics at the end of the semester utilized as a predictor measurement and assessment approach. Chi-squared Automatic Interaction Detection (CHAID), a decision tree analysis, was used to examine the importance of exams, quizzes, homework, and teacher observation in predicting the end-of-semester achievements. As a result of the study, when the first semester is examined in general, quizzes are mostly determined as the best predictor of student success, while exams are generally for the second semester. Implications for teachers and students are stated.

Keywords: Mathematics achievement, prediction of student achievement, measurement, assessment, summative assessment, formative assessment

ÖZET

Karar Ağaçları Kullanarak Ortaokul Öğrencilerinin Matematik Başarılarının Yordanması

Buse Yeter

Eğitim Programları ve Öğretim Yüksek Lisans Programı

Danışman: Dr. Öğr. Üyesi İlker Kalender

Eylül 2023

Günümüzde, Türkiye de dahil olmak üzere pek çok ülkede, eleştirel ve esnek düşünebilen, olaylara farklı açılardan bakabilen, ortak sorunlara yaratıcı çözümler geliştirebilen bireyler yetiştirmek büyük önem taşıyor. Bu amaca ulaşmanın yollarından biri de nitelikli matematik eğitimi vermektir. Bu nitelikli eğitim sürecinde öğrencilerin program hedeflerini ne ölçüde karşıladığının belirlenmesi, birey yetiştirme sürecinin önemli ve temel bir parçasıdır. Nitelikli ölçme ve değerlendirme uygulamalarına sahip bir müfredat kullanılarak çağın gereklerine uygun eğitim verilmelidir. Bu çalışma, yordayıcı bir ölçme ve değerlendirme yaklaşımı olarak kullanılan dört sınıf düzeyindeki ortaokul öğrencilerinin eğitim-öğretim yılının iki yarıyılı sonundaki matematik başarılarını tahmin etmeyi amaçlamaktadır. Öğrencilerin dönem sonu başarılarını tahmin etmede sınavların, kısa sınavların, ödevlerin ve öğretmen gözleminin önemini incelemek için karar ağacı analizi olan Chi-squared Automatic Interaction Detection (CHAID) kullanılmıştır. Çalışmanın sonucunda birinci dönem genel olarak incelendiğinde kısa sınavların öğrenci başarısının en iyi yordayıcısı olduğu, sınavların ise genellikle ikinci dönemin en iyi yordayıcısı olduğu tespit edilmiştir. Öğretmenler ve öğrenciler için çıkarımlar belirtilmiştir.

Anahtar kelimeler: Matematik başarısı, öğrenci başarısının yordanması, ölçme, değerlendirme, biçimlendirici değerlendirme, özetleyici değerlendirme

ACKNOWLEDGEMENTS

I would like to express my gratitude to those who supported me during my master's degree process. First, I want to thank my advisor Asst. Prof. Dr. İlker Kalender for his effort to assist me in patience and devote his time throughout the process of writing this thesis. I am incredibly grateful for his suggestions and support. Additionally, I am thankful to the committee members; Asst. Prof. Dr. Jennie F. Lane and Asst. Prof. Dr. Zerrin Toker, for taking the time to read and give valuable feedback to improve my thesis.

I also feel so lucky to have met Prof. Dr. Margaret K. Sands and Dr. M. Sencer Çorlu on my journey to becoming a teacher.

My grandfather Turan Turgut has always been a role model and inspiration for me both in my academic life and in my life journey. I am grateful for his existence and for being an idol to me.

The most thanks from my heart are to my husband İlker Yeter for his patience and support. The greatest appreciation to my parents, Hülya Erbaş and Nüvit Erbaş and my brother Burak Erbaş. Thank you for your unwavering support and for providing all educational opportunities for me.

Finally, I would like to thank all my friends who were with me with their emotional support throughout the process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Introduction

Galileo Galilei called to scientists and philosophers “Count what is countable, measure what is measurable and what is not measurable, make measurable” (Finkelstein, 1982). Measurement, and the other two crucial and highly related concepts of assessment and evaluation, are used in many areas around the world to make informed decisions about individuals, institutions, and other entities. These decisions include, for example, whether a person passes a driver's test, if someone is qualified to become a public servant, a student's entitlement to enter university, determining the language or academic skill level, and other entities.

Measurement can be defined as determining the quality or quantity of any object (Jaworski et al., 2014). Alternatively, measurement can be defined as assigning numbers or symbols to what is being measured. Assessment is a methodical activity utilized by teachers and students to acquire data, analyze it, interpret it, reach conclusions, make sensible judgments, and take appropriate action in the interest of enhancing teaching and learning. (Berry, 2008). Newton (2007) stated that assessment is used for many different purposes, including assigning grades to students, ensuring national accountability, monitoring systems, allocating resources within a district, monitoring, or placing students, deciding interventions, enhancing teaching, and learning, and giving specific feedback to students and their parents. Finally, evaluation is about making judgments based on the data collected (Ebel & Frisbie, 1991). Evaluation can be formal or informal. A real-life example of informal evaluation is: The sky is grey; it may be raining. A formal evaluation can be made based on information collected from measurement and assessment (Rea-Dickins & Germaine, 1992).

Measurement and assessment activities in schools are generally more difficult and abstract than real-life applications. According to Terman's research (1921) about intelligence and its measurement if the ability to think abstractly is what intelligence is, then tests of intelligence that include the use of language or other symbols are likely to be the most accurate. Another proposal from Terman is that an intelligence test be closely related to "school educability." These include assessments on mathematical reasoning, language completion, and comprehension of challenging passages (Terman, 1921). This difficulty of measurement and assessment in the schools is because the variables such as whether someone is qualified to enter the university or determining the language or academic skill level to be measured such as mind, intelligence, success, interest, and relevance cannot be observed directly. These observations are needed to assess quality or quantity, the outcomes of which help make informed decisions, or evaluations. Ironically, education is one of the most important fields where measurement, assessment, and evaluation are inevitably used. According to Ebel and Frisbie (1991), educational measurement involves assigning a number, a symbol, or any other quantity or quality indicator for the variables such as achievement, interests, attitudes, and aptitudes of students. Educational assessment is systematic collection and documentation from students using oral or verbal exams. And, finally, educational evaluation is a process of judging students' progress (Ebel & Frisbie, 1991). According to Romberg (1992), educational evaluation has changed in time from its exclusive focus on measuring student accomplishment to developing judgments about students to its present, rising interest in providing data to support educational programs and policy decision making. In educational settings, these three terms are used to describe activities or procedures such as giving grades, deciding whether to take the next class or the next course, putting students into different levels of class, and determining the readiness level of the students.

There are various tools used to collect data for observation, analysis, and interpretation of how and how much students learn in education settings. Among them are exams, quizzes, exit cards, homework, projects, presentations, external international exams, and other entities. By using these tools, teachers collect information about, or measure, students' progress during different stages of instruction. Although teachers may have many tools, the selection of measurement and assessment tools and the subsequent evaluation should be made considering several conditions such as students' needs, characters, subjects, topics, or duration.

There are a variety of tools teachers use to measure student performance and assess their knowledge. Exit cards are used at the end of the lesson to assess the knowledge learned in the lesson based on students' needs. The quizzes are used during teaching or, at the end of the topic, to assess the information that includes a few topics and more recent knowledge. Projects are tools that students work on throughout the semester, individually or in groups, and deliver a product at the end of the process. Students are provided with a variety of topics for their interests and characters, and they have the right to choose the subject and topic of the project they will do. Another example of a tool for assessment is more comprehensive exams are applied to assess more subjects and general knowledge. The application times of the exam, quiz, and exit cards also change by decreasing, respectively.

Teachers use the data from different instruments to make a judgment or evaluation about students. According to Rea-Dickins and Germaine (1992), evaluation in practice employs a variety of criteria and takes into consideration a wide range of elements generated from different teaching and learning contexts. The following inquiries are considered in relation to these standards: Why is an evaluation necessary, what will it be used for, and how will it be organized and managed?

The ultimate goal of all measurement, assessment, and evaluation practices is to make informed decisions about students, most of the time, about their achievement levels in different subject areas, using the tools and approaches given above. Quality information from different components may help teachers in the process of structuring the lesson plan by improving goals and objectives, providing extra support for students, informing students and parents, and increasing student engagement and confidence. To this end, having several predictors of achievement may also help teachers with the decisions that are made at the end of the year.

Quality mathematics education is one of the areas in which measurement, assessment and evaluation are of vital importance. Qualified teachers using a well-designed curriculum deliver a qualified education in accordance with the necessities of time. Measurement and assessment, which establish the degree to which students achieve the program's goals, and evaluation which establish judgement based on collected data are crucial components of any program that tries to modify individuals' preferred behavioral patterns. Thus, this thesis focuses on measurement and assessment practices at the middle school level and on determining the importance levels of different measurement results.

Background

Measurement and Assessment

Finkelstein (2003) defined measurement as a crucial enabler in contemporary ideas. It is the process of assigning a quantitative or qualitative value to a fact using an instrument. Measurement is described as assigning symbols to items and events' qualities. When a property of an item is measured, a symbol is linked to it, and other symbols are then allocated to other expressions of the feature using the same method. Then, in the symbolization system chosen, the qualitative connections between the expressions of the properties indicate and are suggested by the logical interactions between the symbols or measurements. Although it was first developed in the physical sciences and technology, its use has expanded to include other fields including the social, behavioral, and decision sciences. Ebel and Frisbie (1991) stated that while measurement involves the use of numbers, judgments of students' scores on the results are not needed. In education, the number of exam questions that each student correctly responds to is counted, and the achievement of each student in the class is given a number using the same procedure. Thus, measurement is helpful in determining the level to which students have abilities.

Measuring learning can be done directly and indirectly. Actual work or work outputs like papers, presentations, integrated tests, and pretests and posttests, serve as direct measurements of student learning to show understanding. On the other hand, indirect measurements are generally gathered via focus groups, interviews, or surveys and represent attitudes or opinions (Price & Randall, 2008). Luce and Kirnan (2016) found that the indirect technique is useful for examining students' perceived knowledge and skills on course subjects, whereas the direct methods, which involve students' test-taking or problem-solving activities, measure real knowledge. Both types of measurements showed improvements from the pretest to the posttest separately, but only the direct measure was linked with the final course grades. Additionally, their results indicated that students who achieved lower scores from direct measurements were too confident in their considered skills and knowledge. Based on their findings, while the indirect measurement was not a reliable indicator of student learning, it might be a useful teaching tool.

Assessment is a broad term and can be interpreted in a variety of ways, especially regarding educational settings. Assessment is the process of collecting data on student learning to guide instructional approaches. Or, as defined by Harlen et al. (1997), it is the process of acquiring, analyzing, recording, and applying data on students' responses to a task is defined as assessment in education.

All types of assessment ensure that the student's response is interpreted according to a certain standard of expectation. This consideration can include using different approaches for different types of assessments. The objective of the assessments must be taken into consideration while choosing between various measurement tools. This standard can also be applied based on the average performance of a specific population or age group of students (Harlen et al., 1992).

There are mainly two types of assessment: Summative and formative assessment. The goal of summative assessment is to provide a decision at the end of the learning process, whereas the goal of formative assessment is to offer feedback during the learning process (Burke, 2010). Classroom assessment methods that incorporate evaluation into the teaching process assist teachers in clarifying students' understanding and misconceptions (Veldhuis & Heuvel-Panhuizen, 2020). Besides, another assessment type definition which was made by Boud (1995), is about diagnostic assessment used to help students and teachers examine their present level of topic knowledge or abilities and ensures that any confusion is cleared up before instruction follows. The diagnostic assessment uses a collection of tasks that teachers and students do to gather data that can be utilized as a diagnostic tool to adjust instruction to better suit the requirements of individual students. All types of assessment tools include a pre-test, self-assessments, interviews, teacher observation, classroom discussion, exit cards, homework, or exams. Formative and summative assessments are the two main categories mostly used in education.

Although summative assessments are more frequently used in practice (Mohr, 1995), a comparison between them suggests that formative and summative approaches are not mutually exclusive and that, in theory, formative assessment should help students perform well on summative exams. Consequently, the most crucial aspects of formative assessment are (a) students' comprehension of their learning goals and targets, (b) the standards by which they will gauge their learning progress, and (c) what should be done next to further learning. Feedback can address the task, the student's processing of the task, recommendations for what to work on next, and scaffolds for the specific student. It is an active element of the process (Hattie & Timperley, 2007).

According to several researchers (Black & Wiliam 1998; Brookhart 1997; Cauley & McMillan, 2010), a positive relationship exists between performance in formative assessment in class and students' motivation and performance on small-group and large-scale assignments. The use of formative assessments, feedback, and instructional correction can be effective strategies to improve student motivation and success. Teachers can learn to adapt their instruction by including more formative assessment strategies in their daily lessons. Students' motivation is supported and their ability to continue their high engagement and success is made possible by the focus on formative assessment of instructional changes and student progress. Effective formative assessment use is essential for student motivation and success (Cauley & McMillan, 2010).

According to Sadler (1989), summative assessment differs from formative assessment in that it focuses on summing up or summarizing a student's academic progress and is designed to be reported at the conclusion of a course of study, particularly for certification reasons. Although it frequently impacts choices that might have significant educational and personal implications for the student, it is mainly passive and frequently has an immediate effect on learning. Instead of timing, the main difference between formative and summative evaluation applies to purpose and effect. Noonan and Duncan (2005) also compared the formative assessment types as the feedback teachers offer to students as they are learning to assess their learning results, in contrast to summative assessments used for promotions or extensive assessments conducted outside of the classroom. Formative assessments are often described as being directed by teachers and classroom based.

Teaching and learning are closely linked with the three concepts discussed above. It is crucial for the teacher since it may offer a variety of information that can be used for the planning of learning activities and students as well as the goals of classroom practice (Rea-Dickins & Germaine, 1992). Actions are taken to give teachers the instruction and resources as a national open bank (for example some early pilot SAT materials or national curriculum assessment data) to be stated in terms of overall goals they need to conduct school assessments with the rigor and dependability that are required for them to be successful in enhancing students' learning. By following these, teachers can become aware of and employ strategies for accessing students' current understandings and difficulties; advancing students' ideas and skills based on the knowledge of their current understandings; conversing with students about their development; and involving students in keeping records of their learning (Harlen et al., 1992).

After 1998, over the years, the use of and research area of formative assessment has grown and developed (Black & Wiliam, 1998; Brookhart, 2005; Burkhardt & Schoenfeld, 2019; Hattie & Timperley, 2007; Wylie, 2020). According to Radmehr and Vos (2020), assessment in math classes during the 20th century was mostly concerned with memorization of calculations and other facts with low cognitive demands. The assessment of mathematics should place more emphasis on higher order thinking in the 21st century.

Measurement and Assessment in Mathematics Education

According to Morgan (2000), an assessment of students' mathematics achievement in schools has a significant influence on both students and teachers as well as societal groupings. The benefits of a "good" assessment are claimed to involve a deeper understanding of students' mathematical thinking both personally and as a group, opportunities for better teaching through the design of lesson plans that consider what is known about students, and the potential to change instructional methods to better connect them with the assessment's goals. Morgan's research supports that assessment in mathematics education that adopts a societal perspective should focus more on understanding how assessment functions in mathematics classrooms and more generally in educational systems, as well as what the implications are for both individuals and social groups, instead of trying to find more effective ways of assessing.

In mathematics education, whether and how activities assess the complicated nature of learning mathematics and the curriculum or standards that are being assessed should be considered. The assessment design should place high importance on problem-solving, modeling, and reasoning in both large-scale and classroom examinations. The assessment should consider the many kinds of activities that are done during education (Suurtamm et al., 2016).

Jitendra and Dupuis (2015) studied the effective mathematical instructional programs provided in general learning environments by elementary and middle school and according to their research, it is significant for students whose ability, success, and motivation vary to develop the mathematical competencies necessary to meet various grade-level requirements. Singh et al. (2002) examined the impact of student and school variables on eighth graders' interest, motivation, and success in both science and mathematics. Attitudes and motivational factors have an impact on the variation in mathematical achievement. They discovered that students who were highly engaged in the class also performed well in mathematics and process following.

The studies that critiqued different types of assessment in mathematics education are highlighted. Cooper (1989) found in his research on the relationship between homework and achievement that there were differences according to age groups. While the effect of homework on the success of primary school students is negligible, Cooper stated that homework positively affects the success of students for high school students. Homework was beneficial in the success of middle school students, who are also in the age group of this study, but it is half as much of high school students' achievement effect. Cooper also examined the effect of the amount of homework by grade level. It was stated that the amount of homework has almost no effect on student success in primary school students, homework that lasts for one or two hours in middle school students increases student success, and more homework brings more success for high school students. De Jong et al. (2000) studied the effectiveness of mathematics homework based on homework frequency, homework problems, homework conditions, and homework time on students' achievement with 1394 students from 28 secondary schools in the Netherlands. They found a lack of policies intended to coordinate a rise in the amount of homework in Dutch schools. However, many math teachers assign homework after each class. The achievement was unrelated to how often and how much time was spent on assignments. The results of the research indicated that the only homework factor associated with achievement was the amount of homework. Students who are given less homework achieve less. This does not, however, suggest that teachers should assign extra homework. There are significant disparities amongst mathematics teachers regarding how or not the curriculum is covered in the homework. The efficacy of different amounts of homework is crucial, and this has a significant impact on achievement.

