AN ANALYSIS OF 4TH GRADE STUDENTS' ROUTINE AND NON-ROUTINE PROBLEM-SOLVING SKILLS USING COGNITIVE DIAGNOSTIC MODELS

by

ZÜLEYHA TAŞTAN

B.S., Undergrad Education in Mathematics Education, Boğaziçi University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfilment of
the requirements for the degree of
Master of Science

Graduate Program in Mathematics and Science Education
BOĞAZİÇİ UNIVERSITY
2022

To my lovely family…

and To my best friend, Sezgin…

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Assoc. Prof. Dr. Serkan Arikan for accepting me as a master student and giving me opportunity to study this very interesting project independently. Working with a consultant who always focuses on progress and solution has been one of the greatest lucks in my life. I am grateful to him for helping me improve my research and academic skills, supporting me to attend conferences, and providing me effective feedbacks at every step of my research.

In addition to my advisor, I would like to thank the other members of my thesis committee, Assoc. Prof. Dr. Engin Ader and Assoc. Prof. Dr. Sedat Şen for their acceptance of being a jury member for my thesis, for their support, encouragement, detailed feedbacks, and constructive comments that helped to improve and finish my thesis. It has been a pleasure working with them.

I would like to thank my dear family for their support, faith, and patience. I am grateful to my father, İsa Taştan, for his love for education and his support for me against everything and everyone. I would like to thank my dear mother, Nebahat Taştan, for her prayers and warm love. I would like to thank my sisters Hatice Taştan and Şevval Taştan for being in my life and for their support. Despite all the difficulties, every success I have achieved today is your work. I am proud of you, and I love you very much.

Also, I would like to thank my dear friends İsmail Elçiçek and Aycan Adıgüzel, who supported me and always encouraged me throughout the thesis process, for being a light and hope for me. I am so lucky to have you in my life.

I would also like to thank Semra Güven, the best friend that the master process brought me, for supporting me throughout the process and being there for me when I needed help. I would like to thank Banu Karyağdı for being a part of my undergraduate and master education and always making me laugh even in the most stressful times. I would like to thank my dear friend Hatice Ünaldı and Mihriban Altun Sert for their help and efforts in the data collection process. I am most sincerely grateful for your great help.

I would like to thank my dear uncle Burhanettin Taştan and his wife Belgin Yazıcıoğlu Taştan for their support throughout my education. Your support means a lot to me. I would also like to thank my dear cousins and motivation sources; Ekin Kızıldağ (EK), Eylül Kızıldağ Çobanoğlu and Betül Taştan Çakır. I would also like to thank all my close friends, İpek Kadı Yıldırım, Melike Okur, Esra Akhüseyinoğlu Güler, Gülnihal Çetin, Çiğdem Özkan, Betül Ürkmez Fışkın, Tuğba Kübra Ladikli, Hakim Fışkın, Miraç Türken, Abdullah Yıldız, who have touched a part of my life at any time and shared my love.

I am also grateful to Boğaziçi University Adaptive Testing Laboratory (BAT Lab) and the entire team for their contribution to test development and data entry processes. In addition, I would like to thank a lot all my teachers and all my professors at the university for their efforts so far.

Last, but never least, I want to thank the best part of my life, Sezgin Mengi. Thank you for being there for me in my toughest times, for being patient with my worst moods, and for your faith and love in me all the time. I don't know what I did to deserve you but thank you for everything. My best friend, I'm glad you're in my life.

ABSTRACT

AN ANALYSIS OF 4^{TH} GRADE STUDENTS' ROUTINE AND NON-ROUTINE PROBLEM-SOLVING SKILLS USING COGNITIVE DIAGNOSTIC MODELS

Assessment is an important part of education that provides feedbacks to both students on their learning and educators on their instruction. Assessments enhance students' learning when providing effective and immediate feedback to learners. In this study, Cognitive Diagnostic Assessments (CDAs) are used to give informative and in-depth feedback to teachers and students. CDAs provide cognitive data about students' strengths and weaknesses in a particular ability. This study investigates routine and non-routine problem-solving attributes of grade 4 students with four operations (addition, subtraction, multiplication, and division) using CDA. The test used in the study is specifically developed with diagnostic purposes regarding 2021-2022 4^{th} grade mathematics curriculum and analysed with Cognitive Diagnostic Models (CDMs). There are 6 attributes and 20 items in the test. A Q-matrix was constructed to show attribute-item relationship. Before administering the test, a group with 10 students were asked to answer the items and interviewed later to evaluate the timing and clarity of the items. The final form of CDA test was administered to 511 students in 4^{th} grade, from various public and private schools. The responses were analysed by the GDINA package in R. As a result of the analysis, each participant was assigned to an attribute profile showing which skills they have mastered and which they have not. The results showed that 75% of the students have mastered routine problem-solving skills, while 17% of the students have mastered non-routine problem-solving skills. The most frequent attribute profiles in the study group are found as follows: 000010, 010010, 111111, 111110, 000000 and 110010.

ÖZET

4. SINIF ÖĞRENCİLERİNİN RUTİN VE RUTİN OLMAYAN PROBLEM ÇÖZME BECERİLERİNİN BİLİŞSEL TANILAMA MODELLERİ KULLANILARAK ANALİZLENMESİ

Ölçme-değerlendirme, hem öğrencilere kendi öğrenmeleri hakkında hem de eğitimcilere öğretimleri hakkında geri bildirim sağlayan eğitimin önemli bir parçasıdır. Değerlendirmeler, öğrencilere etkili ve anında geri bildirim sağlarken öğrencilerin öğrenmesini geliştirir. Bu çalışmada, öğretmenlere ve öğrencilere bilgilendirici ve tanılayıcı geribildirim vermek için Bilişsel Tanılama Değerlendirmeler (BTD) kullanılmıştır. BTD'ler, öğrencilerin belirli bir yetenekteki güçlü ve zayıf yönleri hakkında bilişsel veriler sağlar. Bu çalışma, BTD kullanarak 4. sınıf öğrencilerinin dört işlemde rutin ve rutin olmayan problem çözme özelliklerini araştırmaktadır. Araştırmada kullanılan test, 2021-2022 4. sınıf matematik müfredatına yönelik tanısal değerlendirme sağlamak amaçlı özel olarak geliştirilmiş ve Bilişsel Tanılama Modelleri (BTM) ile analiz edilmiştir. Testte 6 özellik ve 20 madde bulunmaktadır. Çalışmadaki nitelik-madde ilişkisini göstermek için bir Q-matrisi oluşturuldu. Testi uygulamadan önce 10 kişilik gönüllü bir öğrenci grubundan maddeleri cevaplamaları istendi ve daha sonra testin süresi ve maddelerin anlaşılırlığını değerlendirmek için görüşmeler yapıldı. BTD testinin son hali, çeşitli devlet ve özel okullardan 4. sınıflarda öğrenim gören 511 öğrenciye uygulanmıştır.Yanıtlar, R'daki GDINA paketi tarafından analiz edilmiştir. Analizin sonucunda, her katılımcıya hangi becerilerde ustalaşıp hangilerinde ustalaşmadıklarını gösteren bir öznitelik profili atanmıştır. Sonuçlar, öğrencilerin %75'inin rutin problem çözme becerilerinde uzmanlaştığını, %17'sinin ise rutin olmayan problem çözme becerilerinde uzmanlaştığını göstermiştir. Çalışma grubunda en sık görülen öznitelik profilleri ise şu şekildedir: 000010, 010010, 111111, 111110, 000000 ve 110010

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

α_k	Attribute k
g	Guessing parameter
j	Total number of items
$[J \times K]$	Dimensional matrix with the 0 or 1 inputs where J represents the number of items and K represents the number of attributes
$P\ (X_{i\,j})$	The probability of the student who has all the required attributes to answer the item correctly
$P\ (Y_{i\,j})$	Probability of observed score
r_{ik}^{*}	The inverse diagnostic capacity of an attribute for a corresponding item
s	Slipping parameter
$\propto_{lj}^{*}$	Attribute vector
δ_{j0}	the intercept for item j
δ_{jk}	the main effect due to α_k
$\delta_{jkk\prime}$	the interaction effect due α_k and $\alpha_{k\prime}$;
$\delta_{jk\ldots\ldots.K_j^*}$	the interaction effect due to α_1 ……… $a_{k_j^*}$.
$\sum$	Summation Notation
$\eta_{i\,j}$	Total number of items
π_i^*	The probability of solving an item
$\prod$	Product Notation

LIST OF ABBREVIATIONS

ACDM	Additive CDM
AIC	Akaike Information Criterion
BAT	Boğaziçi University Adaptive Testing Lab
BIC	Bayesian Information Criterion
BTD	Bilişsel Tanılama Değerlendirmeleri
BTM	Bilişsel Tanılama Modelleri
CDA	Cognitive Diagnostic Assessments
CDM	Cognitive Diagnostic Models
CDS	Cognitive Design System
CI	Confidence Interval
C-RUM	Compensatory Reparameterized Unified Model
CTT	Classical Test Theory
DCM	Diagnostic Classification Models
DINA	Deterministic-input, noisy-and-gate model
DINO	Deterministic Inputs, Noisy “or” Gate
EAP	Estimates of Attribute Profiles
GDINA	Generalized DINA

GDM	General Diagnostic Model
IDI	Item Discrimination Index
IRT	Item-Response Theory
LCDM	Linear Cognitive Diagnostic Model
LSDM	Least Squares Distance Method
MDS	Multidimensional Scaling
MG-DINA	Multi-Group DINA
NC-RUM	Non-Compensatory Reparametrized Unified Model
NCTM	National Council of Teachers of Mathematics
PISA	Program for Intetrnational Student Assessment
RMSEA	Root Means Square Error of Approximation
R-RUM	Reduced Reparameterized Unified Model
RSM	Rule Space Model
SABIC	Sample-Size Adjusted BIC
SRMSR	Standardized Root Mean Squared Residual
TIMSS	Trends in International Mathematics and Science Study

1. INTRODUCTION

Wiggins (1998) states that main goal of the assessment should be enhancing student performance instead of just scoring or checking their knowledge. Also, Stiggins (2002) asserts that assessments should be used to both determine the state of learning and to advance learning. On the other hand, as accountability became more and more popular, most of the resources were directed into assessments that just audit students' learning rather than offering data that may support instruction and learning (de la Torre, 2009a). Assessments in education are mostly developed to evaluate various domains based on contents, skills, or proficiency (Haberman, Sinharay, & Puhan, 2009; Sinharay, Haberman, & Puhan, 2007) and scores for individuals are assigned to these domains. Although there are many types of assessments to cover a wide domain, assessments in education are dominated by one-dimensional summative tests, in which the total scores of the students show their achievement levels (de Ayala, 2009; Hambleton, Swaminathan, & Rodgers, 1991; Yao & Boughton, 2007). Regarding these assessment tests, each test taker is ranked according to total item score, or a single continuous latent variable and even comprehensive abilities are analysed and reported based on one total score (Wang, 2009).

Even though old assessment models are convenient and beneficial in ranking, comparing, and even predicting learners' future performance, they contain a limited amount of diagnostic information on the strengths and weaknesses of the learners (Choi, 2010; de la Torre & Karelitz, 2009). Knowing their strong and weak points is critical for students who are not satisfied with their performance and want to improve in those areas (Ardıç, 2020). In the "Knowing What Students Know" report of National Research Council (NRC) (2001), emphasis on the need of providing formative and diagnostic assessment to the students was indicated. Similarly, in the project of No Child Left Behind (2001), the importance of giving diagnostic feedbacks is stated to help teacher to address the needs of the students. In conclusion, summative tests that gives one total score to evaluate learners' abilities or knowledge in the domain are found inadequate to give in-depth feedbacks in both individual and classroom level. Therefore, diagnostic models and assessments have become more known and valuable to meet the demands and benefit from large-scale assessments more (Liu, Huggins-Manley, & Bulut, 2017; Rupp, Templin, & Henson, 2010; Sinharay, Puhan, & Haberman, 2010). As a result, assessments should be more explicit about the attributes or abilities that individuals have mastered or have not mastered. It is also crucial that assessments should be formative and diagnostic to direct students on what they need to study and what abilities they should improve to succeed.

Assessment in education has progressed beyond rating one's level of achievement to being diagnostically valuable at all levels of education (Bolt, 2007). Nowadays, there is an increasing demand for more formative information on students' academic performances in assessment tests in the field of education. The existing models are criticized for their incapacity to provide in-depth information on students' strengths and weaknesses in a particular subject. Most of the educators agree on that it would be beneficial to receive diagnostic and detailed information at both the individual and classroom levels beyond what is often provided by large-scale tests. (Huff & Goodman, 2007). These objections revolve on the need for methods to give individuals more comprehensive diagnostic information and relate this data to instructional needs and educational requirements. New assessment and evaluation models have been developed and discussed to provide feedback for learners, teachers, parents, and administrations. Cognitive Diagnostic Assessments (CDAs) are one of these approaches. CDAs are designs that provide data to assess knowledge structure and abilities of the students and to identify cognitive strengths and weaknesses of the students (Gierl, Cui, & Zhou, 2009). CDAs require statistical models called Cognitive Diagnostic Models (CDMs) to extract in-depth and diagnostic data. These models also called as Diagnostic Classification Models (DCMs) (DiBello, Roussos, & Stout, 2007; Rupp, Templin, & Henson, 2010).

Cognitive diagnostic models have been developed to offer diagnostic data on test results by classifying test takers according to their mastery of a particular subject or proficiency in a certain ability (DiBello, Roussos, & Stout, 2007). Cognitive Diagnostic Models (CDMs) or Diagnostic Classification Models (DCMs) are confirmatory latent models (Ma, Iaconangelo, & de la Torre, 2016). These models are useful to describe the relationship between the variables and categorical latent models. In CDMs, the variables are observable responses in the assessment, which are typically referred to as attributes

(Cheng, 2010). The most important and characteristic side of CDM is that it offers in-depth and multi-dimensional representation of individual profiles. Ardıç (2020) also indicated that one of the main reasons of using CDMs in education is to identify the attributes and the skills of the test takers and determine their weaknesses and strengths in the domain. Thus, to evaluate students' strengths and weaknesses on routine and non-routine problem-solving abilities a CDA test is specifically designed and analysed by using CDMs.

1.1. Significance of the Study

The current study provides information on the weaknesses and strengths of 4th graders in the domain of problem solving with arithmetical operations as addition, subtraction, multiplication, and division. The study aims to develop an instrument with CDA approach and exemplify an alternative way for assessment to provide effective feedback. The findings of this study are expected to contribute to the field for many reasons.

First, the items of the study are specifically prepared with diagnostic purposes based on defined attributes. Generally, in the field of education most of the CDA studies are administered using items of large-scale assessments like Programme for International Student Assessment (PISA) and The Trends in International Mathematics and Science Study (TIMSS) (Li et al., 2020; Sen & Arican, 2015; Toker & Green, 2012) instead of designing a test with diagnostic purposes. They define the attributes considering the items that are prepared for large scale assessment purposes by retrofitting. Sen & Arican (2015) pointed out although their research showed that retrofitting a CDM through the DINA model may be highly effective for TIMSS evaluation, it is clear that CDM-based analyses can provide more advantage when tests are created using CDMs in advance. Therefore, current study is expected to contribute to the field by providing a pure cognitive approach.

Secondly, the current study shows the mastery levels of the participants in the cognitive attributes depending on their responses. The study focuses on giving diagnostic feedbacks on problem-solving skills with four operations. According to Van de Walle, Karp, and Bay-Williams (2016), every student should get effective feedback on their problem-solving skills and progress in grasping mathematical procedures. However, most research on problem-solving abilities focus on problem posing (Kılıç, 2013; Tertemiz, 2017) or misconceptions in operational skills (Brandt, Bassoi, & Baccon, 2016; Passolunghi & Pazzaglia 2005; Stoyanova, 2003). Although there are some CDM studies focus on mathematics achievement (Im & Park, 2010; Lee et al., 2013; Sen & Arican, 2015; Toker & Green, 2012), the number of research on developing a CDA test in the domain of mathematical problem-solving is limited (Li, Zhou, Huang,Tu, Gao, Yang, & Li, 2020). There is gap in the field on developing a CDA test in the domain of routine and non-routine problem-solving skills with four operations. Therefore, being aware of what abilities students have and how they differentiate in the name of mastering each attribute can lead the way for the teachers and students.