Swank (1999) investigated the quiz success of students who did and did not do their weekly math homework and found that there was no significant difference in quiz success. However, the limitation of this study is that the sample size is extremely small (22 students). At the same time, this study was conducted with primary school students. This also supports Cooper's study (1989) saying that the positive effect of homework is more in high school and middle school than in primary school.

According to Zhang and Henderson (2015), formative quizzes that are not intimidating can be applied to improve student learning and can help identify students who will not do well on following summative writing examinations. Rosario et al. (2015) examined the impact of three different forms of homework on students' academic achievement. These types are practicing homework to review the topic and to increase skills and timing, preparing homework to prepare students for the next lesson and to let them think about the next topic, and extending homework to improve students’ abstract thinking. This study was conducted with 638 sixth grade students from 27 mathematics teachers. Students' previous year's final math exam grades were used as pre-test data. For six weeks, homework was assigned to students once a week. As a result, while practice and preparation homework had no influence on students' math achievement, extended homework did. However, it was stated that attention should be paid to the frequency of these assignments, the readiness of the students, and the realistic deadline.

The positive and negative effects of homework, quizzes and exams on students' achievement is examined in many studies in the literature. The relationship between these measurement and assessment tools and students' success can be a very important data for teachers in the education process. If teachers know the relationship between these tools and student achievement, they can make more effective preparations such as planning, differentiation and preparing materials.

Problem

Despite teachers’ use of an endless variety of summative and formative assessments both in and out of class, there often exists a discrepancy between performance grades (exit cards, group works, quizzes, classroom participation, and attitude) and formal sit-in exam grades. In other words, formative and summative assessments may not sufficiently predict students’ achievement.

Furthermore, while the types of measurement and assessments are integral parts of educational evaluation, the overall purpose of all measurement and assessment practices is to obtain information about students' learning processes to make valid decisions about students. In parallel to this, the question of which tools or scores from tools are better predictors of student achievement deserves more attention.

Each individual needs different settings and sets up assessments to best showcase their abilities. For example, students who have anxiety problems about sitting in exams or students who prefer the sit in exams because they cannot perform in social settings (i.e., classroom) when they are assessed in a way that does not allow them to show their unique abilities. Students end up having negative connotations to certain subjects and their motivations for success in that subject are negatively affected. Alternatively, students may study harder for some exams, for example, at the beginning of the semester, or oral exams. In this study investigation of the students' grades of different assessment materials on student achievement with a quantitative approach could provide insight into this issue.

Purpose

The purpose of this study is to use decision tree analysis to examine relationships among assessment tools and students' success at the end of the semester used during two semesters for students in grades 5 through 8. Decision trees are a powerful statistical technique for relational analyses and can be used to predict anticipated outcomes. For this study, the formative assessment tools include measurement tools such as quizzes, homework, class participation and attitude, time management, and material and the formative assessment are measurement tools that are used to make end-of-semester evaluations.

Research Question

The main research question for this study is as follows: Among the formative and assessment measurement tools used in mathematics courses in grades 5 through 8, which are the most important predictors of the cumulative semester score?

In addition to this main question, this study also explores if these predictors are the same for each semester and for all grade levels.

Significance

Raising people who can view things from various perspectives, come up with innovative answers to the problems of everyday life, and think flexibly and analytically is now essential, both in Türkiye and throughout the world.

Students are given these skills in the educational process through mathematical thinking. Evaluation of student performance includes assessment and measurement procedures for math achievement. Even while teachers provide a final grade at the conclusion of academic semesters and years, being able to predict students' mathematical achievement earlier in the process might help teachers work more productively.

Students who have the potential to succeed in the semester may benefit from the identification of important predictors. Early detection of successful pupils enables the assignment of various interests to help students advance. Strong predictors that may be used to identify students who could ultimately fail are also crucial. Teachers can observe these students and provide them with the assistance they need to succeed.

CHAPTER 2: REVIEW OF RELATED LITERATURE

Introduction

An outline of the study was presented in Chapter 1. The study's overview, research questions, problem, purpose, and importance were all presented in Chapter 1. In addition, Chapter 1 contains a brief literature review on the measurement and assessment from its use around the world to its use in schools.

This chapter consists of three different subtitles. The purpose of this chapter is to provide the literature review regarding measurement and assessment concepts in the first section. The next second section is about the literature review of the measurement and assessment tools and their use in class. In the last section, the literature review of the prediction of students' achievement was based on several measurement and assessment tools such as national, international, high school or university entrance exams reviewed.

Measurement and Assessment

The terms measurement and assessment need to be clarified before talking about their use in educational settings.

Measurement and assessment are defined in the literature differently. For example, in her definition, Berry (2008) mentioned the quantitative description of specific properties of a group of individuals, things, systems, or events is called measurement. Measurement is any process that enables us to assign numbers to features of individuals, things, and other entities. in accordance with a specified rule. Ferris (2004) stated that measurement is an empirical procedure carried out with an instrument. Instrument in the context of disciplines that use physical measurement refers to hardware that interacts with the manifestation to generate the measure as its output. In other fields, such as economics or education, the term "instrument" refers to a methodical way to carry out measuring process observations. Ferris examined many definitions of measurement and developed a new one: "Measurement is an empirical process, using an instrument, effecting a rigorous and objective mapping of an observable into a category in a model of the observable that meaningfully distinguishes the manifestation from other possible and distinguishable manifestations." (p. 107)

Assessment is defined as a process of information gathered in relation to a known target or goal (Kizlik, 2012). According to Berry (2008), assessment must be viewed as a component of teaching and learning. The idea behind assessment is to describe student learning, determine each student's place in their individual learning curve, identify any learning challenges they may be experiencing, and give guidance to both the teacher and the student on the best ways to advance learning. Assessment for learning is the term used to describe this emphasis on using assessments to facilitate learning rather than archiving documents.

While most of the studies include the positive aspects of evaluation, Black and Wiliam (1998) discussed the negative aspects of assessment in their research. They stated that students are the primary consumers of assessment strategies used to improve learning. This statement has both positive and negative aspects. The negative aspect is that when the focus in the classroom is on points, grades, or class rankings, students are more concerned with getting the top grade than with increasing their learning. The result is that students only spend their time and effort searching for hints of correct responses and lose interest in learning with the aim of getting the right answer or getting good scores.

In terms of classroom assessment, Brookhart (1997) reported results about how teachers develop and apply classroom assessments as well as what they must learn in teacher preparation programs. First, the preparation programs encourage improving the standard of assessments used in classrooms since the effects of assessment methods may go beyond only obtaining accurate measurement data. Second, the performance assessment activities assist in communicating the assessment environment in the classroom to the students and, also, impact effort and success, especially through the objectives' perceived appeal, usefulness, and relevance. Emphasizing the value of including students directly in the assessment process is in its third point. Both the external and individual components of the assessment process should be improved by teachers and students taking equal responsibility. Sharing assessment responsibility with students shows them that teachers have trust in their comprehension and application of performance standards. The students who participate in the assessment procedure must feel more in charge of their education, which must improve their effort and performance.

In-Class Assessment Methods

This section provides information on several assessments. Then, a literature review was conducted on the use of such assessment tools (e.g., quizzes, homework, exit cards, exams) in the classroom. In addition, the types of questions used in the assessment tools based on formative, summative, and diagnostic assessments were examined.

Knowing the students' current skill levels helps in making better plans for instruction and learning procedures, especially in terms of what must be reviewed and covered. Diagnostic assessments usually take place at the beginning of a study unit to collect information about that. According to Chufama & Sithole (2021), diagnostic assessments are prior assessments made of the abilities of each student or each class, interests, experiences, levels of success, and challenges in any field of study that the students decide on following. However, Cornelius (2013) stated that assessments are mostly either summative or formative. Summative assessment is used after teaching with the goal of evaluating how well students understand or show an overview of knowledge, whereas formative assessment takes place during instruction to notify teachers of student understanding and to influence following instructional approach choices (Cornelius, 2013).

As Dixson and Worrell (2016) found, since many exams are developed for formative purposes and may also be utilized for summative purposes and vice versa, the difference between formative and summative assessment is mostly linked to how assessment findings are applied. While formative assessments, like quizzes or exit cards, can also be graded, the goal is to assess student knowledge and instructional efficacy during instruction, therefore evaluations of these assessments are sometimes not taken into consideration when determining final scores.

Several research has conducted on comparing formative and summative assessments. For example, Gezer et al. (2021) analyzed the differences between grade levels in terms of scores from summative assessments and found that students at lower grades perform better on formative assessments but lower on summative assessments. Martin et al. (2016) provided an explanation for this: as students go through the grade levels, they gain a greater subject understanding of mathematics. Another argument from Madison-Harris et al. (2012) suggested that for students who were low achievers and those who had learning difficulties, formative assessment was effective in improving learning outcomes.

Another positive aspect of formative assessments for students is that unfavorable outcomes (such as the students' focus on competition according to their perspective and the fear of failure of low achievers) are not always the result. The culture of formative assessment is effective because it is based on the idea that every student has the potential for achievement. Formative assessment may be a potent tool in this area provided it is communicated well. All students can benefit from a formative assessment, but low achievers benefit the most since it focuses on specific issues with their work and helps them understand what is wrong and how to fix it (Black & Wiliam, 2010).

According to Kincal and Ozan (2018), formative assessment strategies significantly result in improvement in several variables like students' academic performance, attitudes toward the class, and self-regulation abilities. The results of their study with fifth grade students showed that formative assessment affected students to increase their class participation and success according to their own developmental stages without comparing them to one another and focusing on learning rather than grades. They also showed that formative assessment improved students' attitudes toward classes and academic performance by carefully monitoring each student's success in satisfying learning objectives.

Against Bates’ (1979) statement that any summative or formative reward system negatively affects the interest and performance of students, Butler (1988) examined the effect of formative and summative assessments on students’ intrinsic motivation. She gave 12 classes of fifth and sixth grade students interesting tasks and divided them into three groups randomly: the first group received only formative assessment (comment group), the second group got summative assessment (grade group), and the third group took both formative and summative assessments (grade and comment group). She found that after the formative assessments, the interests, and performances of each group of students (low achievers, high achievers, all age groups, and regardless of whether they would receive feedback again in the future) increased. But although similar progress was seen after summative assessments, high achievers remained motivated only for the high grade when they knew they would be summative assessments on a regular basis. In other words, while formative assessment supports intrinsic motivation in every respect since summative assessments are extrinsically motivated for high achievers, students' having a very narrow field of interest reduces the quality, if not the quantity, of their performance. She found that the formative assessment lost its effectiveness in the groups in which the two assessments were used together, as she observed that the students focused on the high grades, not the comments. Biggs (1988), on the other hand, after examining the positive and negative findings of the formative and summative relationship from a large literature, suggested that the teaching process of formative assessment should be examined in a broader context with a multidimensional view.

Teachers, who are the main practitioners of assessments, have both theoretical and practical knowledge and competencies. In Berry's study (2004), conducted with more than 1000 teachers, teachers have theoretical knowledge of assessment according to the 4 categories of the Qualifications of Curriculum Authority (QCA) Assessment for Learning Projects (2000) (involving students in their own learning, modeling quality of learning purpose, giving feedback and self-assessment). However, there is a gap between what teachers know about formative assessment and how they apply their knowledge. Berry found that teachers want professional training, but it should not be forgotten that it is important to include practical examples rather than theoretical knowledge in such training. According to

Heritage (2007), instructors play a crucial role in formative assessment by providing students with the skills and strategies they needed to combine instruction and assessment into effective classroom practices. As opposed to being an essential component of teaching and learning, some teachers saw assessment as an activity that applies only to teaching. Strong planning should be made for in-service professional development, which requires the dedication of administrators at all levels of the educational system.

In the literature, there are suggestions for teachers as how to apply formative assessment in the classroom. For example, three templates were recommended by Cornelius (2013) for collecting formative assessment data in a simple, efficient, and safe way. The first template, *an anecdotal seating chart*, is an easy way to quickly record information on student development and difficulties. The second template, *a daily scorecard*, provides an overview of each student's development through curriculum assignments and in-class activities (participation, physical demeanor, exit ticket) on a single sheet. The third template, *an objectives grid*, keeps track of students' goals and provides chances to record situations.

Also, according to Conderman et al. (2020), various formative assessment examples were given to be applied at the middle school level. *Unison responses* are used to increase active participation in the questions posed to the class and to encourage all students to participate rather than always the same students. For this, the teacher first does a short review with the class and then says that “I will ask a question and give you some time to think about the answer to this question.”, then directs the question to the class and encourages the students without forcing. *Exit slips* (exit cards) are simple assessments that students complete at the end of the lesson on a small card or piece of paper. They are a quick and effective approach to gathering data in the classroom (Patka et al., 2016), and they also clarified effectively how students were understanding the topic and themselves as learners (Leigh, 2012). Some examples for exit cards could be: (1) sum up what you learned today; (2) write the key point and two specifics from today's lesson; (3) explain one new fact you discovered; (4) explain how you can use today's knowledge in your own life; or (5) assume you are the teacher and create a question for the test and the answer according to today's lesson. (Conderman et al., 2020).

Assignments are also an important part of formative assessment and students' achievement according to the literature. Overall, "homework" refers to schoolwork that is taken home (Corno, 1996). The positive effects of homework and time spent on homework on achievement with achievement gains are reported in many studies in the literature (Clinton & Hattie, 2001; Cooper et al., 2006). However, according to Trautweine (2007), the relationship between region, school type, and student level with homework was disregarded in these studies. Trautwein made three different studies using these variables and the findings of all three studies showed that homework has positive effects on achievement (both at class level and student level).

Nevertheless, it is important to know how and when to use the manifold of formative assessments. Instruction and formative assessment are interdependent. It's crucial to pick the right assignments for both classwork and homework. Tasks must be defined in terms of the learning objectives they support, and they only work effectively if preparation includes chances for students to share their developing understanding. All these methods (such as discussion, observation, and grading of written work) can be used to provide students those chances, but it's also important to pay close attention to how students talk, write, and act to see how well they are learning certain concepts (Black & Wiliam, 1998).

Students are motivated to do well on quizzes as a type of formative assessment. According to Zarei (2008) since the quizzes are a source of motivation, students' study more for them. Zarei (2008) also mentions that quizzes help students understand what concepts are crucial from their teacher's perspective and are more likely to be covered in the later assessment such as final exams. Gholami & Moghaddam (2013) examined the final performance scores of high school students with and without weekly quizzes (only those who took midterm exams), while Shirvani (2009) examined the achievement of students who had daily and weekly quizzes. The results of the research found that students with weekly quizzes were more successful than students without quizzes, and the daily quiz effectively improved students' mathematic scores and homework grades. As Dengri et al. (2021) quiz-based learning is clearly more interesting and dynamic for students, and it enhances student curiosity and enthusiasm to learn. The quiz format may be modified to reach certain aims based on the target learning. Quiz-based learning may be made more effective by keeping it dynamic and adaptable, which implies that the structure, material, and challenge level should be customized to the students' requirements, abilities, and limitations. Multiple-choice, true/false, short answers, multiple responses, fill-in-the-blanks, matching, ordering, and other entities. are just a few of the question categories that can be used in quizzes. The use of interactive forms gives students rapid feedback and aids in guiding student discourse. Through active student participation, it helped evaluate the knowledge foundation that already exists and then contributes to its expansion.

Pajares and Miller (1997) conducted a study to determine whether students' assessments of their mathematical self-efficacy were affected by the traditional, multiple-choice approach of mathematical inquiries compared to an open-ended approach. Self-efficacy results did not differ according to the types of assessment. However, students who took the multiple-choice test scored higher than those who took the open-ended test. Students who took the multiple-choice exam could also have scored better because of making some predicted guesses that turned out to be correct. Similarly, according to Funk and Dickson (2011) several test-taking techniques, such as finding correct answers by eliminating incorrect responses or by guessing, can be used to determine the right answers to multiple-choice questions, including recognition, recall, and analysis.

Couch et al. (2018) conducted a study to identify students' incomplete understanding of the multiple-choice question format by using the same questions in a true-false format. The multiple-choice structure, choosing the right answer is challenging for a student whose knowledge of the subject is not clear and who sees the distracting options of the multiple-choice question as well. The multiple-choice questions show a student's chosen response, but they do not directly reveal their determination to choose one of the possibilities. As a result, students might choose the right answer while also thinking that some of the other options are also accurate (but unable to express this). On the other hand, a student could pick the wrong answer yet still believe it to some extent to be true. Because it might make it difficult for teachers to assess how students are thinking about the available response alternatives, the artificial limitations to only choosing one answer in the multiple-choice format present a challenge to teachers. Additionally, questions for assessment are used to promote collaboration among peers. For example, true-false questions encourage students to debate all their alternatives rather than focusing only on the best one. Multiple-true-false questions are an alternative that preserves several advantages and grading advantages of the multiple-choice format. Multiple-true-false questions start with a question stem and then have many prompts or assertions that students must decide whether to accept as true or false. Multiple-true-false questions work similarly to multiple-choice questions, but because multiple-true-false questions require marking both true and false statements, it's a system where students "choose all that apply". Because student replies to a single question stem might include true and false responses, multiple-true-false questions have the power to identify individuals with mixed ideas by asking students to assess every claim (Parker et al., 2012). While the multiple-choice format suggested that students can succeed by selecting the right response from a list of options, the multiple-true-false format showed that learning includes assessing both correct and incorrect thoughts (Richardson 1992). But on the other hand, according to Toppino and Ann Brochin (1989), the findings showed that exposing students to a statement on a true-false question increased the probability of believing the statement was true, regardless of whether the statement was true or not. This occurred for false statements, resulting in a negative suggestion impact.