Additionally, the study highlights the ability differences of the students who have the same total score. Differentiating students' profile is especially important on the domain of problem solving. Assume that there are two different students who get 70 from the test. They have the same number of correct responses from the test in which marking of each question is weighted equally. However, cognitive diagnostic assessment provides another view. Since all items are designed regarding cognitive attributes, every student has a cognitive profile according to their mastered and non-mastered skills. Therefore, even though they get the same score, they may belong totally different profiles. One student might have mastered the order of operations but the other one might have mastered in adding and subtracting integers. Hence, this study suggests providing individual feedbacks for each student according to their strengths and weaknesses.

1.2. The Purpose of the Study

The purpose of the study is to investigate the weaknesses and strengths of 4^{th} grade students on the routine and non-routine problem-solving abilities with four operations using CDMs. As parallel to the aim, a CDA test is specifically designed based on *fine-grained* attributes to provide diagnostic and informative feedback to the students.

1.3. Research Question

The research question of the study:

1. What are the cognitive strengths and weaknesses of 4th grade students for the attributes of routine and non-routine problem-solving skills with four operations (addition, subtraction, multiplication, and division)?

1.4. Problem and Problem Solving

A problem can be defined as a situation that individuals or groups are faced with and requires determination to solve in the case that solution path is not clear or obvious to follow (Polya, 1962). Definition of problem from the view of mathematics can be stated as a circumstance in which a strategy must be found or demonstrated, but the solution is not immediately apparent to the solver (Kayan & Çakıroğlu, 2008).

Furthermore, the problem-solving in mathematics consists of the cognitive processes that require a specific and directed goal to reach, when the solution way is unknown (Van de Walle, 2014). The process of problem-solving is complicated and requiring the coordinated application of a variety of talents (Tertemiz, 2017). It requires some level of creative and critical thinking. The problems that are discussed in mathematics classrooms needs to have some aspects like including both factual and procedural concepts, being engaging and authentic, being appropriate for the level of the students and for the context.

During instruction, problems are important questions that demands students to use both prerequisite knowledge and mathematical thinking. That is, they require some intellectual and cognitive challenges to improve students' mathematical comprehension (National Council of Teachers of Mathematics, 2000). In these kinds of questions, there could be one, none or many different solutions paths to obtain answer. However, problem solving does not aim just finding a correct answer. It also aims to introduce unfamiliar and new situations to the learners, encourage them to find different and flexible solution strategies, and improve useful and aesthetic solutions for the problems (Gail, 1996). Therefore, problem solving is much more than following well-memorized routine steps or recalling facts for every question (Lester,1994), it requires not only mathematical knowledge and computational skills but also creativity, reasoning, and curiosity.

The steps in problem-solving procedure can be ordered as understanding the problem, selecting the relevant data from the options, converting the data to mathematical symbols, and arriving at the solution after completing the required operations. There is no straight path connecting these parts. (Olkun & Toluk, 2004). The first stage in problem solving is understanding what was read, and if this step is not completed, it is assumed that the person would fail selecting the correct answers or numbers from the problem.

After reviewing the literature, two types of problems may be found: routine problems and non-routine problems (Tertemiz, 2017). The problem types are explained below.

1.4.1. Routine and Non-routine Problems

As the perspective toward problem-solving changes in time, the features and types of the problems are also affected by that change. It resulted in the emergence of several problem types and strategies over time. With a shift in perspective, problems are classified as either routine or non-routine. While regular problems are computational problems with well-known techniques (Altun, 2001), non-routine problems need mathematical reasoning and criticism (Mullis & Martin, 2017) that extends beyond routine problems.

A routine problem is one that a learner would encounter in textbooks or in regular classes and that requires the use of a typical, well-known technique (Arikan, Erktin, & Pesen, 2020). A formula, an equation, or a well-known process can be used to address routine problems (Polya, 1957). Routine problems rely mostly on determining which computations will be used and then performing arithmetic operations. Therefore, it can be said that routine problems tend to require a low level of critical and very limited creative thinking. Altun (2001) defined routine problems as practical and requiring at least one of the four arithmetic operations or ratios. In light of the definitions, routine problems might be considered weak problems in the context of mathematical reasoning. Math educators, on the other hand (Xin, Lin, Zhang, & Yan, 2007), believe that routine problems are equally as important as non-routine problems. Polya (1966) pointed out, however, that routine problems can be basically a need and beneficial at the right time, but only when used appropriately.

Compared to routine problems, non-routine problems are generally focus more on mathematical thinking, originality, and empowering students in the name of mathematical understanding. A non-routine problem is one that necessitates the use of one's thinking abilities and has a novel context to which a learner cannot find a solution by following a standard approach (Arikan et al., 2020). When the literature is evaluated, it is discovered that non-routine problems often contribute the problem-solving ability, and non-routine problem-solving skills improve the ability to use them in real life scenarios (Polya, 1957; London, 2007; Schoenfeld, 2016). London (1993) stated that as students solve more non-routine problems, they also experience and gain additional problem-solving skills like data organization, interpretation of given data, designing algorithms, or transformation of complex equations to another easier form. London (1993) also indicated that the experience with non-routine problems encourage students to act like a mathematician, while they are solving problems because non-routine problems reinforce them to think mathematically, be creative, and think critically.

1.4.2. Importance of Problem Solving

Problem solving is crucial in mathematics and mathematics education. It is not simply a mathematical ability, but also a survival skill that increases with time and experience. In recent decades, mathematical problem solving has received more attention, and it has come to be seen as an integrated aspect of mathematics learning rather than an isolated component of mathematics curriculum (National Council of Teachers of Mathematics, 2000). The reason of this importance is that problem-solving skills are not needed just for daily life problems and success but also for the improvement of the society (Brown, 2003). NCTM (2000, p.54) indicated:

> “Students can learn about, and deepen their understanding of, mathematical concepts by working through carefully selected problems that allow applications of

> mathematics to their contexts, and these well-chosen problems can be particularly valuable in developing or deepening students' understanding of important mathematical ideas".

Since the problems lead students to think deeper about the concepts in mathematics, they can change or support their mathematical knowledge and concepts and they can be more open to different ways of learning mathematics in that way (Steele & Widman, 1997). Therefore, it can be said that problem solving provide students a new and dynamic perspective toward mathematics and they can organize their ideas, construct concepts and be more engage in discussions (Santos-Trigo, 1998).

1.4.3. Assessment of Problem Solving

Problem solving is an excellent way to observe students' mathematical concepts and demonstrate how they link or relate to one another (Brown, 2003). Furthermore, problem solving is not limited to a mathematical framework. The importance of problem-solving stems from its many aspects and large impact on problem-solving skills to create a better future and contribute to the improvement of society (Brown, 2003). As a result, problem-solving instruction have become more popular in the twenty-first century, and many countries began to place more emphasis on it.

However, these changes must reach students to achieve the purpose of the educational improvements in the domain of problem-solving. Students learn from their experiences, which are largely provided by the teacher in the classroom setting. These encounters have the potential to alter students' mathematical understandings and problem-solving abilities (Arabeyyat, 2004). Additionally, the creation of curriculum materials, assessment tools, manipulatives, activities, and difficulties by instructors demonstrates the importance of teachers in bringing about educational reforms (Mason, 2003). Therefore, the perspective of the teachers to the problem-solving, the instruction in the classroom, the materials and assessment style have an impact on students' problem-solving skills. According to some research (Nathan & Koedinger, 2000; Schimidt & Brosnan, 1996), there are some important points to consider while teaching and assessing problem-solving. For example, the feedback and process are valuable than answers, excellent computational skills are not required to be able to solve problems and spending some time in a problem should not be seen as a waste of time (Nathan & Koedinger, 2000; Schimidt & Brosnan, 1996). However, still there are some educators who assess the problem-solving skills by the number of the correct answers instead of the process (Arabeyyat, 2004). Since the current study focuses on assessing problem-solving skills and providing feedbacks, importance of the assessment and types of the assessments will be shortly explained in following parts.

1.5. A Historical View for the Definition of Assessment

Assessment is an inseparable part of education that is directly related to both learning and teaching. Assessment can be basically defined as the process of collecting, analyzing, recording, and using data of learners' answers on an educational test or task (Harlen et al., 1992). The general idea of assessment is to collect information that reflects students' content knowledge and cognitive skills.

Interpretations can be made based on the collected data regarding where the learner is in his or her learning process, what knowledge level is aimed to achieve, and how best to achieve the goal (Black & William, 2009; Brown, 2004). Wiggins (1998) indicated the aim of assessment as not only checking students' content knowledge but also teaching and achieving further learning. Also, Stiggins (2002) pointed out that assessment needs to be used also to improve learning but detecting students' achievement level.

Considering the domain of the assessment definitions, a wide range of assessment tools and tests are developed for different purposes. Even though assessment has covered teaching and learning, it is mostly used in one direction. Assessment is a way to measure learning and identify students' content knowledge and cognitive abilities. But it is commonly used for summative purposes like ranking, scoring, and certificating. Assessment is a two-way ticket that can be beneficial also for teaching. It can be used to evaluate the effectiveness of the programs or instructions (Harlen et al., 1992). de La Torre (2009) highlighted that especially with the increment on the importance of accountability, the expectation for assessment has been changed. Assessments are expected to not only measure conceptual knowledge but also provide feedback for students, teachers, and instruction. Also, in 2001 NRC reported that the need for more research and funding for assessment to facilitate learning. In that way, it is aimed to provide more, informative, and diagnostic data to support learning and teaching by using assessment. In this part, it is presented various definitions for assessment and emphasized the need for cognitive assessments. Now, two the most common types of assessment will be explained, and cognitive diagnostic assessment will be summarized.

1.6. Summative and Formative Assessment

Assessment is important in learning and education since it is closely tied to both the learning and teaching processes. Although there are several types of assessment, such as formative assessment, summative assessment, and cognitive diagnostic assessment, determining which type of assessment will be employed depends on the educational purposes (Shute & Rahimi, 2017). The most common two assessment types in the education are summative assessment which is assessment of learning and formative assessment which is assessment for learning.

Summative assessment is utilized for grading and accountability since it involves the using information for certification, grades, or GPAs. It is simpler to use cumulative groupings, yet it requires solid proof for each student. Summative evaluation allows for the comparison of student performances, trustworthy data, accountability, and is employed by authorities in common and entrance exams (Pellegrino, Chudowsky, & Glaser, 2001). However, it is also possible to argue that its contribution to learning is restricted because it is done after learning with the primary goal of grading (Shute & Rahimi, 2017).

Formative assessment, on the other hand, is used to promote and improve learning by offering feedback. The needs of the learner guide instruction in formative assessment (Shute & Zapata-Rivera, 2010). Formative assessment is offered based on the learner's knowledge level and learning outcomes. Formative assessment informs educators about how well students are learning. As a result, it provides a chance for the learner to improve his or her learning by offering feedback (Reigeluth & Karnopp, 2013). Therefore, it is used to support both teaching and learning. In other words, formative assessment is a more learner-centered methodology than summative evaluation (Pellegrino, Chudowsky, & Glaser, 2001).

To sum up, different assessment types can be preferred based on the aim of the assessment. Each assessment provides information about students' learning in different levels. Summative assessment and formative assessment are well-known assessment types in the field of education. Summative assessment is assessment of learning and applicable for large-scale groups at the end of their learning process. On the other hand, formative assessment is assessment for learning which focus on providing feedbacks and continue during the learning (Black & William, 2009). On the other hand, the current study focuses on the Cognitive Diagnostic Assessments (CDAs) which can be used to obtain qualitative data from quantitative tests (Leighton & Gierl, 2007).

1.7. Cognitive Diagnostic Assessments (CDAs)

Cognitive Diagnostic Models (CDMs) are statistical models that are needed for Cognitive Diagnostic Assessments (CDAs) to obtain diagnostic information from the student responses (de La Torre & Minchen, 2014). CDAs are designed to measure learners' knowledge structures, attributes, and processing skills in order to provide cognitive feedbacks in terms of weaknesses and strengths (Leighton & Gierl, 2007). CDAs are directly related to representation of students from different achievement levels with various processing skills. It is also related to improvement of the knowledge levels and adjustments-based profiles of students over time through interventions that is based on the results of the assessments. In that sense, it can be said that most of the need for CDAs comes from this potential to identify the learning level, diagnose the learning needs, and manipulate teaching to improve students' learning. Thus, it brings us to the integration of three basic elements which are curriculum, instruction, and assessment (Leighton & Gierl, 2007).

CDAs provide in-depth and diagnostic information. With the help of this diagnostic information, students will not only learn where they are in their learning process, they will also learn what they need to study to improve. CDAs are also useful for the teachers since they offer an opportunity to review their instruction and curriculum depending on the results. CDAs provide informative data for the curriculum since it shows which attributes or skills have been achieved (Yamaguchi & Okada, 2018). In that way, teachers can identify the attributes that need to be supported for the classroom and also individuals (Yamaguchi & Okada, 2018). Therefore, CDAs are beneficial to provide effective and individual feedbacks. That is why, CDAs are one of the potential future assessment models in the field of education.

1.8.The Need for Cognitive Assessments

In order to help students to improve their academic skills and knowledge, assessing current content knowledge is an important step. It is very useful to inform students, teachers, and parents about the learning process. In that way, assessment will facilitate to follow the academic process of the learners and give feedback to improve their cognitive skills. One of the most common educational assessments that is used in schools to measure students' content knowledge is administering an achievement test (de La Torre, 2009). Achievement tests are necessary to measure what students already know (Yamaguchi & Okada, 2018). Many psychometric models like Classical Test Theory (CTT) and Item Response Theory (IRT) have been designed to evaluate the data based on students' learning processes from these achievement tests. However, since these models have not been developed to extract diagnostic information from educational assessment tests, they have some limitations like unidimensionality which restricts to show students' latent ability (Yamaguchi & Okada, 2018). Therefore, achievement tests and other psychometric models might not be appropriate to represent various attributes, which is needed for educational diagnosis.

While providing feedback for further learning, having detailed and cognitive information on learner profile is essential. However, according to NRC (2003) assessments are mostly used as a part of school system to measure what is taught, instead of providing diagnostic and detailed information. Assessing learners' content knowledge and knowledge status is one of the basic aims of the assessment. But, providing effective feedback for teachers, parents, following up students' learning and evaluating curriculum effectiveness are also crucial roles of assessment (Yamaguchi & Okada, 2018). Based on school assessments, the need for cognitive assessments has been an issue in education for years. In order to extract diagnostic and detailed information from educational assessments, Cognitive Diagnostic Models (CDMs) are shown to be effective (Gierl, Cui, & Zhou, 2009).

1.9. Psychometric Models for Educational Assessment

There are numerous psychometric models for educational assessment that serve different purposes considering the aim of the assessment. The purpose of the assessment has an important role to determine an appropriate model that fits in. Ranking, finding out the conceptual knowledge, detecting misconceptions, certification or identifying weaknesses and strengths are some examples for the purpose of the assessment (Brown, 2004). For example, while ranking students in a single administration, it may be very convenient to have one common score to see the order of the students. In that case, Classical Test Theory (CTT) might be used. CTT models are used to estimate true scores of the students. CTT models are quite simple and easy to apply. But for large-scale assessment, when the assumptions of Item-Response Theory (IRT) could be satisfied, it is more effective to use IRT models. IRT models are very helpful to estimate latent ability of students independent from the test (Deng & Hambleton, 2013). On the other hand, IRT models are not efficient for all purposes. They do not provide detailed cognitive feedback for participants about their weaknesses and strengths. CDMs can be beneficial to answer this problem. They are designed to identify students' weaknesses and strengths while creating various learner profiles using attained attributes. Therefore, all psychometric models have their own advantages and provide better results for various test purposes, and they correspond to a need in the field. Hence, deciding which model is the most beneficial and appropriate depends on purpose of the test.

Assessment is important in learning and education since it is closely tied to both the learning and teaching processes. Although there are several types of assessment, such as formative assessment, summative assessment, and cognitive diagnostic assessment, determining which type of assessment will be employed depends on the educational purposes (Shute & Rahimi, 2017). The most common two assessment types in the education are summative assessment which is assessment of learning and formative assessment which is assessment for learning.

Summative assessment is utilized for grading and accountability since it involves the using information for certification, grades, or GPAs. It is simpler to use cumulative groupings, yet it requires solid proof for each student. Summative evaluation allows for the comparison of student performances, trustworthy data, accountability, and is employed by authorities in common and entrance exams (Pellegrino, Chudowsky, & Glaser, 2001). However, it is also possible to argue that its contribution to learning is restricted because it is done after learning with the primary goal of grading (Shute & Rahimi, 2017).