Prediction of Student Achievement

The literature has a variety of prediction studies of student achievement. For example, students' high school cumulative grade point average is a predictor of future academic success. According to studies (Geiser & Santelices, 2007; Olani, 2009), high school overall grade point average may be used as the strongest predictor of students' academic success in subsequent educational levels.

Using prospective research of 4660 middle-school students from 24 schools, Casillas et al. (2012) investigated the effects of prior academic success, psychological, behavioral, demographic, and school-related factors on the high school grade point average. The results showed that academic achievement indicators which were calculated as the mean of students’ English, mathematics, science, and reading subject grades are the best predictors of future academic success. Bowers (2010) and Mac Iver (2010) also found parallel results with other studies, reporting students' previous academic achievement was a crucial predictor of their high school and graduation achievement.

Dinçer (2020) examined the relationships between high school entry and exit performance of laboratory high school students in Turkey. In the study, student performance on an admission exam (Turkish language, mathematics, science, English, non-verbal reasoning, Turkish essay writing) was regressed on both national (university entrance exam results, and high school cumulative grade point) and international (International General Certificate of Secondary Education [IGCSE], International Baccalaureate Diploma Programme [IBDP] scores) performance indicators average scores as exit performance was analyzed. He found that the admission exam's Turkish, science, and mathematics components are the best strongest predictors of students' cumulative grade point averages. The Turkish component of the admissions test is the most accurate predictor, math and science also predict well to IGCSE results. Other factors (writing a Turkish essay and the interview) did help predict students' academic performance in the IGCSE exams, although their effects were extremely small.

Alghamdi and Al-Hattami (2016) tested students' university achievements based on the first three years (students at this level completed more than half of their educational program) university grade point averages were predicted by using the high school grade point averages, General Aptitude Test (150 multiple choice general skill questions, not subject-oriented), and Achievement Test (Science graduates are requested to enter the university). The findings showed that university achievement is significantly predicted by high school grade point average and achievement test scores.

Two different and important admission exams in Türkiye are the high school entrance exam (LGS which was named TEOG before 2014) and the university entrance exam (YKS). LGS is taken at the end of eighth grade and YKS is taken at the end of 12th grade. For three distinct school types in Turkey, Science High Schools, Anatolian Teacher Training High Schools, and Anatolian High Schools, Bahar (2013) investigated the correlation between TEOG and YKS results scores. Analysis was done on the test results of 34.479 students from 570 schools. She found Turkish, mathematics, social sciences, and natural sciences scores of TEOG and YKS had different mean scores in each school type. According to YKS and TEOG results, the highest average difference increasingly was observed in mathematics in all three school types (Anatolia High School, Anatolian teacher training high school, and Science High School). Although Bahar (2013) did not examine the degree to which YKS scores were predicted by TEOG scores, the differences determine that the LGS results are probably not strong predictors of the results of the YKS.

Kontaş and Özpolat (2017) studied how well Turkish, Mathematics, Science, Religion, Foreign language, and History written exam results predicted TEOG results. The records of 1035 students in eighth grade who were enrolled in the first term were used to gather the research data. The relational approach was used, and they found a strong correlation between exam scores and TEOG exam scores.

Çelikel and Karakuş (2017) investigated the relationships between the TEOG examination and eighth grade students' academic success in mathematics as well as the impact of the mathematics lesson on the teaching methodology. In the study, it was discovered that there was a correlation between students' performance on the TEOG examinations and their school's written exams. In 3 private and 5 public schools, the study involved 17 mathematics teachers. According to the research, the teachers found that the success of the students in the TEOG exam was generally compatible with their success in the mathematics exams at school. At the same time, one of the aims of education given in schools is to ensure that students succeed in central examinations. For this reason, exams are held in parallel with the central exams in schools, and students are evaluated with grades according to the results of these exams. Although teachers determine the difficulty level of written exams in parallel with the student level in some schools, the central exams are prepared for the level of all students. While students have difficulties in the exams made according to student success in successful schools, they can solve the central exam questions more easily. Similarly, in schools with low achievement levels, central exams are difficult for students who are used to easy questions. Both cases can also cause students' achievements in school and in central examinations not to be perfectly matched.

A similar study was done by Kingston and Anderson (2013) in the USA using a sample of more than 2000 students from 20 various high schools. The researchers wanted to compare the math and reading outcomes that students had achieved in a community college with the state assessment math and reading results that they had obtained from the Center for Educational Testing and Evaluation (KSA) at Kansas University. The findings reveal a significant relationship between state assessment mathematics scores and a college algebra course and a slight correlation between KSA Reading scores and English course grades.

Higher education institutions primarily use internationally valid exams as one of the factors in their student choice procedures, such as the Scholastic Assessment

Test (SAT), Advanced Placement (AP), and American College Test (ACT). According to Pitman (2016), there is a significant correlation between students' achievement on these tests and their performance in higher education. Also, for satisfactory predictions for long-term measures such as graduation and GPA, the combination of SAT scores and high school GPA is stated by Burton & Ramis (2001) to be a stronger predictor than only SAT scores or high school GPA.

The International General Certificate of Secondary Education (IGCSE) and the International Baccalaureate Diploma Programme (IBDP) are also other international examinations. The correlation between a student's IGCSE and IBDP results was also studied by Corlu (2014). He found a significant correlation between students' IGCSE and IBDP results.

Summary

- Formative assessment is more effective for low-achieving students.
- While measurement is a concept that is generally done with a tool and expressed with numbers or symbols, this concept can also be done with an "observation" as a tool in more abstract situations such as education.
- Assessment should be seen as a component of teaching and learning. Teachers should also be well-equipped in this subject and should examine the evaluation documents not as grading and archiving, but as a communication material and developing the learning process.
- Homework has a positive effect on students’ academic achievement.
- True-false questions offer many of the advantages of the multiple-choice format, as well as create a productive discussion environment among students and provide an option that includes the advantages of grading. But, as a disadvantage,

students might tend to be true regardless of whether the answer is true or false if they are exposed to false statements.

- Quizzes are a sort of formative assessment, and thus students are encouraged to do well on them. Math test results and homework grades are successfully raised by quizzes for students.

CHAPTER 3: METHOD

Introduction

This chapter presents the methodology of the study. First, the research design and context were explained; then participants and instrumentation were described. Finally, details were given about how data was collected and how it was analyzed.

Research Design

This current study is mainly relational. A relational design emphasizes relationships between things and works under the presumption that everything is interconnected (Reynolds, 2022). This study used dependent and independent values to explore the relationship among students' math achievement indicators collected over an academic year and their end-of-semester math performance grades.

In this study, relational analyses were based on decision trees, a statistical approach with strong visual components. The decision trees are frequently used in educational research because they determine a variety of relationships between these predictors for various sets of predictors and different subgroups (De Ville & Neville, 2013). Given that indicators of students' mathematics performance can be highly skewed, decision tree analysis outperforms conventional least squares regression when analyzing performance grades and end-of-year grades. Finally, researchers can communicate their findings in a non-statistically intensive context with clarity of understanding and simplicity of visual presentation (Olson et al., 2012; Wu, 2009)

Context

The school where the data of this study came from is a private school located in Istanbul, Turkey. According to the information provided by the school's admissions office, students in this school could be characterized as children from families with high socioeconomic status with international experience, providing their children academically.

There are five classes of each grade level. For example, in grade 5, the sections are 5A, 5B, 5C, 5D, and 5E. There are nine mathematics teachers in the mathematics department and three or four of them teach in total for each grade. Each teacher has a maximum of two different grade levels. At the beginning of the year, a responsible teacher for each grade level is assigned. A two-hour grade level meeting is held every week. Lesson plans and materials are decided at these meetings by all teachers entering that grade. The materials to be prepared are shared among the teachers and the due date is determined. Care is taken to make a difference among the teachers who prepare the assessment and evaluation materials throughout the year. As the content of the material is decided at the meeting, all level teachers check after the material is prepared. Then the answer key and rubric are prepared. The teacher in charge of the level moderates the meeting, writes the decided plan, and follows the progress of the level.

In the school, each class has a "support" system for one math lesson per week. A support mathematics teacher who is one of the other mathematics teachers in the school is determined for each class in the weekly lesson schedule. The support teacher teaches the students' deficiencies with the determined student or small group of students by the class mathematics teacher in a different empty classroom. These support hours can be used at the initiative of the teacher, depending on the needs of the class or students in the class.

Measurement and assessment tools used in mathematics lessons throughout the year for each grade of middle school are readiness, exit cards, quizzes, performance tasks, homework, presentations, and projects. An academic year consists of two semesters and each semester three performance grades and two exam grades are given from the mathematics lesson (based on the weight of the weekly lesson hours). Each of the three performance grades consists of homework given and quizzes applied until that date, and teacher observation.

The school creates the performance assessment according to the criteria given in Table 1. Homework, quizzes, and teacher observation constitute this assessment component. Performance grades are calculated out of 100 using the weighted percentages of the criteria given in Table 1. Information about the assessments that were used in this study as sources of data is provided in the next section. Other forms of assessment (teacher observation and projects) are described here but were not part of the d

Table 1

Performance Assessment Criteria

Criteria	Weight
Homework	30%
Quizzes	30%
Teacher Observation	
Active participation	20%
Time management	5%
Material	5%
In-class attitude	10%
Total	100%

There are four different criteria under the title of teacher observation out of 40 points. This means that 40% of the students' performance grades, excluding homework and quizzes, are composed of teacher observation. These criteria are active participation out of 20 points, time management out of 5 points, materials out of 5 points, and in-class attitude out of 10 points. Active participation is assessed by teacher observation through the student's regular participation in the lesson, making meaningful contributions to the lesson, asking questions about the topic does not understand, and listening to and questioning other students' questions and answers. The time management criterion is assessed by the student's timely arrival to the lessons, starting the in-class studies in the given time, or completing them within the given time. The material criterion is given according to the student's bringing the necessary materials to the lesson, delivering the products, and using the course materials in accordance with the instruction. The in-class attitude criterion is assessed according to the student's regular follow-up of the instructions, active listening to the teacher/friends, respecting the ideas of others, and speaking by taking the floor. In this thesis, these four criteria were taken together as a single variable.

According to Ministry of National regulations, the project is the work that students do individually or in groups under the guidance of the course teacher to examine, research, and comment on a subject determined in line with their wishes; students reach new information, produce original thoughts, and produce a product because of inferences. Furthermore, according to the Ministry of National Education regulations, each student must take at least one project from the desired course in the first or second semester of the academic year.

Table 2

The Number of Students Took a Mathematics Project by Semesters

Grade	First Semester	Second Semester
Grade 5	10	6
Grade 6	19	6
Grade 7	16	4
Grade 8	15	6

The reason for the decrease in the number of students taking projects in the second semester may be that the students interested in mathematics have already chosen a project in the first semester, the in crease in the students' schedule such as the intensity of external exams, competitions, trips, or the decrease in the school motivation in the second semester. The project grades are measured out of 100, but the distribution between the project grades of 82 students is a very narrow range. For example, in the first semester, 10 of the fifth grade students got a math project and only two of them got 83 and 87 while the others got 100 points. The sixth grade students, 19 received a math project, and only two of them received 70 and 75, while the others received 100 in the first semester. Also, in the second semester, sixth grade students received a math project, and only one of them got 95 points, while the other five students got 100. Since there is a similar situation at all levels in every period, project grades were not used in this study.

Sources of data

The data for this study came from the formative and summative assessments of students from fifth to eighth grades of the school in the 2021-2022 academic year.

Information on the number of students whose grades are used in this study is given in Table 3. Mathematics grades of all students in the school were included in this study.

Table 3

Information on the Number of Students Whose Grades Are Used

Grade	Boys	Girls	Number of Students
Grade 5	54	55	109
Grade 6	52	57	109
Grade 7	61	49	110
Grade 8	66	48	114
Total	233	209	442

The data used in this study consists of the numerical scores from different assessment tools for each student. Quizzes and application frequency are applied according to the decision of the mathematics department. Quizzes are announced one week before the application. In this study, for each grade level, 55-65 homework grades, 12 quiz grades and four exam grades were used as averaged according to different calculations or percentages.

Homework

Mathematics homework is planned in such a way as to focus on the topic taught in that week and the previous topic. Homework is prepared by the grade-level teachers according to the weekly division of labor. Homework questions are prepared in such a way that they are ordered from simple to complex and include challenging questions. Weekday homework is shorter and practice intensive. Weekend homework is more intense. Mathematics homework is given once or two every weekday. The due date of the homework during the week is one, two, or three days depending on the content of the homework or the intensity of the student's school schedules for that week. Another homework is given every weekend; the deadline is set at 10.00 pm every Sunday for each grade. Homework is given digitally as an eco-friendly school and students are expected to upload them to the digital platform within the given time. No extra math homework is given to students during the holidays. Intensity is given to reading activities, and mathematics homework continues as a weekend routine. Homework is graded by each mathematics teacher with the rubric up to four points each week.

Homework is given to reinforce the topics covered in the lesson. The time given for homework is one to three days, depending on the intensity of the homework or the schedule intensity of the students in the respective week. Homework grades are mainly awarded based on correctness and responsibility, careful, timely submission, as checked by teachers. Homework is measured with the rubric given below.

4-points: On-time and completed submission

3-points: Completed content but late submission

2-points: Incomplete submission

1-point: Submission without calculations (Just typing results or ticking the box)

0-points: No submission

The homework criterion is calculated as the average grades of all homework up to each performance grade date (three performance grades in each semester) converting the grade out of 100. Generally, students have two homework each week, one of them is assigned for weekdays and the other for the weekend. This routine may change during exam weeks and school holidays. A total of 55-65 homework is given at each level throughout the academic year.

Quizzes

Quiz topics, question types, and the number of questions in the quiz are determined in line with the decisions taken at the level meetings according to the mathematics lesson and mathematics homework performances of the students. The date and topics of the quizzes are announced to the students in the weekly post (which includes all academic content and activities lesson by lesson) one week before the quiz. It is the student's responsibility to follow the weekly posts. The duration of the quiz is usually 20-25 minutes, depending on the content. Each mathematics teacher grades their students' quizzes using the prepared answer key and rubric up to any total points but then each quiz is converted to 30 points.

Quizzes are applied at the end of one or two subjects to measure how well subjects are learned. Quizzes include open-ended, true-false, checkbox, fill-in-the-blank, or multiple-choice question types. Quizzes can be prepared on different total points. Each of them is converted out of 100 while included in the calculation of the performance grade. Two quiz grades are considered for each grading date as their average, so a minimum of 12 quizzes are applied throughout an academic year. If it is more than 12, the averaging and performance scoring system is applied in the same way. Homework and quizzes are always included in the average calculation process after converting them into a scale out of 100.

Written Exams

According to the Ministry of National Education, students in fifth to eighth grades are required to take at least two written, standardized exams from each subject area regardless of the number of weekly class hours. The number and dates of exams are determined at the beginning of each semester and announced after the school principal's approval. Consequently, in the school where this study was conducted, two mathematics exams are administered each semester.

Exams are prepared jointly by all teachers who enter the relevant grade. At least two weeks before the exam, the exam objectives list is created, and all teachers write sample questions about the selected objectives. One week before the exam objective lists are shared with students. When this exam question pool is completed, the teachers come together and select the questions according to their low, medium, and high difficulty levels, distinguishing or critical thinking criteria, and the ratio of topic distribution. The responsible teacher, who is determined at the beginning of the year for each exam, writes the exam with the selected questions and solves it by keeping time, then makes the corrections found, and delivers it to the other teacher. This process ends with the grade teachers first, then all the teachers in the mathematics department, and finally the approval of the head of the department. The two exams consist mainly of open-ended questions and include multiple choice, multiple true-false, and fill-in-the-blanks questions, graded out of 100.

The exams are graded jointly by the teachers who teach that grade level. The questions to be graded are shared equally according to the difficulty of grading, and each teacher grades the selected question of all students at that level. (For example, in a 10-question fifth grade exam, teacher X grades questions 1-2-5, teacher Y grades questions 3-7-8, and teacher Z grades questions 4-6-9 of all grades 5 students' exam.)

End-of-Semester, Cumulative Grade

The assessment and evaluation system of the mathematics department, from which data of this study was obtained, is designed to include four standardized exams and six performance grades in an academic year. For calculating the end-of-semester grades, the arithmetic average of the three performance grades out of 100 points for each is calculated first. Then, the end-of-semester grade is generated for each student by calculating an average of performance grades and two exam grades.

End of term grade = (performance grades average + exam 1 + exam 2) / 3

In this study, students with an end-of-semester grade which means a cumulative score. According to the Ministry of National Education, students need to be successful in at least 70 percent of their grades to receive a certificate of achievement. Therefore, for this school students who earn 70 points and above are identified as successful and students with a score below 70 are unsuccessful.

Method of Data Collection

All mathematics teachers who record the mathematics grades in this study save the data in an Excel file. As the main system, each teacher opens titles or sheets in their files in the form of homework, quizzes, exams, teacher observation, performance, and to save data and calculate and see the end-of-semester cumulative final grades. This is how the teachers save all data. The teachers convert all measurement and evaluation tools based on different scores into percentages related to the formulas created in this Excel file. The data of this study were collected and used by combining the data in the files taken from each mathematics teacher in a way that all grade levels came together.