Formative assessment, on the other hand, is used to promote and improve learning by offering feedback. The needs of the learner guide instruction in formative assessment (Shute & Zapata-Rivera, 2010). Formative assessment is offered based on the learner's knowledge level and learning outcomes. Formative assessment informs educators about how well students are learning. As a result, it provides a chance for the learner to improve his or her learning by offering feedback (Reigeluth & Karnopp, 2013). Therefore, it is used to support both teaching and learning. In other words, formative assessment is a more learner-centred methodology than summative evaluation (Pellegrino, Chudowsky, & Glaser, 2001).

To sum up, different assessment types can be preferred based on the aim of the assessment. Each assessment provides information about students' learning in different levels. Summative assessment and formative assessment are well-known assessment types in the field of education. Summative assessment is assessment of learning and applicable for large-scale groups at the end of their learning process. On the other hand, formative assessment is assessment for learning which focus on providing feedbacks and continue during the learning (Black & William, 2009). On the other hand, the current study focuses on the Cognitive Diagnostic Assessments (CDAs) which can be used to obtain qualitative data from quantitative tests (Leighton & Gierl, 2007).

1.10. Cognitive Diagnostic Models (CDMs)

In recent years, a new psychometric model known as the Cognitive Diagnostic Model (CDM) has been created to address the issues that CTT and IRT are unable to address. CDMs are alternative psychometric models that give extensive and cognitive information based on the test performance of the learners. CDMs mainly aim to identify students' mastery or non-mastery attributes (de La Torre & Rutgers, 2011). Thus, cognitive models allow us to learn more about learners' weaknesses and strengths to provide information for instruction and learning. Furthermore, CDMs provide the information on students' performances and the reasons of their performances (Ravand, 2016). They break the tests down into strategies, knowledge and processes needed to complete each task correctly, allowing teachers to identify their students' mistakes or misconceptions (Embretson, 1983).

Haagenars and McCutcheon (2002) grouped CDMs as latent class models because they classify students into groups based on how similar their responses to items are. CDMs are named as *restricted* latent class models since the number of attributes in test items limits the number of latent classes (Ravand & Robitzsch, 2015). Students are divided into 2^K latent groups based on K characteristics underlying test performance (the number 2 implies that each attribute has two alternative outcomes: mastery or non-mastery). For example, for the current study, there are 6 defined attributes necessary to complete the assessment test successfully, so students are divided into 2^6=64 latent classes. To summarize, CDMs use a set of discrete/dichotomous attributes to diagnose students' proficiency. In CDMs, latent (unobservable) categorical variables to estimate the likelihood of an observable skill are used. Skills, subskills, attributes, abilities, and processes have all been used to describe these latent variables (Ravand, 2016).

According to Rupp and Templin (2008b), CDMs are “probabilistic, confirmatory multidimensional latent variable models with a simple or complex loading structure” (p. 226). Each CDM describes students’ performances in terms of the likelihood of mastery of each characteristic independently, or the possibility of each student belonging to each latent class (Lee & Sawaki, 2009). Therefore, CDMs can be classified as probabilistic models. CDMs are also confirmatory since they, like confirmatory factor analysis models, have latent variables created a priori using a Q-matrix (Ravand & Robitzch, 2015). Q matrix is kind of framework for CDM (Tatsuoka, 1985) because a Q-matrix consists of assumptions for the necessary attributes to answer each item correctly (Li, 2011). As the last key feature, it can be highlighted that CDMs are also multidimensional latent variable models because unlike one dimensional IRT models where a single score is assigned, CDMs assign respondents to multidimensional skill profiles by categorizing their skills included in the test proficiency as mastery vs non-mastery.

To sum up, there are different types of statistical models that can be used for data analysis. The purpose of the analysis has an impact on deciding the best model for the analysis. CDMs are one of these models. CDMs are designed to provide diagnostic and in-depth data from a quantitative designed assessment (Leighton & Gierl, 2007). Therefore, CDMs were used to analyze the data in the presented study. The aim of the study was to investigate the weaknesses and strengths of the grade 4 students in the domain of routine and non-routine problem-solving with four-operations by using CDMs. In the light of the aim, a CDA test is designed to identify students’ weaknesses and strengths.

2. LITERATURE REVIEW

2.1. CDM Studies

Regarding the literature, CDMs have been administered in two different ways as retrofitting (post hoc analysis) of an existing non-diagnostic test (Dogan & Tatsuoka, 2008; Duong Thi & Loye, 2019; Im & Park, 2010; Lee et al., 2011; Lee et al., 2013; Sen & Arican, 2015; Toker & Green, 2012) or designing a group of items or task for diagnostic purposes in the first place (Demir & Koç, 2018; Li et al., 2020; Lin, Xing & Park, 2020). Majority of the CDM applications are administered using retrofitting to extract diagnostic information from non-diagnostic tests. In this chapter, retrofitting studies and studies designed by diagnostic purposes are presented.

2.1.1. Retrofitting Studies

There are several studies (Dogan & Tatsuoka, 2008; Duong Thi & Loye, 2019; Im & Park, 2010; Lee et al., 2011; Lee et al., 2013; Sen & Arican, 2015; Toker & Green, 2012) have been administered to provide feedbacks to the students by using retrofitting. In retrofitting studies, the attributes are defined by using the items of the tests or exams that are prepared before (Ravand, 2016). These studies are important, and they are very useful while providing feedbacks for large-scale assessments like TIMSS or PISA (Sen & Arican, 2015). Some of the retrofitting studies in the field of mathematics are provided below.

Toker & Green (2012) investigated cognitive abilities of 8th graders' mathematical achievement in an international exam. The purpose of the study was to validate cognitive attributes on the TIMSS-2007 mathematics test items as regards the cognitive attributes developed by Tatsuoka & her colleagues (1983). In the study the least squares distance method (LSDM) was used to analyse the data. The attribute identification is applied 4498 8th grade students from seven regions of Turkey. TIMSS-2007 mathematics test included 179 questions 96 of which were multiple-choice questions. Two of the items were cancelled to provide validation since everyone answered correctly. 20 attributed were defined and Q-matrix was designed to analyse the data. The findings of the study showed that the attributes that were defined for each item provided important information in terms of item difficulty. However, the relationship between items and Q-matrix was not that successful because according to the finding 14 items were not related to the defined attributes. So, in order to obtain better findings from the study some items needed to be changed.

Sen & Arican (2015) conducted a study to compare the math scores of Korean and Turkish students in TIMSS. The purpose of the study was to analyse international large-scale assessments using CDM approaches. In that sense, their study aimed to provide diagnostic feedback for the large-scale assessments instead of a single score. The study was designed to give feedback to the samples about their weaknesses and strengths. Data for the study were taken from TIMSS 2011 8th grade mathematics tests. Therefore, the study was designed as retrofitting research and it was analysed by using DINA (**t**he deterministic, inputs, noisy, “and” gate) model. In the tests, different kinds of items like multiple-choice and constructed response were asked. The domains of the items were Number (30%), Algebra (30%); Geometry (20%); and Data and Chance (20%). For TIMSS 2011, out of 14 mathematics assessment blocks just 6 of them were open to public. The number of the items and sample sizes were different in each booklet for Korean and Turkish students. Considering these differences, Booklet 2 was chosen for the study to assess the responses of the students. There were some reasons for the researchers to choose Booklet 2. It had relatively more topics, the distribution of the topics was almost equal, the cognitive domains were also matched equally for knowing, applying, and reasoning. To sum up, Booklet had 32 items as 15 multiple-choice and 17 constructed response items. The sample size of the study was 856 students. 368 of these students were Korean and 488 of them were Turkish. Their results indicated that Turkish students did not master in the attributes comparing to Korean students. According to the results Turkish students had some difficulties in fractions and decimals especially equivalent fractions and ordering fractions. On the other hand, Korean students mastered in these attributes. Also, in geometry items the weaknesses of the Turkish students were in the topics of drawing, constructing, and describing geometrical figures. However, both Turkish and Korean students mastered at data and solving data analysis items. According to the study, another important result was about the types of items. Study showed that the performance of the

Turkish students was low on the open-ended items in comparison to multiple-choice items. So, the findings of the study showed not only the weaknesses of the students individually it also led the way for the changes in the Turkish curricula and instruction.

Dogan & Tatsuoka (2008) reanalysed the math performance of eight grade Turkish students on TIMSS-R in 1999 and compare the performances of 2900 Turkish and 4411 American students. One of the aims of the study was to show how to compare different student groups who are taking part in an international exam by using cognitive diagnostic analysis. The Rule Space Model is used to analyse the performance of the students. 23 attributes were identified and divided into three groups as content (five attributes), process (nine attributes) and skill/item type (nine attributes). 162 items were coded according to these attributes. Mastery level of each student was identified for these specified attributes. According to the results, algebra and probability/statistics were shown to be poor areas for Turkish students. Additionally, they had weaknesses in several abilities including applying algebraic rules, approximation/estimation, solving open-ended problems, identifying patterns and relationships, and quantitative reading. One of the most significant results of the study was that both Turkish and American students tend to master thinking skills and complex problem-solving abilities after becoming proficient in routine problems, numbers, and geometry but before becoming proficient in algebra-related abilities. This finding suggests that for these students, regarding their development teaching advanced and complex skills in the context of numbers and geometry may be better than teaching them via algebra.

Im and Park (2010) conducted a study to compare the math performance of Korean and American 8^{th} graders in TIMSS 2003 using cognitive diagnostic models. Another purpose of the study was to identify relationships between instruction of the teachers and mastery of knowledge and attributes of students. The samples of the study were 1179 pupils as 740 students from the US and 439 from Korea who took the Booklet 3 mathematics exam. In the study, 10 attributes and 43 items related to these attributes are identified. Rule Space Model was used to analyse the data. In the study, different data sets are integrated to examine students' mathematical abilities and knowledge in relation to characteristics relevant to instruction that were included in the TIMSS 2003 data set. These data sets are students’ score data, teachers’ and students’ survey data, the data from Rule

Space Model and teacher and student link data. According to the results of the study, there is a significant difference between the performances of Korean and American students in terms of problem restructuring and reasoning, geometry, and measurement. On the other hand, data, and probability, translating word into equations and visualizing/utilizing figures and graphs were considered as difficult attributes for all students. They also showed that encouragement of students' individual and independent problem solving was the most effective approach for American and Korean students. For the US students, reviewing, reteaching, and clarifying the material were particularly beneficial.

As another retrofitting study example, Lee et al. (2011) carried out a study to compare the fourth-grade students' math achievement levels in Massachusetts and Minnesota to the national results excluding these two regions on the TIMSS 2007. The aims of the study were to use CDM to determine item characteristics including discrimination, slip-and-guessing parameters and to assess students' proficiency of attributes and their ability to effectively apply them in an exam context. By doing this, it might be easier to look at how the two benchmark states and the rest of the US are similar or different in terms of attribute mastery of the students. In the study, TIMSS 2007 4^{th} grade math assessment was used. The assessment test had content domain items and cognitive domain items. Totally, 25 items as 15 multiple-choice and 10 constructed response items were used in the study. 15 attributes were identified, and DINA model was used to analyse the collected data. According to the results of the study, the performance of the students in Massachusetts and Minnesota was higher than the US overall except for one attribute which is data display. Therefore, teachers should make sure that students in Massachusetts and Minnesota thoroughly comprehend and have the mastery of interpreting data from tables, pictographs, bar graphs, and pie charts to prevent them from guessing. Additionally, Lee et al. (2011) presented diagnostic data about the performances of the students which they claimed may be precisely applied to instruction. They benefited from item parameter estimates such as slipping and guessing to provide suggestions for the curriculum on how to increase students' math performances.

Lee et al. (2013) also implemented a retrofitting study and develop a multi-group DINA (MG-DINA) model to reanalyse the math performance of 8^{th} grade students in TIMSS 2007. The aim of the study is to identify cognitive strengths and weaknesses of students' math performances of the countries using visualization methods. To achieve this aim 17 countries with at least 8 benchmark participants, including the United States, were chosen among 59 countries. Two math instructors created a Q-matrix for the items that comprises nine attributes or abilities evaluated by TIMSS exam. Four subject domains were determined as numbers, algebra, geometry and data and chance. The provided data which includes 88 item responses from 17 different countries, was fitted using MG-DINA model to produce a total of 176 item parameters as 88 guessing and 88 anti-slipping parameters. The model also showed a total of 153 attribute prevalence proportions which consist of 9 attribute proportions for each country. Another aim of the study was to analyse the similarities and differences in how attribute prevalence varied across countries since the chosen model allowed to estimate attribute prevalence. The study relied on multidimensional scaling (MDS) and clustering methods to assess and explain the data. These methods were used to determine the similarity or difference in attribute prevalence that is required to answer the items correctly on the assessment. The results of the methods that were used in the study offered various perspectives on how to examine similarity of the countries. It is clear to see that Taipei, Korea, Singapore, Hong Kong, and Japan perform better in any attribute level.

Even though this group of countries may be considered as "high-achieving", the weaknesses and strengths of each country considerably varied. Russia and Israel, for instance, have higher level of success than Korea, yet they have similar success patterns. However, there are several countries having similarities in both success levels and success patterns such as the US and Hungary or England and Scotland. On the other hand, Turkey is a country that does not have either a similar level of success or success pattern with other countries examined in the study. The results of the study may suggest that there is no one optimum curriculum or set of teaching techniques without taking into consideration specific educational conditions of the countries. It is also important to indicate that the study showed Japan has very high results for some attributes and they are considerably better at these abilities while the majority of the other countries struggle.

2.1.2. Studies Designed by Diagnostic Purposes

A limited number of studies (Demir & Koç, 2018; Li et al., 2020; Lin, Xing & Park, 2020) have been implemented to provide feedbacks to the students by designing a test or task with diagnostic purposes. These studies have been provided effective and diagnostic feedback on students' learnings and performances. Since they are designed by diagnostic purposes, they can pinpoint the needs of the students and deficiencies of the instruction. Some of the studies that are designed with diagnostic purposes are discussed below.

Li et al. (2020) administered a study on kindergarteners' mathematical problem-solving skills. The purpose of the study was to assess the mathematical problem-solving ability of kindergarteners using CDAs. Through the study it is also aimed that to develop an instrument for the students to assess children's problem-solving abilities in the domain of numbers and operations. The sample size of the study was 747 kindergarteners from 12 kindergartens. The samples were chosen from both rural and urban areas to obtain a representative sample. A test with 11 attributes and 38 items were developed to use for the study. For the test 30 minutes were given. Before the test to check validation, interviews were administered. The items were coded using 0-1 score system. 0 was used for the attributes that were not mastered, 1 was used for the attributes that were mastered. The results of the study indicated that cognitive diagnostic tests are effective. Also, the instrument that is developed is reliable for the assessment of kindergarteners' mathematical problem-solving skills in number and operations.

Lin, Xing and Park (2020) conducted a study to evaluate the development of skill mastery and assess the impacts of attribute-level interventions over time. This study suggests longitudinal CDMs integrate latent growth curve modelling and covariate extensions. The research illustrated implications of unconditional and conditional latent growth CDMs using data from the real-life. The study was divided into three sections as one real-life data and two simulation studies. In the study, real-life data was used to illustrate how the model may be used, support the justification for the latent growth framework, and track adjustments in the skill mastery of students and intervention effects. The simulation studies were conducted independently to analyse the parameter recovery of the suggested models. In this way, simulation studies with diverse longitudinal design elements offer thorough inference for a range of variables, including sample sizes, time points, and covariate specification. Simulations demonstrated consistent parameter recovery and latent class classification for various sample sizes. These results imply that applications of covariate based longitudinal CDM can be useful to see the impact of explanatory variables and intervention on the development in attribute mastery. These applications can be built upon the well-established growth modelling frameworks.

3. METHODOLOGY

In the methodology section, the sample of the study, data collection instrument, steps of CDA design, construction of Q-matrix, data collection process, data analysis and data analysis methods used in the study like model-fit indices, item parameters, accuracy and consistency are explained.