Methods of Data Analysis

In this study, considering the skewed distribution of the scores, non-parametric approaches were conducted. Almost 75% of the predictors had skewness values outside the range of -2 and +2.

Decision tree analysis, a multiple variable analysis, systematically divides data into segments depending on the identified inputs until the segments include homogeneous facts. It does this by using algorithms to discover an important set of inputs in predicting the target variable (Maimon & Rokach 2014). It generates a reversed tree form with a root node at the top, followed by branches and child nodes, which is why it is termed a decision tree. The segments that were separated from the parent node based on a splitting rule are represented by child nodes, whereas branches indicate splitting rules. The tree then grows because each child node becomes a parent node. When every segment has homogenous facts or when it exceeds ending criteria (such as the maximum number of branches or the minimum number of facts at a node) determined by researchers, the tree ends up growing. The leaf nodes, or last generation nodes, are the segments of homogeneous facts after the tree ends growing. At the same time, decision trees produce visual results and are therefore very useful for interpretation. The use of Chi-squared Automatic Interaction Detection (CHAID) Analysis or decision trees can also be seen in student, teacher, material, or exam evaluations in many studies. Park and Dooris (2020) chose decision tree analysis to provide a more comprehensive and rational method for examining and understanding student evaluations of instruction and course grades.

A decision tree is generally created from a training data set whose classes are learned through the inductive approach and has the shape of a tree (Dondurmacı, 2011). These trees are formed by dividing large data sets into small groups using decision rules. Using decision criteria in line with the goal variable, the decision tree structure separates extreme heterogeneous data sets into homogeneous groupings

(Berry & Linoff, 2004). As this study has a very limited sample size, creating training data was not possible.

CHAID organizes data by identifying independent or predictor variables that differentiate the target variable from the dependent or target variable on which classification is based. While CHAID and regression both discover factors associated with a target variable, CHAID has the distinct benefit of uncovering important factors that maximize differences across subgroups (Kalender, 2014). The first level of a classification tree is defined by the predictor variable that explains the greatest proportion of the target variable, and the first-level clusters are formed up of different values of that variable. By breaking up the clusters created in the first classification into related clusters, a different predictor variable that makes the second-largest contribution to explaining the variations in the target variable is chosen for the second-level classification. Decision trees generate visual outcomes, making them ideal for interpretation. The depth of the tree can be determined before the CHAID analysis which is one of several settings that can be established. The minimal number of node cases can also be altered to produce various tree layouts. According to Kalender (2014), particularly when the sample size is small for clusters to have an acceptable number of instances, the values should be set appropriately. In this study to make the leaves large enough to be realistically useful, the minimum leaf size (i.e., the minimum number of students that a leaf may have) was set to 10.

Table 4

Variables Used in CHAID Analyses

Predictors	Meaning
quiz_1_30	Mean of the first and second quizzes in the first semester
quiz_2_30	Mean of the third and fourth quizzes in the first semester
quiz_3_30	Mean of the fifth and sixth quizzes in the first semester
quiz_4_30	Mean of the seventh and eighth quizzes in the first semester
quiz_5_30	Mean of the ninth and tenth quizzes in the second semester
quiz_6_30	Mean of the eleventh and twelfth quizzes in the second semester
hw_1_30	Average of homework given up to the first performance grade in the first semester
hw_2_30	Average of homework given up to the second performance grade in the first semester
hw_3_30	Average of homework given up to the third performance grade in the first semester
hw_4_30	Average of homework given up to the fourth performance grade in the second semester
hw_5_30	Average of homework given up to the fifth performance grade in the second semester
hw_6_30	Average of homework given up to the sixth performance grade in the second semester
summative_1	First written exam grade in the first semester
summative_2	Second written exam grade in the first semester
summative_3	First written exam grade in the second semester
summative_4	Second written exam grade in the second semester
cumulative_1_binary	End-of-semester achievement in the first semester
cumulative_2_binary	End-of-semester achievement in the second semester

In this study, 16 predictors (quizzes, homework, and exams) and two cumulative grades from two semesters as dependent variables are used in CHAID analysis (see Table 4). Since the data of active participation (20%), in-class attitude

(10%), material (5%), and time management (5%) were quite high and the range in the groups was very narrow, they were not used in this study. Only few scores existed in these predictors. Since the quizzes are prepared over any point and the assignments are graded over 4 points, all the assessments are first converted to 100 points and then the end of the semester grades are calculated by taking 30% of them. And then end-of-semester grades are labeled as successful or unsuccessful according to the criteria of 70 out of 100 points in the end-of-semester grades. For example, homework_1_30 means the grade out of 30 of the average of all homework grades given up to the first performance grade, quiz_1_30 means the grade out of 30 of the average of all two quiz grades applied to the first performance grade because each of the homework and quiz grades makes up 30% of the performance grade, and summative_1 means first written exam grades up to 100. Two different models were tested with and without the written exams.

The dependent variable of all analyses was the end-of-semester score, analyzed separately with two different independent variable groups. One of the groups consists of eight variables and these are three homework grades, three quiz grades, and two exam grades. Since students study more for written exams, the other group consists of six variables by excluding two exam grades and these are three homework grades and three quiz grades. Eight (six after removing summative) predictors were used in CHAID made for any period of any class level: three quizzes, three homework, and two summative. Classification efficiency is also examined in this study.

The analyses were conducted for both semesters using data from students in grades 5 through 8. Furthermore, CHAID analysis was conducted separately for each grade level and for each semester. Due to the low sample size, parent and child nodes were set by 10 to have a minimum of 10 students in each node. That is if a child node has 9 or fewer students, no new nodes were sought to be created.

CHAPTER 4: RESULTS

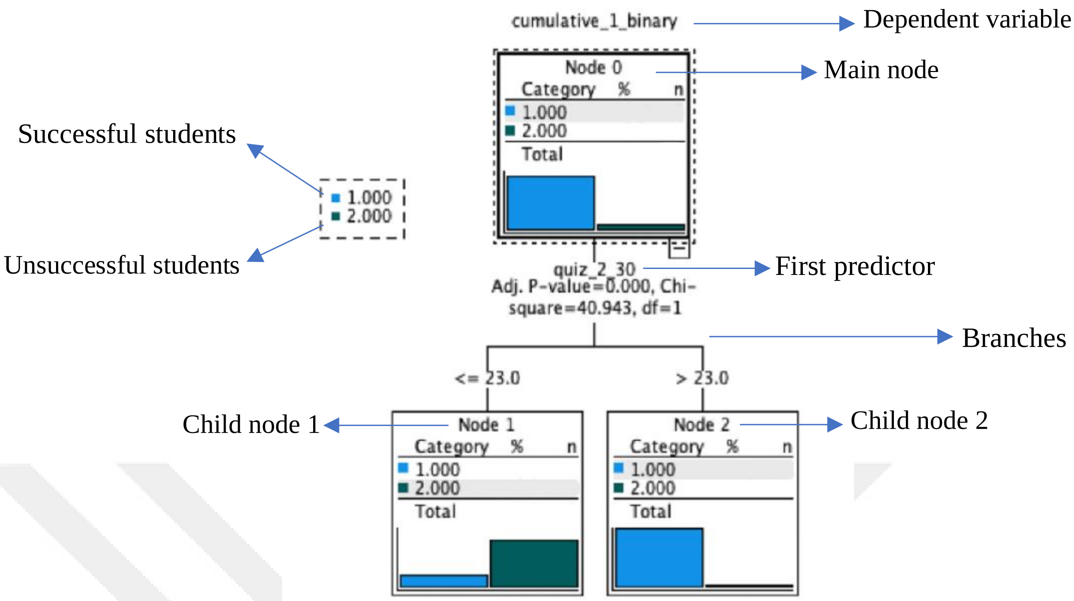
Introduction

The main purpose of this research is to predict the achievement of private secondary school students by using scores they collected throughout the year.
In this chapter, the results of the CHAID analyses are presented in decision trees. CHAID is used to identify the determinants of subgroups or segments as an experimental tool. CHAID clusters data by identifying independent or predictor factors that distinguish the target variable from the dependent or target variable on which classification is based (Kalender, 2014).

The decision tree diagram's nodes show the percentage and count frequencies of the target variable in relation to the selected predictors, together with the overall number of cases examined at each node. Also, the correct classification rates were examined as successful and unsuccessful concerning end-of-semester mathematics achievement. The performance of the decision trees is determined by their overall classification rate (Pal & Mather, 2003). According to Li et al. (2013), the percentage of data sets that the model successfully classifies is the classification accuracy, which is used as the benchmark for comparing results. The algorithm's overall performance was measured by determining the average classification accuracy. An example decision tree diagram and related terms are given below (see Figure 1).

Figure 1

An Example Decision Tree Diagram and Related Terms



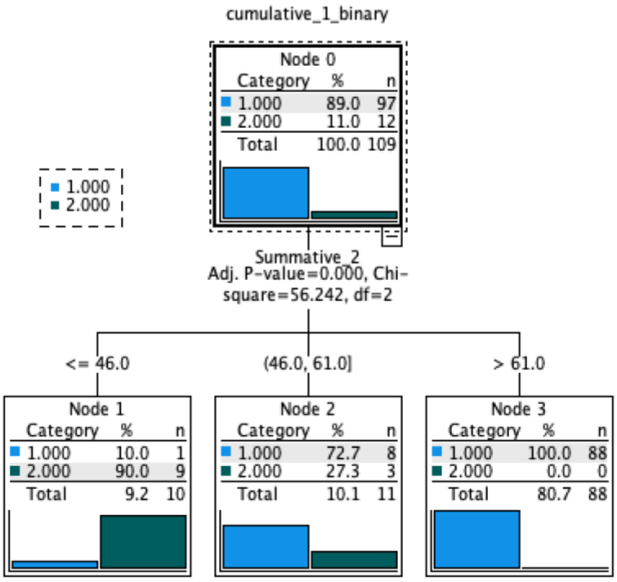
First Semester Decision Trees for Grades 5, 6, 7, and 8

Decision Trees for Grade 5

Figure 2 presents the decision tree analysis of the data (formative and summative assessments) from 109 grade fifth grade students. The main node of these students has 97 (89%) successful and 12 (11%) unsuccessful students. According to the CHAID analysis, which includes eight independent values, the strongest predictor is the second written exam (summative_2) with a mean of 79.83 for 109 grade 5 students. This predictor creates three nodes with no child nodes. Node 1 includes 10 students and only one (10%) is successful.

Figure 2

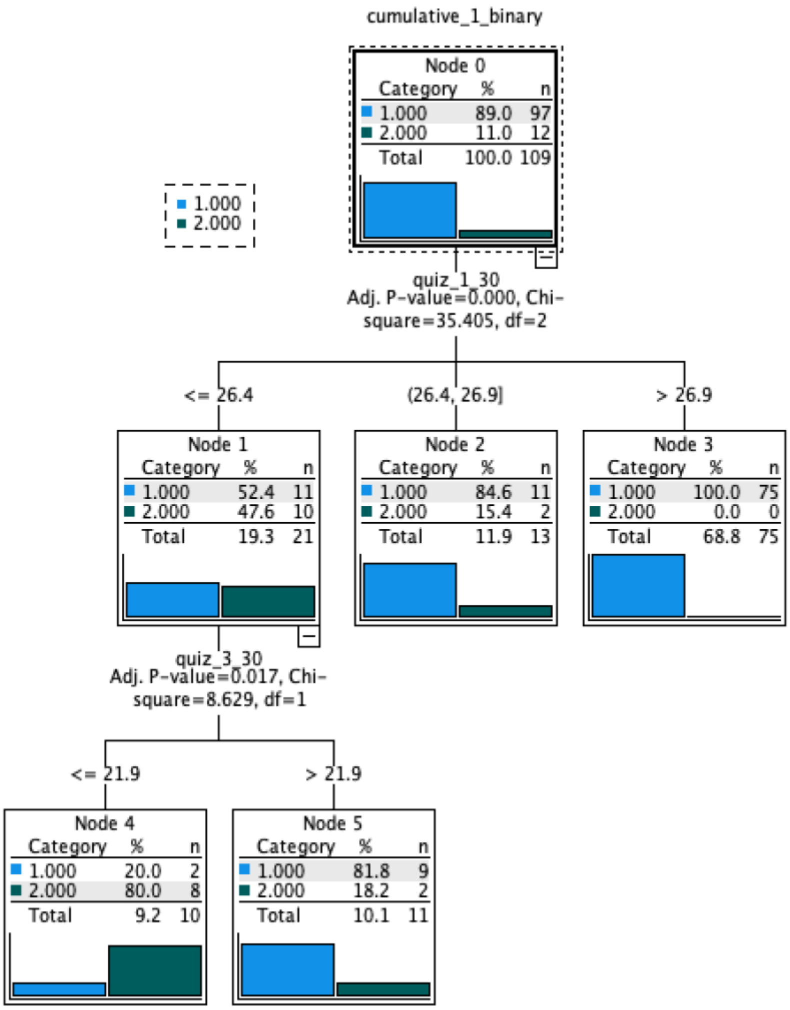
Decision Tree for the Prediction of Grade 5 Students' First Semester Grades with Summative



The second exam grades (summative_2) of students in this node are equal to or less than 46. Node 2 includes 11 students, 8 of whom are successful (72.7%) and the rest of them (27.3%) are unsuccessful. In this node, students' results from the second written exam are between 47 and 61, inclusive. Node 3 includes the remaining 88 students and all of them (100%) are successful and get over 61 on the second written exam. There is no student who is unsuccessful in this node (see Figure 2). The overall classification rate for this grade level is 96.3%. The correct classification percentage is 99% for successful students and 75% for unsuccessful students. Thus, the decision tree including a single variable seems to predict future cases which is the second exam (summative_2).

Figure 3

Decision Tree for the Prediction of Grade 5 Students' First Semester Grades without Summative



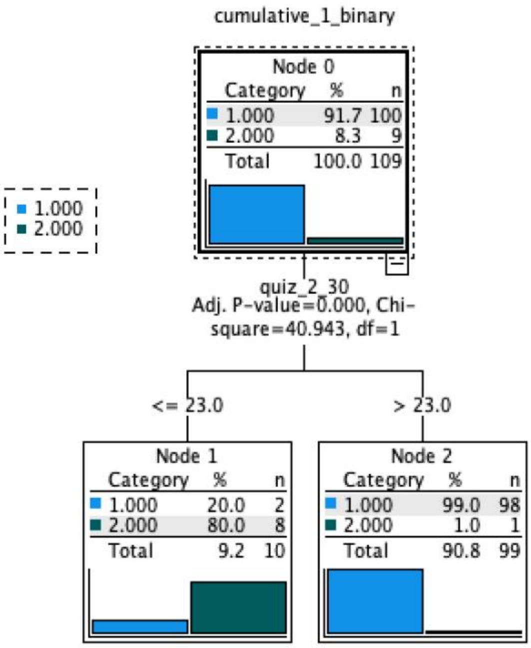
The same analysis is performed without summative with six independent values as shown in Figure 3. The main node has 97 (89%) successful and 12 (11%) unsuccessful students, which is same as the analysis with eight independent values including two written exams. But, in this case, the average of the first and second quiz grades (quiz_1_30) for the first performance grade were found to be the strongest predictor of the student's achievement at the end of the first semester. This analysis also created three nodes. Node 1 includes 21 students, 11 of them (52.4%) are successful and 10 of them are unsuccessful (47.6%). The average of the first and second quiz grades (quiz_1_30) of students in this node was 26 or lower out of 30. The mean of the quiz_1_30 variable is 27.96. The child nodes of node 1 (node 4 and node 5) consisted of the average of the fifth and sixth quiz grades (quiz_3_30). In node 4 there are 10 students, two of them (20%) are successful and eight of them (80%) are unsuccessful. The average of fifth and sixth quiz grades (quiz_3_30) of students in this node is 21 or lower. Node 5 includes 11 students, 9 of them (81.8%) are successful and two of them (18.2%) are unsuccessful and get over 21 from the quiz_3_30. Node 2 for quiz_1_30 includes 13 students, 11 of them (84.6%) are successful and two of them (15.4%) are unsuccessful. Node 3 includes 75 students and all of them (100%) are successful and get over 26 from the quiz_1_30. There is no student who is unsuccessful in this node. For this grade level, the overall classification rate is 94.5%. The correct classification percentage is 97.9% for successful students and 66.7% for unsuccessful students. This tree cannot classify unsuccessful students into the correct group as much as the previous tree, especially unsuccessful students.

Decision Trees for Grade 6

The first CHAID analysis with eight independent variables for 109 sixth grade students is shown in Figure 4. The main node of sixth grade students has 100 (91.7%) successful and 9 (8.3%) unsuccessful students. The most dominant predictor is the average of third and fourth quiz grades (quiz_2_30) with a mean of 26.92. This predictor creates only two nodes with no child nodes. Node 1 includes 10 students and two of them are successful in quiz_2_30. The average of third and fourth quiz grades (quiz_2_30) in this node are 23 or lower. Node 2 includes 99 students, 98 of whom are successful (99%) and only one of them (1%) are unsuccessful. In this node, students get over 23 from quiz_2_30 (see Figure 4). The overall classification rate is 97.2%. The correct classification percentage is 98% for successful students and 88.9% for unsuccessful students.