3.1. Sample

The sample size influences the recovery of CDM model parameters, since item parameter recovery gives better results with larger sample sizes (Başokçu, 2014; Sen & Cohen, 2021). To determine the sufficient sample size, Sen and Cohen (2021) indicated that if the sample size is less than 200, the results are found poor. To achieve exact estimates, the sample size needs to be at least 500 for C-RUM, DINA, DINO, and reduced-LCDMs. Therefore, the sample size of the current study is determined as 511 4^{th} grade students. Details about participants are shared in Table 3.1. As shown in Table 3.1, the samples are selected from public and private schools by the given amounts.

Table 3.1. The number of students according to schools (N=511).

School Type	**Schools**	***n***	**Total**
Public	A	206	400
	B	194	
Private	C	23	111
	D	6	
	E	73	
	F	9	

In the current study, the purposive sampling method was used. Public and private schools were selected regarding the success of the school, their location and socio-economic level of the neighbourhood that school located in. Having a representative sample group for Istanbul was aimed. Therefore, schools were chosen from various regions of both Asian and European side in Istanbul like Etiler, Ulus, Çekmeköy, Ataşehir, Ümraniye and Sultangazi. Before administering the test, 10 students were selected voluntarily to conduct an interview about items. While the current study was approved by the Ethical Committee of the Institute of Graduate Studies in Science and Engineering of Bogazici University. Also, legal permission was taken from Turkish Ministry of Education to collect data.

3.2. Instrument

In the line with the aim of the current study, a cognitive diagnostic mathematics test was developed and was administered to 4th grade students. The test consists of 6 attributes and 20 items. All the attributes and items were developed in the scope of 4th grade mathematics curriculum. The items of the test cover the routine and non-routine problem-solving skills with four operations in natural numbers (addition, subtraction, multiplication, and division). In order to develop the items, 2021-2022 education year fourth grade math books are reviewed, the questions shared by ministry of education were examined and various textbooks with routine and non-routine questions were analysed in detail. For the non-routine problems, the researcher collaborated with Boğaziçi University Adaptive Testing Lab (BAT). First, 30 items were written, then regarding the timing issue, some of them were eliminated. In the elimination process, the attributes required by the items, appropriateness of the students' level and clearness of the items were considered. At the end a test with 20 item was created. The items of the study consist of routine and non-routine problems that require four arithmetic operations with natural numbers. To finalize the instrument, four booklets were created for the study. While items in each booklet remained the same, the order of the items was changed by the researcher. In that way, students were prevented from cheating. Also, teacher-class interactions were aimed to decrease since the teacher may lead or mislead the students while answering the items.

3.3. Steps of CDA

Educational tests which are designed with cognitive diagnostic purposes differs from traditional tests. Unlike traditional approaches, cognitive diagnostic assessment tests do not only depend on taxonomies or objectives (Leighton & Gierl, 2007). CDAs directly focus on the mechanisms that test takers use while answering the items or tasks. CDAs make assumptions about the learners regarding the knowledge level and the processes, how to use the knowledge, how to improve the learning processes, the differences of the learners than others in terms of their weaknesses and strengths (Nichols, 1994).

There are different suggested paths while designing a CDA test. According to Nichols (1994), the first step is to devise a model or theory that illustrates the assessment's target knowledge structures, abilities, and processes. The second one is design selection like choosing the test items considering the cognitive processes and knowledge structures that is identified in the first step. Next is test administration which includes deciding on the item format and test setting. Then, the following step is response scoring which activates the assessment design identified in test administration step. The last step is design revision to check whether it fits in with the model or theory.

Embretson (1994) also suggested Cognitive Design System (CDS) approach to emphasize the part of cognitive theory in CDA test development. CDS mainly includes three parts which is developing the items, writing, and analysing (Gorin, 2007). Embretson (1994) identified seven steps for test development: identifying aims of the assessment, deciding the features of the assessment task, developing a cognitive model, generating the items, evaluating, and checking the model for the generated test, keeping the items with cognitive accuracy and validation. Gorin (2007) especially emphasize the model-fit checking for the tests and changing the items if it is necessary.

Regarding these two significant pathways to develop a CDA test, the steps for the current study are determined as follows:

1. Identifying aims and attributes of the study
2. Design Selection
3. Writing Items
4. Q-matrix construction
5. Test Administration
6. Model-fit and validity Check
7. Revision (if necessary)

3.3.1. Identifying Cognitive Attributes

Leighton and Gierl (2007) defined cognitive attributes as the conceptual knowledge and procedural skills required to achieve a task or test in a particular domain. In this study, cognitive attributes are referred to the conceptual knowledge in the domain of four operations required to solve each test item and conceptual skills required to perform routine and non-routine mathematical problem-solving skills. In the beginning of the attribute identification process, 4th grade mathematics curriculum of Turkey was examined. The learning areas, subjects, objectives, explanations, and questions were identified. Additionally, the tests prepared by the ministry of education and several current test books are analysed. After the examinations, the attributes are identified as “adding, subtracting, multiplying, dividing, routine problem-solving and non-routine problem-solving”. The objectives of ministry of education are shown in Table 3.2.

Table 3.2. MEB Objectives.

-Students will be able to solve addition problems. *(M.4.1.2.4. Doğal sayılarla toplama işlemi gerektiren problemleri çözer.)*
-Students will be able to solve subtraction problems. *(M.4.1.3.4. Doğal sayılarla toplama ve çıkarma işlemi gerektiren problemleri çözer.)*
-Students will be able to solve multiplication problems. (*M.4.1.4.6. Doğal sayılarla çarpmaa işlemi gerektiren problemleri çözer.)*
-Students will be able to solve division problems. (*M.4.1.5.6. Doğal sayılarla en az bir bölme işlemi gerektiren problemleri çözer.)*

The attributes were related to problem-solving skills with 4 operations as addition, subtraction, multiplication, and division. The cognitive domains included in the assessment are, routine and nonroutine problems. While deciding the attributes, similar studies in the literature was checked to make sure about routine and non-routine problem-solving attributes (Lee, Park & Taylan, 2011; Su, Choi, Lee, Choi & McAninch, 2013). Attributes required by the test items are presented in Table 3.3.

Table 3.3. Attributes of the Study.

Attribute 1	A1	Adding 2-, 3-, and 4-digit numbers
Attribute 2	A2	Subtracting 2-, 3-, and 4-digit numbers
Attribute 3	A3	Multiplying 2-, 3-, and 4-digit numbers
Attribute 4	A4	Dividing 2-, 3-, and 4-digit numbers
Attribute 5	A5	Solving routine problems
Attribute 6	A6	Solving non-routine problems

3.3.2. Design Selection

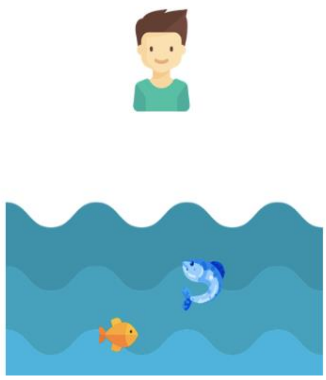
It has been determined that all items would be presented as multiple-choice questions. At the beginning the number of the items was decided as 30. However, regarding the timing and the targeted age group it would be tiring for the students. Because of these issues some items were eliminated. During the elimination process, it is regarded to have an equal number of items for each attribute, appropriate level for students and clear explanations and pictures. Regarding all the criteria above, the number of the items was decreased to 20.

3.3.3. Writing Items

After determining the subject and identifying the cognitive attributes, multiple-choice test items were prepared by the researcher who is currently a mathematics teacher and taught 4th graders in previous years. Some of the non-routine items are chosen among the questions of Boğaziçi University Adaptive Testing Lab, with a permission and then revised. For the other items, literature was searched and the items in similar studies were considered as examples. Then, items were developed regarding the level of students and defined cognitive attributes. These items were also evaluated by a faculty member from the field of measurement and evaluation.

3.3.3.1. Revision of the Items. After the proposal presentation of the presented study, some revisions are done on the items based on the recommendations of the thesis jury. For example, in the initial form of the item 9 the distance was not straightforward, thus the item was unclear. Then, the item was revised (see Figure 3.1) regarding the feedback from the jury. The final form of the item is provided in Figure 3.2.

Suya dışarıdan bakan bir gözlemci, suyun altında bulunan cisimlerin kendisine olan mesafesini, gerçek mesafeden daha yakın olarak algılar. Gözlemci tarafından algılanan bu mesafe ‘’görünür derinlik’’ olarak isimlendirilir. Örneğin, gerçekte 20 metre derinlikte bulunan bir balığa dışarıdan bakan bir gözlemci için görünür derinlik 20 metreden daha azdır.