Figure 4

Decision Tree for the Prediction of Grade 6 Students' First Semester Grades with Summative



The second CHAID analysis with two written exam scores (summative_1 and summative_2) are removed and the predictors are not changed while the decision tree is determined the same for sixth grade students with eight independent variables.

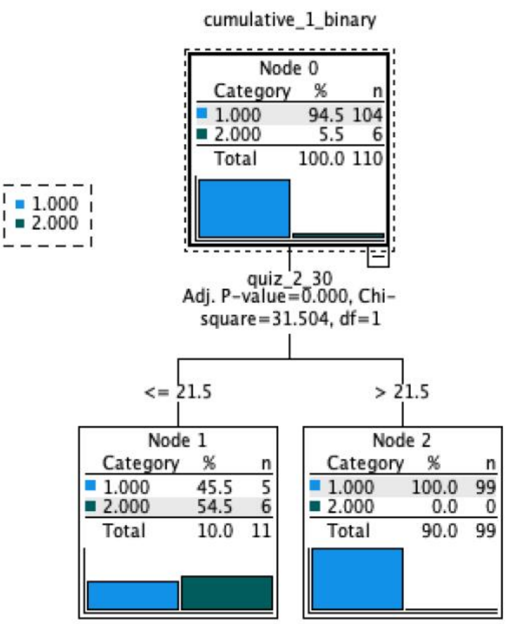
Decision Trees for Grade 7

Figure 5 shows the decision tree results of 110 seventh grade students. The main node of seventh grade students has 104 (94.5%) successful and six (5.5%) unsuccessful students. At this grade level, students are mostly successful, the decision tree analysis identifies 11 unsuccessful students in the same node. The mean (25.82) of the third and fourth quiz grades (quiz_2_30) is the most significant predictor, according to the CHAID analysis, which is conducted with eight independent values. This predictor creates only two nodes with no child node. Node

1 includes 11 students, and five (45.5%) are successful. The mean of the quiz_2_30 of students in this node is 21 or lower. The 99 students in node 2 are all successful (100%) and there is no unsuccessful student in this node. The students get over 21 from quiz_2_30 (see Figure 5) in node 2. The overall classification rate for the analysis of seventh grade students is 95.5%. The correct classification percentage is 95.2% for successful students and 100% for unsuccessful students.

Figure 5

Decision Tree for the Prediction of Grade 7 Students' First Semester Grades with Summative



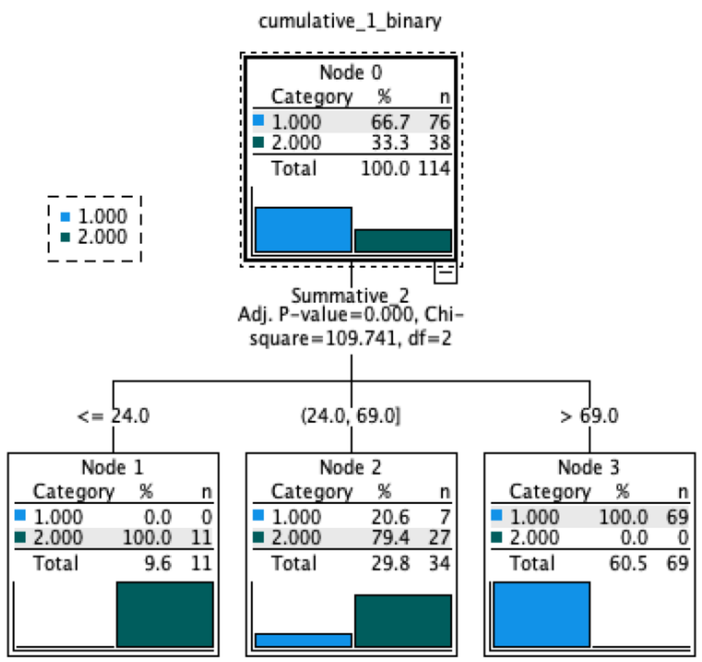
The same analysis is then repeated with six independent values and the predictors stay the same while the decision tree for seventh grade students is constructed.

Decision Trees for Grade 8

The main node of eight grade students has 76 (66.7%) successful and 38 (33.3%) unsuccessful students (see Figure 6). According to the CHAID analysis with includes eight independent predictors the strongest predictor is the second written exam (summative_2) with a mean of 70.83. Three nodes produce by this predictor. Eleven students are a part of node 1, and none of them are successful. The second written exam grades of students in this node are 24 or lower. Node 2 includes 34 students, seven of whom are successful (20.6%) and the rest of them (79.4%) are unsuccessful. The second written exam grades of students in this node are between 25 and 69, inclusive. There are 69 students in Node 3, and all of them (100%) are succeed and score higher than 69 on the second written exam. There is no student who is unsuccessful in this node as can be seen below in Figure 6.

Figure 6

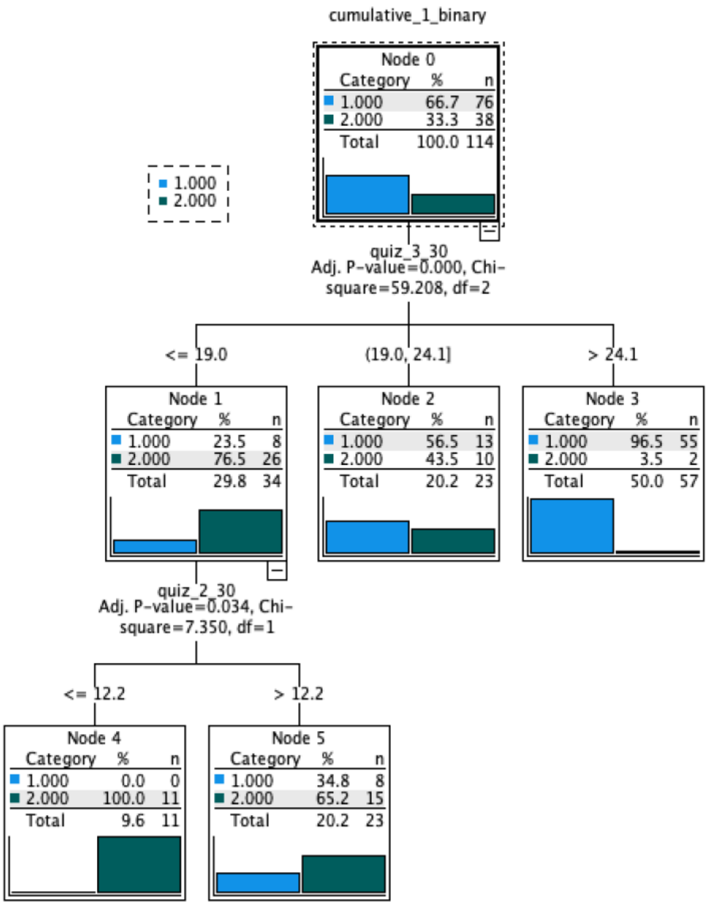
Decision Tree for the Prediction of Grade 8 Students' First Semester Grades with Summative



The overall classification rate for eighth grade analysis is 93.9%. The correct classification percentage is 90.8% for successful students and 100% for unsuccessful students.

Figure 7

Decision Tree for the Prediction of Grade 8 Students' First Semester Grades without Summative



In Figure 7, the same analysis is done by removing two summative with six independent values like the previous analysis with eight independent values the main node has 76 (66.7%) successful and 38 (33.3%) unsuccessful students. But in this case, the mean of fifth and sixth quiz grades (quiz_3_30) can predict the student's success at the end of the first semester. This variable creates three nodes. Node 1 includes 34 students, eight of them (23.5%) are successful and 26 of them are unsuccessful (76.5%). The fifth and sixth quiz grades (quiz_3_30) of students in this node are 19 or lower. The second leaves of Node 1 which are Node 4 and Node 5 consisted of the average of third and fourth quiz grades (quiz_2_30). In node 4 there are 11 students and all of them (100%) are unsuccessful. The mean of quiz_2_30 in this node is 12 or lower. Node 5 includes 23 students, eight of them (34.8%) are successful and 15 of them (65.2%) are unsuccessful and get over 12 from quiz_2_30. Node 2 includes 23 students, 13 of them (56.5%) are successful and 10 of them (43.5%) are unsuccessful. In this node the mean of quiz_3_30 was between 18 and 24, inclusive. There are 57 students in node 3, and 55 of them (or 96.5%) complete the fifth and sixth quizzes (quiz_3_30) successfully and get marks higher than 24 on average. There are two students (3.5%) who are unsuccessful in this node (see Figure 7). For this grade level, the overall classification rate is 82.5%. The correct classification rate is 89.5% for students who are successful and 68.4% for those who are not. As compared to the previous analysis this tree is less effective at grouping unsuccessful students into the appropriate group.

For the first semester, there is no difference between the correct classification rates of the six variable and eight variable analyses in the sixth and seventh grades, while a change is observed in the correct classification rates of the analyzes made in the fifth and eighth grades. In both grades, the rate of correct classification is higher with eight variable analyses.

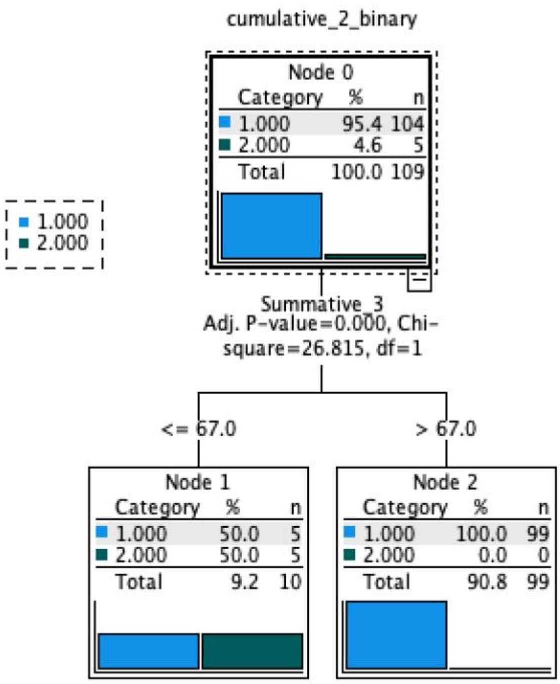
Second Semester Decision Trees for Grades 5, 6, 7, and 8

Decision Trees for Grade 5

The outcomes of 109 fifth grade students' decision trees are displayed in Figure 8. There are 104 (95.4%) successful students in the main node of fifth graders and five (4.6%) are unsuccessful ones. According to the CHAID analysis, which considers eight independent values, the third written exam grades (summative_3) are the most major predictor with a mean of 88.88. Only two nodes without any child nodes are produced by this predictor. Ten students are in Node 1, and only half of them are successful. Students in this node receive scores on their third written exam (summative_3) of 67 or less out of 100. There are 99 students in Node 2, and all of them are successful (100%). The third written exam (summative_3) result for the students in this node is over 67 (see Figure 8).

Figure 8

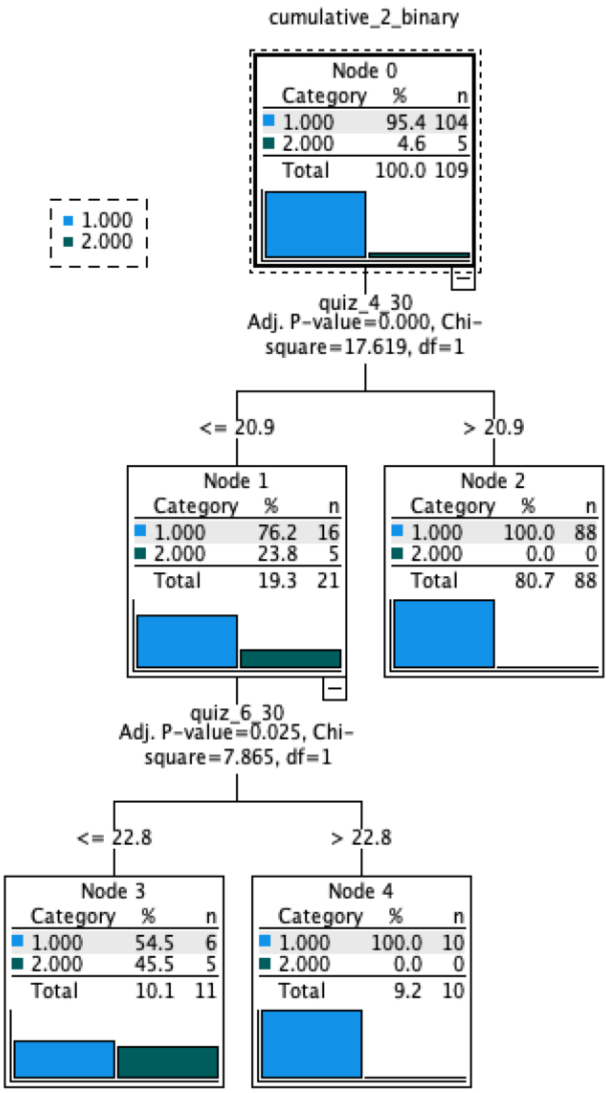
Decision Tree for the Prediction of Grade 5 Students' Second Semester Grades with Summative



For the analysis of students in the fifth grade, the total classification rate is 95.4%. For all 104 students, the correct classification percentage is 100%.

Figure 9

Decision Tree for the Prediction of Grade 5 Students' Second Semester Grades without Summative



In Figure 9, two written exam scores (summative_3 and summative_4) are dropped. The same analysis is subsequently performed using six independent values, this time, the results are combined as the main node of 109 fifth grade students have 104 (95.4%) successful and 5 (4.6%) unsuccessful students (see Figure 8). The mean of the seventh and eighth quiz grades in the second semester (quiz_4_30) is 25.84 found to be the most significant predictor. Two nodes are produced by this variable. There are 21 students in Node 1, and 16 (76.2%) of whom complete successfully, while the remaining seven (23.8%) do not. Students in this node score 20 or below on the average of the seventh and eighth quizzes (quiz_4_30). Quiz_4_30 grades make up the second leaves of Node 1, which are Node 3 and Node 4. There are 11 students in Node 3, six of whom (54.5%) are unsuccessful, and five of whom (45.5%) are successful. Students in this node have average quiz grades in quiz_6_30 that were 22 or below. Ten students make up Node 4, and all of them (100%) achieve success by scoring higher than 22 on the mean of their eleventh and twelfth quizzes (quiz_6_30). Node 2 includes 88 students and there is no unsuccessful student in this node, all of them (100%) are successful.

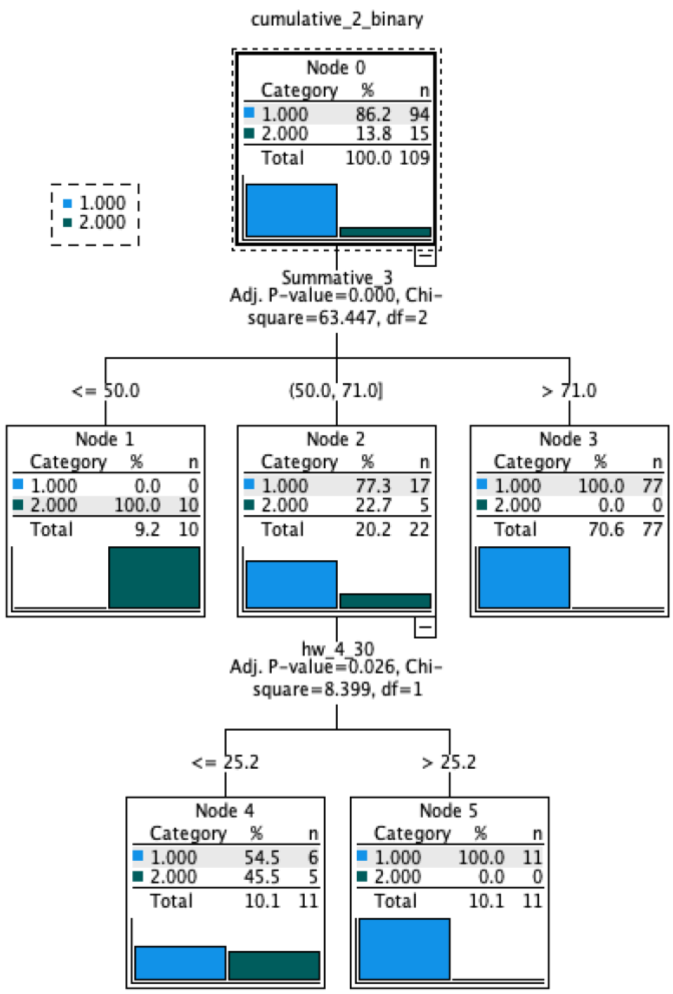
Students in this node have quiz_4_30 that are greater than 20 (see Figure 9). The overall classification rate for fifth grade student level is 95.4%. The percentage of successful students who receive the right categorization is 100%. Both studies show the same rate of classifying students into the correct category as the prior tree, which includes eight independent predictors.

Decision Trees for Grade 6

The decision tree findings from 109 sixth graders are shown in Figure 10. There are 94 (86.2%) successful students in the main node of sixth graders and 15 (13.8%) of them are not successful. The third written exam (summative_3), which is included in the CHAID analysis along with eight independent values is the strongest predictor with a mean of 77.85. The predictor produced three nodes. Ten students are in Node 1, and there is a success rate of 0%. Students in this node have scores of 50 or below on their third written exam (summative_3). 22 students are enrolled in Node 2, 17 (77.3%) of whom are successful, while five (22.7%) are unsuccessful. Results from the third written exam (summative_3) for the students in this node ranged from 51 to 71, inclusive. The average of the homework (homework_4) that is assigned for the first performance grade of the second semester makes up the second leaf of Node 2, which is composed of Node 4 and Node 5. There are 11 students in node 4, six of whom (54.5%) are successful, and five of whom (45.5%) are not. Students in this node receive 25 or less on average on their homework. 11 students make up Node 5, and all of them (100%) succeed in and receive more than 25 on homework_4. There are 77 students in Node 3, and all of them (100%) achieve success and score higher than 71 on the third exam. In this node, there is not a single unsuccessful student (Figure 10).

Figure 10

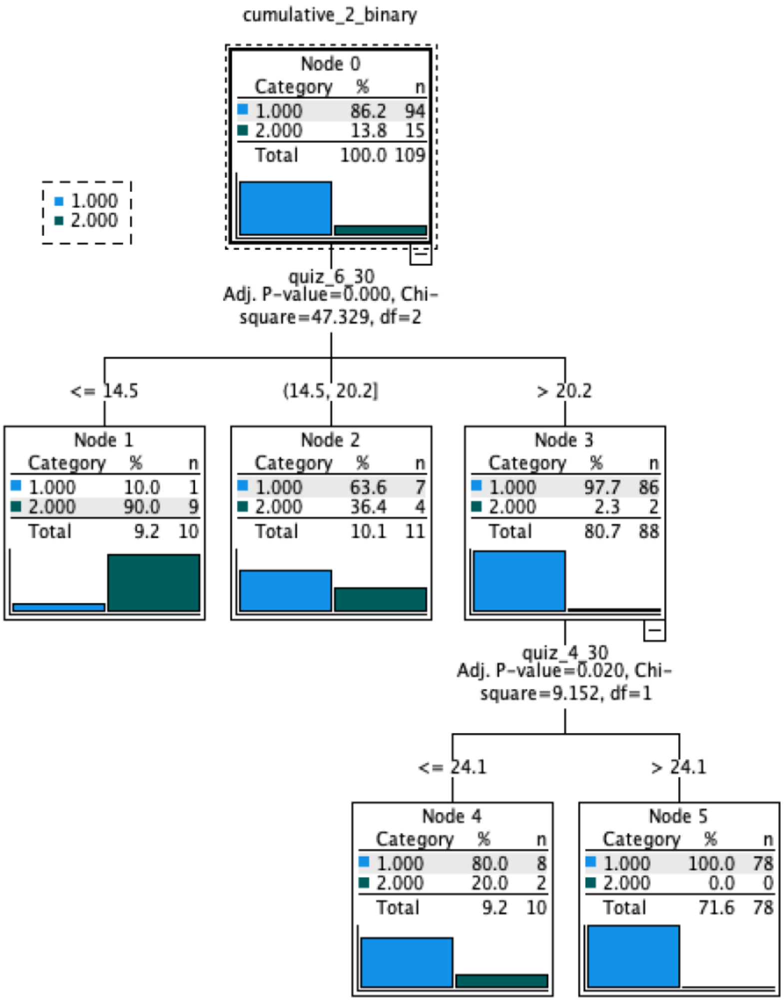
Decision Tree for the Prediction of Grade 6 Students' Second Semester with Summative



For this grade level, the overall classification rate is 95.4%. The accurate categorization percentage is 100% for students who are successful and 66.7% for those who are not.

Figure 11

Decision Tree for the Prediction of Grade 6 Students' Second Semester Grades without Summative



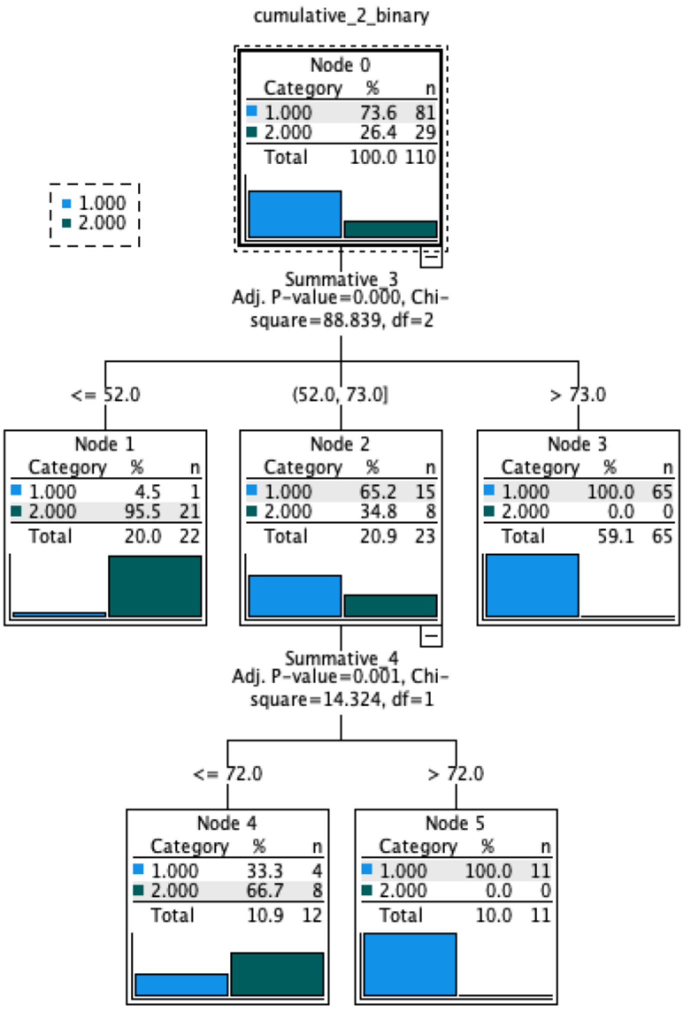
The same analysis is then carried out using six independent values by disregarding summative_3 and summative_4, as shown in Figure 11. The main node has 94 (86.2%) successful and 15 (13.8%) unsuccessful students as in the previous tree. However, in this case, the most accurate indicator of the student's success is the average of the eleventh and twelfth quiz grades (quiz_6_30) generated for the final performance grade of the second semester. The mean of quiz_6_30 is found 24.37. This study also produces three parent nodes. Only one (10%) of the 10 students in Node 1 is successful, while nine (90%) are unsuccessful. The mean quiz grade for students in this node in the quiz_6_30 is 14 or below out of 30 points. There are 11 students in Node 2 for quiz_6_30 for the last performance grade; seven of them (63.6%) are successful, and four (36.4%) are unsuccessful. The quiz_6_30 in this node ranges from 15 to 20, inclusive. Only two (2.3%) of the 88 students in Node 3 are unsuccessful in quiz_6_30 and received less than 20 points. The average of the seventh and eighth quiz grades (quiz_4_30) is 26.88, which is calculated for the fourth performance grade of the second semester, making up the second leaves of Node 3, which are Node 4 and Node 5. Ten students are enrolled in Node 4, eight (80%) of them are successful, and two (20%) are unsuccessful. Students in this node scored an average of 24 or below on quiz_4_30. There are 78 students in Node 5, and all of them (100%) achieve success and score higher than 24 on quiz_4_30. The overall classification rate for sixth grade is 93.6%. The correct classification rate for successful students is 98.9%, whereas for the unsuccessful ones it is 60%. This tree is less effective than the preceding tree at grouping students into the appropriate grouping.

Decision Trees for Grade 7

The outcomes of 110 seventh grade students' decision trees are displayed in Figure 12. In the main node of seventh graders, there are 81 (73.6%) successful students and 29 (26.4%) unsuccessful ones. Based on the CHAID analysis, which considers eight independent values, the third written exam grades (summative_3) are the most crucial predictor which has a mean of 75.22. The predictor produces three nodes. 21 students are enrolled in Node 1, and only one (9.1%) of them is successful. Students in this node have third-exam grades (summative_3) 52 or less. There are 23 students in Node 2, 15 of them (65.2%) achieve success, while the remaining eight (34.8%) do not. Results from the third exam for the students in this node range from 53 to 73, inclusive. The fourth written exam (summative_4) results with a mean of 74.49 for the second semester are contained in the second leaves of Node 2, which are composed of Node 4 and Node 5. There are 12 students in Node 4, four (33.3%) of them achieve success, and eight (66.7%) of them do not. Students in this node receive scores of 72 or lower on their fourth exams (summative_4). There are 11 students in Node 5, and all of them (100%) complete the fourth exam (summative_4) successfully. Successful students in this node score more than 72 out of 100 on summative_4. In Node 3, there are 65 students, and all of them (100%) succeed and achieve scores greater than 73 on the third written exam (summative_3). This node in Figure 12 above has no unsuccessful students and no child nodes.

Figure 12

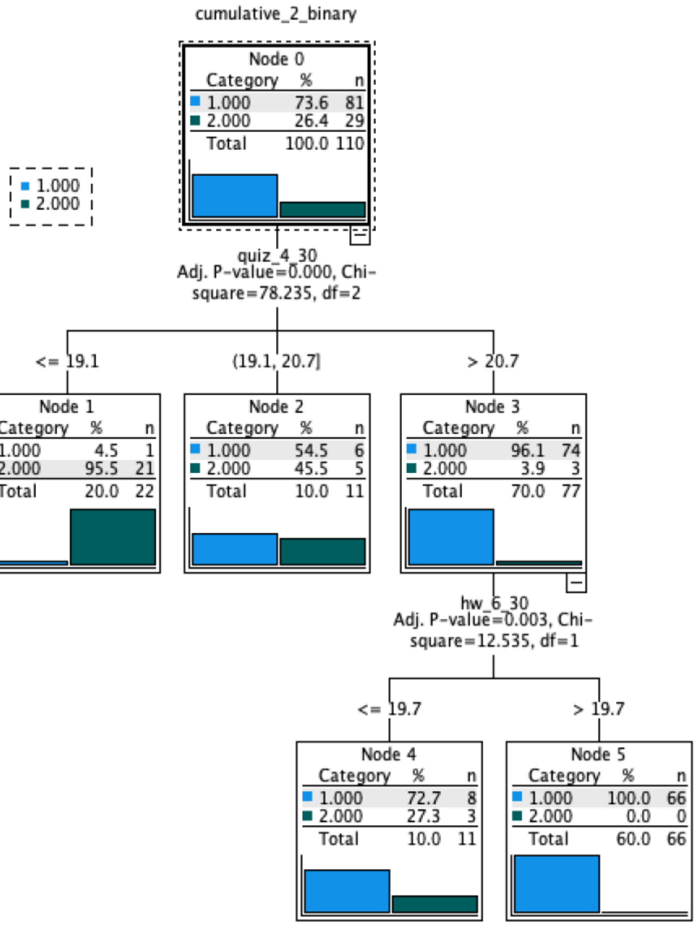
Decision Tree for the Prediction of Grade 7 Students' Second Semester Grades with Summative



The seventh grade overall classification rate is 94.5%. The correct classification percentage is 92.7% for those who are successful and 100% for those who aren't.

Figure 13

Decision Tree for the Prediction of Grade 7 Students' Second Semester Grades without Summative



The same analysis is conducted (see Figure 13) using the six independent variables by removing the two written exams (summative_3 and summative_4), from the previous analysis values in Figure 13. Figure 13 shows the main node has 81 (73.6%) successful students and 29 (26.4%) unsuccessful ones. However, in this instance, the student's achievement in the second semester may be predicted using the mean of the seventh and eighth quiz grades (quiz_4_30). The variable produces three nodes. 22 students are enrolled in Node 1, one of whom (4.5%) complete quiz_4_30 successfully, while the remaining 21 (95.5%) are not. Students in this node scored 19 or below on quiz_4_30. 11 students are represented by Node 2, six of

whom (54.5%) are successful and five (45.5%) are not. The average quiz grade of the seventh and eighth quizzes (quiz_4_30) for students in Node 2 was 20. In Node 3, there are 77 students; 74 of them (96.1%) successfully finish quiz_4_30 and get on greater than 20. In this node, three students (3.9%) are not successful (see Figure 13). Node 3 has two child nodes which are Node 4 and Node 5 based on the mean of homework (hw_6_30) that is given for the sixth (last) performance grade of the second semester in Figure 12. The mean of hw_6_30 is 24.24 out of 30 points.
In Node 4, there are 11 students, eight of them (72.7%) succeed while three of them (27.3%) are unsuccessful. The average grade for hw_6_30 given to students in Node 4 is 19 or less. There are also 66 successful students (100%) in Node 5 and none of them is unsuccessful. Node 5 consisted of 66 students, and all (100%) of them were successful and received more than 19 on hw_6_30. The overall classification rate is 90.9% for this grade level. The correct percentage is 97.6% for successful students and 71.4% for those who are not.

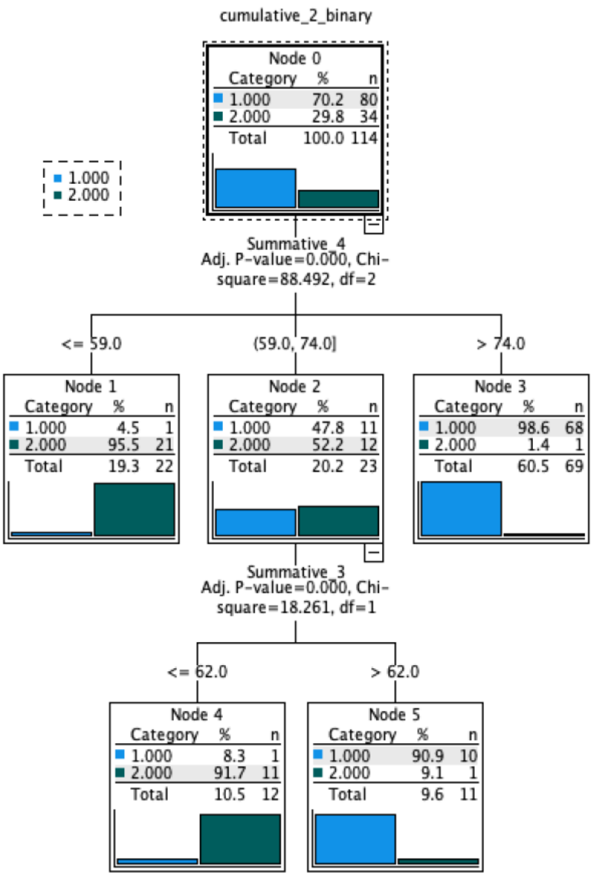
Decision Trees for Grade 8

The results of 114 eighth grade students' decision tree in the second semester are shown in Figure 14. There are 80 (70.2%) successful students and 34 (29.8%) unsuccessful students in the main node of eighth graders. The fourth written exam grades (summative_4) are the most important predictor with an average of 76.67, according to the CHAID analysis, which takes eight independent values into account. The predictor generates three nodes. 22 students are a part of Node 1, and only one of them (4.5%) is successful. 21 students (95.5%) are unsuccessful in this node. The fourth written exam grades (summative_4) of students in this node are 59 or lower. Node 2 includes 23 students, 11 of whom are successful (47.8%) and the rest of them (52.2%) are unsuccessful. The fourth written exam grades of students in

Node 2 were between 60 and 74, inclusive. The second leaves of Node 2 which are Node 4 and Node 5 consist of the third written exam grades (summative_3) of the academic year. In Node 4 there are 12 students, only one of them (8.3%) is successful and the rest of them (91.7%) are unsuccessful. The third exam grades (summative_3) of students in this node are 62 or lower. Node 5 included 11 students, only one of them (9.1%) is unsuccessful and 10 of them (90.9%) are successful and got over 62 from the third written exam. There are 69 students in Node 3, and only one of them (1.4%) is not successful and scored less than 75 on the fourth written exam grades (summative_4). There are 68 students who are successful in this node seen above in Figure 14.

Figure 14

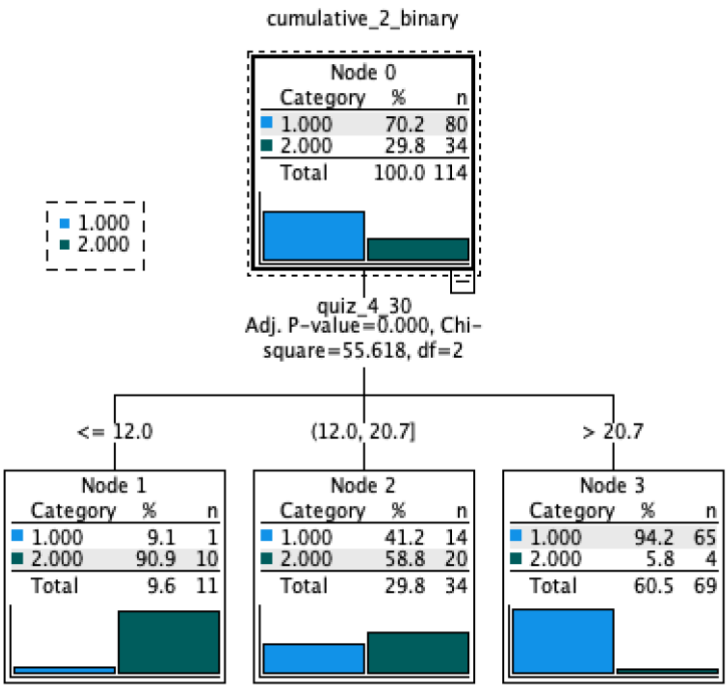
Decision Tree for the Prediction of Grade 8 Students' Second Semester Grades with Summative



For the eighth grade analysis, the total classification rate is 96.5%. The correct classification rate for successful students is 97.5%, while for those who are unsuccessful is 94.1%.

Figure 15

Decision Tree for the Prediction of Grade 8 Students' Second Semester Grades without Summative



Two written exam grades (summative_3 and summative_4) are excluded and the same analysis is conducted using six independent values, as shown in Figure 15. The main node has 80 (70.2%) successful and 34 (29.8%) unsuccessful students. The average of the seventh and eighth quiz grades (quiz_4_30) calculated with the average of 22.03 is the strongest predictor of the student's achievement in the second semester. This predictor creates three nodes with no child nodes. Node 1 includes 11 students and only one of them (9.1%) is successful. The mean of seventh and eighth quiz grades (quiz_4_30) in this node are 12 or lower out of 30. Node 2 includes 34 students, 14 of whom are successful (41.2%) and the rest of them (58.8%) are unsuccessful. In this node, students' results from quiz_4_30 are between 13 and 20, inclusive. Node 3 includes 69 students and four of them (94.2%) are successful and get over 20. There are four students who are unsuccessful in this node (see Figure 15). The overall classification rate for eighth grade level is 83.3%. The correct classification percentage is 81.3% for successful students and 88.2% for unsuccessful students.

Summary

While no child nodes are formed in any of the grades in the first semester with eight variable analyses, child nodes are formed at three grades (except grade 5) of the second semester with eight variable trees. In the analyses with six and eight variables, no change is observed in the sixth and seventh grades’ decision trees without child nodes, while all decision trees in the second period are formed with different nodes and child nodes.

Based on semesters, the overall classification rate is over 93% in all eight variable analyses in the first semester. The lowest overall classification rate among all six variable analyses is 82.5% which is in the first semester of eighth grade decision tree. Based on the grade levels and semesters, although the resulting decision trees are different from each other in terms of the number of child nodes, predictors, or averages the only tree that has the same grouping table, and the correct classification percentage is 0% is for five unsuccessful fifth grade students in both six and eight variables in the second semester.

CHAPTER 5: DISCUSSION

Introduction

This chapter aims to provide a summary of the study and its major findings. Then, based on the discussions and the results, implications for practice and implications for future research are presented. Limitations to the study can be found at the end of this chapter.

Overview of the Study

The importance of developing students' analytical thinking, problem-solving, and flexible and divergent thinking skills is increasing day by day around the world. Students are developing these skills throughout their education, including in their mathematics classes.

To ensure students have acquired these skills, mathematics teachers, like all teachers, use measurement tools to assess student learning. Typically, student achievement is evaluated and reflected in a final, cumulative grade. Although teachers give a final grade at the end of the academic semesters and years, predicting students' mathematics achievement at earlier stages can also enable teachers to work more effectively on planning and differentiating.

Teachers have a variety of summative and formative assessments that they can use both in and out of class; however, not all scores can be considered equal in predicting students’ future achievement. Thus, determining more significant predictors can help identify students’ potential for success or likelihood of failure. Knowing which students are succeeding and which are struggling gives teachers insights into how they can adjust learning activities and resources to be more suitable to the students’ learning needs. If successful students are detected early, different studies (called early finishers at school in this study) can be assigned to students, or students can be prepared by directing them to areas such as advanced mathematics, academic Olympics, and international projects. Likewise, differentiated studies can be assigned to students with the potential to fail, students can be directed to additional work hours after school, parents can be contacted, and support can be requested at home.

The formative and summative assessments in this study were used in grade 5 through 8 mathematics courses at a private school in Istanbul. The study included assessments from two semesters. The assessments included homework, quizzes and written exams. Of particular interest in this study was the predictive power of homework as this formative assessment required extensive teacher time to prepare, administer, and grade. Furthermore, students seemed not to value the assignments and it would be helpful to learn how effective they were in giving insights into students’ academic success.

In this study, Chi-squared Automatic Interaction Detection (CHAID) was used to create decision trees to distinguish students as successful and unsuccessful using several predictors. To this end, cumulative mathematics grades were selected as the criterion score. Based on the school grading policy, students who scored 70 and above were labeled as successful, and those scoring below 70 were labeled unsuccessful. The decision tree analyses were conducted for grades 5, 6, 7, and 8 level using students' mathematics scores they got during the academic year, based on each grade with overall grades which were calculated end-of-semester and examined for each semester separately. Specifically, the predictor variables were formative and summative. The formative included six homework grades which were calculated by averaging all the homework given before each of the six performance grades and six quiz grades which are calculated by averaging the two quizzes applied before each of the six performance grades. The summative assessments were four written exam grades. The cumulative, end-of-semester grade (the dependent variable, was the total score of the formative and summative assessments).

CHAID analyzed the relationships between the independent variables (the formative and summative assessments) and the dependent variable (the end-of-semester, cumulative grade). The purpose of using CHAID analysis is to determine which variables have more predictive power for the dependent variables (either continuous or categorical) by maximizing differences for predictor variables to create subgroups. Predictors with the highest predictive power are considered earlier in the CHAID analysis. In this way, predictors are ordered in terms of their significance or power in predicting the dependent variable.

The study offers responses to the following research question:

Among the formative and assessment measurement tools used in mathematics courses in grades 5 through 8, which are the most important predictors of the cumulative semester score?

The analysis showed that generally, quiz grades for the first semester can predict the success of the end of the semester. When it comes to the second semester, the exams are better predictors. There is not much difference among the different grade levels. Detailed results were given in the discussion of major findings.

Discussion of Major Findings

The main findings of this study are that while quizzes and written exams can be powerful predictors of end-of-semester, cumulative grades, homework provided no evidence of predictability. Before reviewing important insights gained from the findings about homework, a discussion of the more predictive variables is discussed.

General Findings Regarding Quizzes and Exams: It Depends on the Grade Level

In general, the first semester achievement mostly seems to be predicted by quizzes. The analysis revealed, however, that in the second semester, for grades 5 and 8 written exams became a more powerful predictor while quizzes remained more powerful predictors for students in grades 6 and 7. In the next section, highlights of the findings for the grade levels are presented and compared. Following the grade level discussion, findings regarding the written exams and the potential predictive role of quizzes in this study are discussed. Table 5 provides an overview of the findings for this discussion.

Table 5

Key Predictors Found in This Study

Grade Level	Semester 1		Semester 2	
	Eight predictors	Six Predictors	Eight predictors	Six Predictors
5	second exam	first and second quizzes; fifth and sixth quizzes	first written exam	first and second quizzes; fifth and sixth quizzes
6	third and fourth quizzes	third and fourth quizzes	first written exam; fourth homework	fifth and sixth quizzes; first and second quizzes
7	third and fourth quizzes	third and fourth quizzes	first exam; second exam	first and second quizzes; sixth homework
8	second exam	fifth and sixth quizzes; third and fourth quizzes	second exam; first exam	first and second quizzes

For grade 5 student scores, the best predictor is the first two quizzes in the first semester of the fifth grade and student success can be predicted by looking at the first exams in the second semester. In particular, the third written exam was the strongest predictor of the condition of 5 out of 5 students (100%) for the grade 5 students who were unsuccessful at the end of the semester. Exams may be a stronger predictive tool as fifth grade students started to get used to the secondary school homework and quiz system as of the second semester. In the research conducted by Kıncal and Ozan (2018) with fifth grade students, it was found that formative assessment strategies increased students' academic performance and achievement. The improvement of performance grades of the fifth grade students in this study may have enabled the summative assessment exam grades to be a strong predictor of student achievement.

In grade 8, the most powerful predictor was the written exams given at the end of the second semester. In fact, the study revealed that at the end of the second semester, student points increased on average in both exams compared to the scores in the first semester (first semester exams: 24 and 69 points; second semester exam points: 59 and 74 points). The reason for this increase in predictor averages may be the personal grade point average target, external exam targets, or graduation motivations that the secondary school senior students set for themselves.

For grades 6 and 7, quizzes were the strongest predictors for both semesters. These findings will be discussed further below. In addition, the discussion will note that when the written exams are removed from the predictors, quizzes play a powerful role for all grade levels.

Written exams

The first exam of the first semester was found not to be predictive at any of the grade levels. On the other hand, the first exam of the second semester is the strongest predictor for each grade in the second semester. Since the students' motivation at the beginning of the first semester/academic year is high and the number of topics covered in the exam is low, the selectivity of the first exam may be lower in the first semester, but the second exam is more selective in predicting student success in fifth and eighth grade levels. According to Butler (1988), summative assessments are extrinsically motivated for high achievers, students' narrow field of interest reduces the quality, if not the quantity, of their performance. The findings reported in this study might support Butler's motivational comment based on students' selectivity of assessment material. Furthermore, one of the reasons for having the second semester's first exam as the strongest predictor may be that the students who passed the first semester unsuccessfully started the second semester with higher achievement goals and motivation.

A key finding of this study is that the second exams of the first semester are effective predictors in the fifth and eighth grades. It should be noted that when the exams were removed, the quizzes came out as predictors. Thus, it can be determined that quizzes are still strong predictors depending on semesters, grades, or independent predictors. The fifth and eighth grade levels are the first and last levels of middle school. The fifth graders start the semester with the goal of being successful, and while some of the eighth-grade students increase their exams by focusing on the graduation score, the other part of eighth grade students focus more on closed-ended questions for the nationwide transition to high school exam (TEOG) goals instead of open-ended questions. This difference in student goals may be the reason for the change in strong predictors. But it is clearly established that quizzes and exams can predict student achievement better than homework given during the first semester.

Quizzes

A valuable finding of this study is the power of quizzes to be strong predictors for overall student achievement. According to Dengri et al. (2021), the advantages of quiz-based learning are that teachers to tailor the subject matter and difficulty level to the individual needs, skills, and limitations of students, and Shirvani (2009) promoted weekly quizzes to improve students' achievement. Zarei (2008) supported the students studying more for the quizzes since quizzes are a source of motivation for them. This is an important prediction for the teacher.

In this study, this was especially true for sixth and seventh grade where the predictive quizzes were administered halfway through the first semester. In particular, it was the average of the third and four set of quizzes in these grades that were most predictive. When these quizzes were compared to the average of the first two set of quizzes, it was determined that there was no difference in the variety of question types that make up the quizzes. The types of questions for all the quizzes were mostly open-ended and consisted of multiple choice, fill-in-the-blanks, comparison, and true-false questions. The topic of the predictive exams for sixth grade students was sets and for the seventh grade it was rational numbers. These topics may support the predictive power of the quizzes as they challenge the habits of associating arithmetic with symbols and abstract thinking for some students.

Quizzes also had some predictive value for students in grade 5, however for them it was the first two sets of quizzes that could determine 10 out of 12 (83%) of the students who failed at the end of the semester. This may be because students who score low on the first two quizzes may not be able to develop their motivation and study sufficiently, or as fifth grade student who has just started middle school, they may not be familiar with these assessment or question types. Another reason may be that fifth grade students are in the process of getting used to the middle school system, measurement and assessment materials, and the variety of questions for the fifth grade students. For example, Funk and Dickson (2011) mentioned that using various test-taking techniques such as finding the right answers by eliminating or guessing the wrong answers, using the skills of recognition, recall, and analysis can be used to determine the correct answers to the multiple-choice questions. Since such techniques will be new for fifth grade students, they may have caused differences in the predictors of achievement. According to the growth and cognitive development of the students, there will be changes in the variety of questions and educational approaches according to grade levels, but such predictive changes occur because these changes affect the success of the students. According to Jitendra and Dupuis (2015), it is essential for students whose ability, success, and motivation differ to acquire the mathematical skills required to satisfy different grade-level criteria.

For eighth grade students, the best predictors in both semesters were exams, but in the case of analysis with six predictors by removing exam predictors, the strongest predictors were only quizzes. At this level, the paired averages of the last four quizzes of the first semester were determined as predictors, while the average of only the first two quizzes of the second semester was determined as the strongest predictor. The reason for this may be that eighth grade students start the second semester more motivated after the first semester as they approach graduation. Singh et al. (2002) investigated the crucial role of individual and school variables on eighth grade students' interest, motivation, and achievement in mathematics and science. The variation in mathematical success is influenced by attitudes and motivating factors. They found that students with high motivation towards the lesson also had high success in process tracking and mathematics.

Homework: Pitfalls and Potentials

As discussed above, quizzes and written exams served as more powerful predictors of students' overall success. One homework may not be a strong predictor is while quizzes and exams are applied at a certain set time in class under standardized conditions, there can be flexibility in time, place, and setting for homework. As Corno (1996) labeled homework as “schoolwork that is taken home” (p. 22) some students may want to complete schoolwork at school and manage their own time at home according to their interests.

In addition, while correct and incorrect statements are graded in quizzes and exams, responsibility tracking is assessed rather than right and wrong in homework. The care taken on homework by the fifth grade students in the first year of middle school and the target of a high general point of average for the eighth grade students may have contributed to the general increase in homework grades. From another point of view, the eighth graders' focus on the graduation process, the increase in their absenteeism, or their focus on external exams may have caused a general decrease in homework grades. Therefore, in this grade level of fifth and eighth, homework was not determined as a strong predictor. Homework predicted student achievement only for sixth and seventh grade students in the second semester.

The reason why homework is not as strong a predictor as other assessment tools is an important finding for education. One of the reasons why homework is the weakest predictor of student achievement may be the homework scoring system. No matter how late students turn in their assignments, they get three points out of four points from that assignment. The aim of the school in this system is to support students' motivation to teach and to give them an opportunity to reinforce the subject by doing homework even if a student learns a topic late. However, when the homework grades were examined, the rate of doing homework in some of the students, especially in the second semester, decreased significantly. Students may have decreased their intrinsic motivation to do homework.

There were some indications that in grades 6 and 7, homework has the potential to predict student achievement, especially in the second semester. Perhaps in the first semester, students are becoming more aware of the importance of attendance and being responsible for their homework. These habits carried over to the second semester, thus enabling homework to be a predictor in the second semester. Another finding of the study is that in grade 7 the sixth homework assignment was found to be predictor for around 80% of the students who were successful. This success rate based on homework is supported by the findings of Cooper (1989) that homework lasting one or two hours affects middle school success. De Jong et al. (2000) also found the effectiveness of mathematics homework and they found that the method of incorporating the mathematics curriculum into homework and the amount of homework given were effective on student achievement. In the school where this study was conducted, the content of homework was prepared by the teachers specifically and the frequency was planned according to the intensity and success of the students, so success in the sixth and seventh grades became a significant predictor. On the contrary, the decrease in the rate of doing second-term homework, in general, may have provided this selectivity.

Implications for Practice

This study was launched to explore how various forms of assessment can help teachers predict student success. The outcomes of the study can help teachers better plan their semesters to apply assessments effectively. Teachers can predict students who have the potential to fail at the end of the semester with their quiz grades in the first semester and their exam grades in the second semester. Assessment material can be ordered and examined closely as to why they have lower predictive power. Instead of realizing student success at the end of the semester, teachers can plan for both successful and unsuccessful students in the middle of the semester. That is, teachers can invite students with the potential to be unsuccessful to additional after-school working hours, assign reinforcement work, or communicate with their parents to plan cooperation on the follow-up and responsibility of additional work that students need to do.

Within the context of the current study, the role of homework was of particular interest. There was concern that teachers were spending an inordinate amount of time preparing for and grading homework which might not be useful to provide insights into student learning. This study had positive and negative findings regarding homework. The schools and teachers must carefully balance homework systems based on student responsibility tracking, homework motivation, and homework scoring systems. Teachers spend a lot of time and effort on homework preparation and checking them. Therefore, the homework system of schools is very important for both students' learning perspective and teacher motivation.

Finally, this study has shown that teachers can use CHAID, or decision trees, in general, to get predictive information about student’s end-of-semester achievement levels at earlier stages of the academic years. If teachers' knowledge about statistical analysis is not sufficient, supporting them with training on this subject will be efficient for teachers to plan both in interpreting the data consisting of student grades and in predicting achievement. The software which is called Statistical Package for the Social Sciences (SPSS) online training might be the best way to learn how to use CHAID or decision trees.

Implications for Further Research

The quantitative technique was used in this study. Further research may use the CHAID analysis made can be replicated with a large-scale sample. Grade data of student groups can be examined in a longer-term instead of an academic year.

In addition, by interviewing the teachers who prepare the assessment materials and teach at the school, the relationships between the achievements of the students and teacher observation, which is one of the formative assessment tools that cannot be used in this study, can be learned.

Other variables related to students can be added. If there are international and external exams applied during or at the end of the year, they can be used as a dependent or independent predictor in the data group as student achievement data. A similar study can be done with other main subjects (Science, English, or Turkish) or the overall year-end success of the students can be examined together with other academic performances.

Analysis can also be applied using training data. For example, the collected data can be split into 20% and 80%. CHAID analysis is performed on 20% of data and predictors are determined. Then, using these predictors in 80% of the students, the performance of the students as successful or unsuccessful is examined.

Limitations

In this study, mathematics quizzes, homework, and exam grades were used to predict the academic success of middle school students in one school. At this point, it should be underlined that the predictors defined by the CHAID analyses are only the strongest predictors. This does not mean that other predictors, not included in the trees, do not have predictive power. However, in this study, the number of students at each level is limited to 109 (number of fifth grade students) is the minimum number of students in all grade levels, and 114 (number of eighth grade students) is the maximum number of students. Thus, the number of the child nodes was limited by the research not to get subgroups with few students within. In other words, since the sample size was small in this study, it was difficult to get new offshoot node data from the nodes. If more offshoots could be taken, homework would also appear as a predictor. CHAID analyses order the predictors in degree of predictive powers, not in a binary classification such as predictor and not predictor.

According to the results of the decision trees, the overall classification rate gave a successful result (average 95% of correct percent) for all analyses. Especially strong results were obtained in identifying successful students, but the results for unsuccessful students are weaker. In this regard, it is useful to be cautious when trying to identify unsuccessful students with CHAID. For students identified as unsuccessful by CHAID, it may be considered to seek additional evidence such as assessments of other disciplines, external exam success, or projects to support their probability of failure.

It has a very limited sample size, so creating training data was not possible. Using a larger sample size with more than one school can provide more reliable results for the research. Since the school where the research was conducted is a private school, the population is less than in public schools. Extending the sample size may assist in producing results that are more in-depth and reliable.

However, there is a poor number of students at each level in this study. In other words, it is challenging to extract new child node data from the nodes due to the small sample size. Other predictors would also be a predictor if more child nodes could be taken. Additionally, CHAID ranks predictors in order of importance. Because no new child nodes are created due to sample size, variables that might be (less) important were not displayed here.

On the other hand, students in this research may be described as coming from high-income families, and they are generally children of academically and socially supportive parents. Other factors, such as socioeconomic position, school type, and family expectations and support, could impact students' academic achievement. White (1982), Liu and Lu (2008), and Long and Pang (2016) worked on this argument. They claimed that family socioeconomic situation and parental expectations are important variables in educational success.

REFERENCES

Akış, D. (2020). *Predicting high school students' academic achievement in international and national programs based on admission examination: a case study from Turkey* [Unpublished master thesis]. Bilkent University.

Alghamdi, A. K. H., & Al-Hattami, A. A. (2016). The accuracy of predicting university students' academic success. *Risālat al-Tarbiyah wa-'Ilm al-Nafs, 186*(3855), 1-10.

Bates, J. A. (1979). Extrinsic reward and intrinsic motivation: A review with implications for the classroom. *Review of Educational Research*, *49*(4), 557-576. https://doi.org/10.3102/00346543049004557

Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management*. John Wiley & Sons.

Berry, R. (2008). *Assessment for learning* (Vol. 1). Hong Kong University Press.

Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, *5*(1), 103-110. https://doi.org/10.1080/0969595980050106

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Granada Learning. https://doi.org/10.1177/003172171009200119

Boud, D. (1995). Assessment and learning: contradictory or complementary?. In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 35-48). Kogan.

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research*, *103*(3), 191-207. https://doi.org/10.1080/00220670903382970

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education*, *10*(2), 161-180. https://doi.org/10.1207/s15324818ame1002_4

Brookhart, S. M. (2005). The quality of local district assessments used in Nebraska's School-based Teacher-led Assessment and Reporting System (STARS). *Educational Measurement: Issues and Practice*, *24*(2), 14-21. https://doi.org/10.1111/j.1745-3992.2005.00007.x

Burke, K. (2010). *Balanced assessment: From formative to summative*. Solution Tree Press.

Burkhardt, H. & Schoenfeld A. (2019). (2019). Formative assessment in mathematics. In G. J. Cizek, H. L. Andrade, & R. E. Bennett (Eds.) *Handbook of formative assessment in the disciplines* (pp. 3-19). Routledge.

Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980.* (Research Report No. 2001-2). College Entrance Examination Board. https://files.eric.ed.gov/fulltext/ED562836.pdf

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1-14. https://doi.org/10.1111/j.2044-8279.1988.tb00874.x

Casillas, A., Robbins, S., Allen, J., Kuo, Y. L., Hanson, M. A., & Schmeiser, C. (2012). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*, *104*(2), 407. https://doi.org/10.1037/a0027180

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1-6. https://doi.org/10.1080/00098650903267784

Çelikel, F., & Karakuş, M. (2017). The analyzing the relevance of TEOG exam to academic achievement and the effects of TEOG exam on teaching process of math class. *Necatibey Faculty of Education Electronic Journal of Science & Mathematics Education*, *11*(2).

Chufama, M., & Sithole, F. (2021). The pivotal role of diagnostic, formative and summative assessment in higher education institutions' teaching and student learning. *International Journal of Multidisciplinary Research and Publications*, *4*(5), 5-15.

Conderman, G., Pinter, E., & Young, N. (2020). Formative assessment methods for middle level classrooms. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *93*(5), 233-240. https://doi.org/10.1080/00098655.2020.1778615

Cooper, H. (1989). Synthesis of research on homework. *Educational Leadership*, *47*(3), 85-91.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*(1), 1-62. https://doi.org/10.3102/00346543076001001

Corlu, M. S. (2014). Which preparatory curriculum for the International Baccalaureate Diploma Programme is best? The challenge for international schools with regard to mathematics and science. *International Review of Education*, *60*, 793-801.

Cornelius, K. E. (2013). Formative assessment made easy: Templates for collecting daily data in inclusive classrooms. *Teaching Exceptional Children*, *45*(5), 14-21. https://doi.org/10.1177/004005991304500502

Corno, L. (1996). Homework is a complicated thing. *Educational Researcher*, *25*(8), 27-30. https://doi.org/10.3102/0013189X025008027

Couch, B. A., Hubbard, J. K., & Brassil, C. E. (2018). Multiple–true–false questions reveal the limits of the multiple–choice format for detecting students with incomplete understandings. *BioScience*, *68*(6), 455-463. https://doi.org/10.1093/biosci/biy037

De Jong, R., Westerhof, K. J., & Creemers, B. P. (2000). Homework and student math achievement in junior high schools. *Educational Research and Evaluation*, *6*(2), 130-157. https://doi.org/10.1076/1380-3611(200006)6:2;1-E;F130

De Ville, B., & Neville, P. (2013). *Decision trees for analytics: Using SAS enterprise miner*. SAS Institute. SAS Institute Inc.

Dengri, C., Gill, A., Chopra, J., Dengri, C., Koritala, T., Khedr, A., & Jain IV, S. (2021). A review of the quiz, as a new dimension in medical education. *Cureus*, *13*(10).

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory into Practice*, *55*(2), 153-159. https://doi.org/10.1080/00405841.2016.1148989

Dondurmacı, G. (2011). *Veri madenciliğinde regresyon ağaçları ile sınıflandırma ve bir uygulama* (Classification with regression trees in data mining, and an application) [Unpublished doctoral dissertation]. Mimar Sinan Güzel Sanatlar Üniversitesi.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice-Hall.

Finkelstein, L. What is not measurable, make measurable. *Measurement and Control*, *15*(1), 25-32. https://doi.org/10.1177/002029408201500104

Finkelstein, L. (2003). Widely, strongly and weakly define measurement. *Measurement*, *34*(1), 39-48. https://doi.org/10.1016/S0263-2241(03)00018-6

Finkelstein, L., & Leaning, M. S. (1984). A review of the fundamental concepts of measurement. *Measurement*, *2*(1), 25-34.

Funk, S. C., & Dickson, K. L. (2011). Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology*, *38*(4), 273-277. https://doi.org/10.1177/0098628311421329

Geiser, S., & Santelices, M. V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes.* Research & Occasional Paper Series. https://files.eric.ed.gov/fulltext/ED502858.pdf

Gezer, T., Wang, C., Polly, A., Martin, C., Pugalee, D., & Lambert, R. (2021). The relationship between formative assessment and summative assessment in Primary grade mathematics classrooms. *International Electronic Journal of Elementary Education*, *13*(5), 673-685.

Gholami, V., & Moghaddam, M. M. (2013). The effect of weekly quizzes on students' final achievement score. *International Journal of Modern Education and Computer Science*, *5*(1), 36.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365-379. https://doi.org/10.1080/0969594970040304

Harlen, W., Gipps, C., Broadfoot, P., & Nuttall, D. (1992). Assessment and the improvement of education. *The Curriculum Journal*, *3*(3), 215-230. https://doi.org/10.1080/0958517920030302

Hattie, J., & Clinton, J. (2001). The assessment of teachers. *Teaching Education*, *12*(3), 279-300. https://doi.org/10.1080/10476210120096551

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81-112. https://doi.org/10.3102/003465430298487

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, *89*(2), 140-145. https://doi.org/10.1177/003172170708900210

Jaworski, J. M., Bek, J., & Fiok, A. J. (2014). Impact of modern instrumentation on the system of basic concepts in metrology. *Acta IMEKO*, *3*(1), 41-46. https://doi.org/10.21014/acta_imeko.v3i1.194

Kalender, I. (2014). Profiling instructional effectiveness to reveal its relationship to learning. *The Asia-Pacific Education Researcher*, *23*, 717-726.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *29*(2), 119-127. https://doi.org/10.2307/2986296

Kayri, M., & Boysan, M. (2007). Using chaid analysis in researches and an application pertaining to coping strategies. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, *40*(2), 133-149. https://doi.org/10.1501/Egifak_0000000182

Kingston, N. M., & Anderson, G. (2013). Using state assessments for predicting student success in dual-enrollment college classes. *Educational Measurement: Issues and Practice*, *32*(3), 3-10. https://doi.org/10.1111/emip.12014

Kizlik, B. (2012). Measurement, assessment, and evaluation in education. *Retrieved October*, *10*, 2015.

Kontas, H., & Özpolat, E. T. (2017). Investigating the written exam scores' prediction power of TEOG exam scores. *International Journal of Evaluation and Research in Education*, *6*(1), 31-37.

Leigh, S. R. (2012). The classroom is alive with the sound of thinking: The power of the exit slip. *International Journal of Teaching and Learning in Higher Education*, *24*(2), 189-196.

Li, X., Chan, C. W., & Nguyen, H. H. (2013). Application of the neural decision tree approach for prediction of petroleum production. *Journal of Petroleum Science and Engineering*, *104*, 11-16. https://doi.org/10.1016/j.petrol.2013.03.018

Liu, X., & Lu, K. (2008). Student performance and family socioeconomic status: Results from a survey of compulsory education in Western China. *Chinese Education & Society*, *41*(5), 70-83. https://doi.org/10.2753/CED1061-1932410505

Long, H., & Pang, W. (2016). Family socioeconomic status, parental expectations, and adolescents' academic achievements: A case of China. *Educational Research and Evaluation*, *22*(5-6), 283-304. https://doi.org/10.1080/13803611.2016.1237369

Luce, C., & Kirnan, J. P. (2016). Using indirect vs. direct measures in the summative assessment of student learning in higher education. *Journal of the Scholarship of Teaching and Learning*, *16*(4), 75-91. https://doi.org/10.14434/josotl.v16i4.19371

Mac Iver, M. A. (Ed.). (2010). *Gradual disengagement: A portrait of the 2008-09 dropouts in the Baltimore city schools*. Baltimore Education Research Consortium.

Madison-Harris, R., Muoneke, A., & Times, C. (2012). *Using formative assessment to improve student achievement in the core content areas.* Southeast Comprehensive Center.

Maimon, O. Z., & Rokach, L. (2014). *Data mining with decision trees: Theory and applications* (Vol. 81). World scientific.

Maimon, O. Z., & Rokach, L. (2014). *Data mining with decision trees: theory and applications* (Vol. 81). World Scientific.

Martin, C. S., Polly, D., Wang, C., Lambert, R. G., & Pugalee, D. K. (2016). Perspectives and practices of elementary teachers using an internet-based formative assessment tool: The case of "assessing mathematics concepts". *International Journal for Technology in Mathematics Education*, *23*(1), 3-11.

Mohr, L. B. (1995). *Impact analysis for program evaluation*. Sage.

Morgan, C. (2000). Better assessment in mathematics education? A social perspective. *Multiple Perspectives on Mathematics Teaching and Learning*, 225-242.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, *14*(2), 149-170. https://doi.org/10.1080/09695940701478321

Noonan, B., & Duncan, C. R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research, and Evaluation*, *10*(1), 17. https://doi.org/10.7275/a166-vm41

Olani, A. (2009). Predicting first year university students' academic success. *Electronic Journal of Research in Educational Psychology*, *7*(3), 1053-1072.

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, *52*(2), 464-473. https://doi.org/10.1016/j.dss.2011.10.007

Ozan, C., & Kincal, R. (2018). The effects of formative assessment on academic achievement, attitudes toward the lesson, and self-regulation skills. *Educational Sciences-Theory & Practice, 18*.

Pajares, F., & Miller, M. D. (1997). Mathematics self-efficacy and mathematical problem solving: Implications of using different forms of assessment. *The Journal of Experimental Education*, *65*(3), 213-228. https://doi.org/10.1080/00220973.1997.9943455

Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, *86*(4), 554-565. https://doi.org/10.1016/S0034-4257(03)00132-9

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

Parker, J. M., Anderson, C. W., Heidemann, M., Merrill, J., Merritt, B., Richmond, G., & Urban-Lurain, M. (2012). Exploring undergraduates' understanding of photosynthesis using diagnostic question clusters. *CBE—Life Sciences Education*, *11*(1), 47-57. https://doi.org/10.1187/cbe.11-07-0054

Patka, M., Wallin-Ruschman, J., Wallace, T., & Robbins, C. (2016). Exit cards: Creating a dialogue for continuous evaluation. *Teaching in Higher Education*, *21*(6), 659-668. https://doi.org/10.1080/13562517.2016.1167033

Price, B. A., & Randall, C. H. (2008). Assessing learning outcomes in quantitative courses: Using embedded questions for direct assessment. *Journal of Education for Business*, *83*(5), 288-294. https://doi.org/10.3200/JOEB.83.5.288-294

Radmehr, F., & Vos, P. (2020). Issues and challenges for 21st century assessment in mathematics education. In L. Leite, E. Oldham, A. S. Afonso, F. Viseu, L. Dourado, H. Martinho (Eds.), *Science and mathematics education for 21st century citizens: Challenges and ways forwards* (pp. 437-462). Nova Science.

Rea-Dickins, P., & Germaine, K. (1992). *Evaluation*. Oxford University Press.

Reynolds, M. (2022). A relational approach to learning environments: Learning from Pacific wisdom. *New Zealand Journal of Educational Studies*, 57(1), 143-158. https://doi.org/10.1007/s40841-021-00236-x

Richardson, R. (1992). The multiple choice true/false question: What does it measure and what could it measure? *Medical Teacher*, *14*(2-3), 201-204. https://doi.org/10.3109/01421599209079488

Romberg, T. A. (Ed.). (1992). *Mathematics assessment and evaluation: Imperatives for mathematics educators*. SUNY Press.

Jitendra, A. K., & Dupuis, D. N. (2015). The role of tier I mathematics instruction in elementary and middle schools: Promoting mathematics success. In S. Jimerson, M. Burns, A. Van Der Heyden (Eds) *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (pp. 215-233). Springer.

Rosário, P., Núñez, J. C., Vallejo, G., Cunha, J., Nunes, T., Mourão, R., & Pinto, R. (2015). Does homework design matter? The role of homework's purpose in student mathematics achievement. *Contemporary Educational Psychology*, *43*, 10-24. https://doi.org/10.1016/j.cedpsych.2015.08.001

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-144.

Shirvani, H. (2009). Examining an assessment strategy on high school mathematics achievement: Daily quizzes vs. weekly tests. *American Secondary Education*, 34-45.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research*, *95*(6), 323-332. https://doi.org/10.1080/00220670209596607

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., & Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*. Springer Nature.

Swank, A. L. G. (1999). *The effect of weekly math homework on fourth grade student math performance* [Unpublished master thesis]. Johnson University. https://files.eric.ed.gov/fulltext/ED433234.pdf

Terman, L. M. (1921). Intelligence and its measurement: A symposium--II. *Journal of Educational Psychology*, *12*(3), 127. https://doi.org/10.1037/h0064940

Toppino, T. C., & Ann Brochin, H. (1989). Learning from tests: The case of true-false examinations. *The Journal of Educational Research*, *83*(2), 119-124. https://doi.org/10.1080/00220671.1989.10885940

Toppino, T. C., & Ann Brochin, H. (1989). Learning from tests: The case of true-false examinations. *The Journal of Educational Research*, *83*(2), 119-124. https://doi.org/10.1080/00220671.1989.10885940

Trautwein, U. (2007). The homework–achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction*, *17*(3), 372-388. https://doi.org/10.1016/j.learninstruc.2007.02.009

Veldhuis, M., & Van den Heuvel-Panhuizen, M. (2020). Supporting primary school teachers’ classroom assessment in mathematics education: Effects on student achievement. *Mathematics Education Research Journal*, *32*(3), 449-471.

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment*. Psychology Press.

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*(3), 461-481. https://doi.org/10.1037/0033-2909.91.3.461

Wu, D. (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications*, *36*(5), 9105-9112. https://doi.org/10.1016/j.eswa.2008.12.039

Wylie, E. C. (2020). Observing formative assessment practice: Learning lessons through validation. *Educational Assessment*, *25*(4), 251-258. https://doi.org/10.1080/10627197.2020.1766955

Zarei, A. A. (2008). The effect of quiz frequency on course performance and its relationship with class attendance. *Journal of English Language Pedagogy and Practice*, *1*(Inaugural Issue), 162-180.

Zhang, N., & Henderson, C. N. (2015). Can formative quizzes predict or improve summative exam performance? *Journal of Chiropractic Education*, *29*(1), 16-21. https://doi.org/10.7899/JCE-14-12