Ali, suya dışarıdan bakmaktadır.

- Sarı balık ve mavi balık arasındaki gerçek mesafe: 800 mm
- Ali için sarı balığın görünür derinliği: 1500 mm

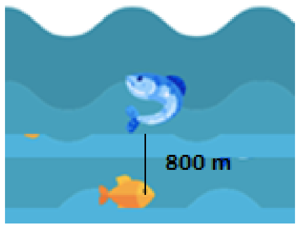
Yukarıda verilen bilgilere göre mavi balığın gerçek derinliği kaç mm olabilir?

A) 560 B) 601 C) 699 D) 701

Figure 3.1. Initial Form of Item 9

Suya dışarıdan bakan bir gözlemci, suyun altında bulunan cisimlerin kendisine olan mesafesini, gerçek mesafeden daha yakın olarak algılar. Gözlemci tarafından algılanan bu mesafe ‘’görünür derinlik’’ olarak isimlendirilir.

Örneğin, gerçekte 20 metre derinlikte bulunan bir balığa dışarıdan bakan bir gözlemci için görünür derinlik 20 metreden daha azdır.



Ali, suya dışarıdan bakmaktadır.

- Sarı küçük balık ve mavi büyük balık arasındaki gerçek mesafe: 800 mm
- Ali için sarı küçük balığın görünür derinliği: 1500 mm

Yukarıda verilen bilgilere göre mavi büyük balığın gerçek derinliği kaç mm olabilir?

A) 560 B) 601 C) 699 D) 701

Figure 3.2. Final Form of Item 9

3.3.4. Construction of Q-Matrix

One of the most important steps in cognitive diagnostic models is to create a Q-matrix which shows the link between attributes and items. Q-matrix is developed to define each attribute that is measured by the assessment. Tatsuoka (1983) was the first to develop a Q-matrix for a CDM, his Rule-Space Model (RSM), and apply it to 5th grade Fraction-Subtraction problems. To limit the number of allowed attribute profiles, attributes were arranged in a hierarchy. Since then, a Q-matrix has served as the starting point for CDM building. Tatsuoka (1985) pointed out that Q-matrix is a binary item-by-attribute matrix. The process of specifying the number of attributes and their interactions is referred to as attribute structure specification. The definition of item-attribute alignments is the process of appropriately determining which items are used to measure which characteristics (Dogan & Tatsuoka, 2008). Shortly, Q-matrix is a representation that shows the relationship between items and attributes (Koyuncu, 2020). A lot of diagnostic information can be obtained by using a well-designed Q-matrix.

Q-matrix is created as $[J \times K]$ dimensional matrix with the 0 or 1 inputs where J represents the number of items and K represents the number of attributes (de La Torre & Chiu, 2016). In Q-matrix, rows include items and columns include attributes. Each item is coded by 0 or 1 considering the existence of the defined attribute. 0 means that the attribute is not required to be mastered and 1 means that the attribute is required to be mastered to answer the question correct. As an example, for a $[J \times K]$ Q-matrix, demonstration can be shown as follows:

Table 3.4. A Q-matrix example with Three Attributes for Cognitive Diagnostic Models.

	Attributes		
	Attribute 1	Attribute 2	Attribute 3
Item 1	1	0	0
Item 2	0	1	0
Item 3	0	0	1

Based on Table 3.4, to be able to answer item 1, only attribute 1 is required. Similarly, to be able to answer item 2 correctly, the participants need to master attribute 2, for item 3 they need to master attribute 3. Table 3.4 is a basic example for Q-matrix design. In a different example, each item may require more than one attribute as shown in Table 3.5.

The total number of possible learner profiles depends on the number of attributes and their hierarchical relationship. If all attributes are independent which means that each item corresponds only one attribute (Sun et al., 2013), then the number of all possible learner profiles depends on the number of attributes. If "K" attributes are defined by binary system (0,1), the total number of possible learner profiles is 2^k. Based on the Table 3.4 example, there are three independent attributes, so the total number of possible learner profiles is 2^3. This means that there are 8 different latent classes as (0 0 0), (1 0 0), (0 1 0), (0 0 1), (1 1 0), (1 0 1), (0 1 1), (1 1 1). These latent classes show the weaknesses and strengths that each group member has. For example, the learners that is grouped in (0 0 0) latent class does not have any defined attributes. As another example, the members of latent group (1 0 0) have achieved only attribute 1 but they need to improve attribute 2 and 3. On the other hand, the learners who are classified in (1 0 1) group are good at attribute 1 and 3 but they do not have attribute 2. In that way, using Q-matrix is helpful to diagnose the strengths and weaknesses.

Table 3.5. A Q-matrix example with Five Attributes for Cognitive Diagnostic Models.

	Attributes				
	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
Item 1	1	0	0	0	1
Item 2	0	1	0	1	1
Item 3	0	0	1	1	1
Item 4	1	1	1	1	0
Item 5	1	1	1	1	1

3.3.4.1. Construction of Q-matrix for the Current Study. The initial Q-matrix of the current study was developed by the researcher. The Q-matrix of the current study had 20 multiple-choice items for 6 attributes. To evaluate the Q-matrix, the items and attributes were asked to match by one academic, one homeroom teacher and two math teachers. As a result of their suggestions, some revisions are made on the Q-matrix. The final form of the Q-matrix is provided in Table 3.6. Table 3. 6 also shows the number of the items that requires mastering each of the attributes.

Table 3.6. Final Form of Q-Matrix.

	Attributes					
items	A1	A2	A3	A4	A5	A6
1	1	0	0	0	1	0
2	1	0	0	0	1	0
3	1	0	0	0	1	0
4	1	0	0	0	0	1

Table 3.6. Final Form of Q-Matrix (cont.).

5	1*	1	0	0	0	1
6	0	1	0	0	1	0
7	0	1	0	0	1	0
8	0	1	0	0	1	0
9	0	1	0	0	0	1
10	0	1	0	0	0	1
11	0	0	1	0	1	0
12	0	1	1	0	1	0
13	1	0	1	0	1	0
14	0	0	1	1	0	1
15	0	0	1	0	0	1
16	0	0	0	1	1	0
17	1*	0	0	1	1	0
18	0	0	0	1	1	0
19	0	0	0	1	0	1
20	0	0	1*	1	0	1
frequency	7	7	6	6	12	8

*= *revised values*

The items of the study are given in Appendix 1. Two of the items for the exam are shown in Figure 3.3 and Figure 3.4:

Karavanla tatile çıkan bir aile tatilin,

• birinci haftasında 2230 km,

• ikinci haftasında 1985 km,

• üçüncü haftasında 2368 km yol gitmiştir.

Buna göre üç haftanın sonunda gidilen yol toplam kaç kilometredir?

A) 2368 B) 4215 C) 4353 D) 6583

Figure 1.3. Example for Routine Problems.

Hatice, gittiği bir manavdan her biri 150-gram olan 6 tane elma ve her biri 10-gram olan çileklerden bir miktar alacaktır. Manav; elmaları terazinin sağ kefesine, çilekleri terazinin sol kefesine koyduğunda yandaki gibi bir görüntü elde etmiştir.

Buna göre Hatice en az kaç tane çilek almıştır?

A) 89

B) 90

C) 91

D) 92

Figure 3.4. Example for Non-Routine Problems.

3.4. Data Collection Process

To evaluate the clarity of the items, the test was administered to a group of 10 volunteer students from 4th grade. The answer sheets of the participants were checked and the semi structured interviews with the participants were completed to ensure the clarity of the items, timing, and other issues with same group. Interview questions are provided in Appendix 2. The clarity of the items and timing were asked to the students. After the interviews, revisions were made, and the final form of the items was prepared. Regarding these feedbacks, the duration of the test was decided to be 60 minutes.

The final form of the test was administered to 511 students from various schools in Istanbul. Detailed information about schools and students were provided in the samples section. As in the trial test, the place of the items in the test was changed and the test was organized as four forms. These 4 booklets were named as A, B, C, and D.

The participants answered 20 items in 60 minutes in their own classroom. For the test, two consecutive lesson hours were allocated during the day. The test administered under the invigilation of homeroom teacher and the researcher. Before answering the items, the test instructions were explained to the students. Students were told that correction for guessing would not be used, thus it was tried to encourage students to provide an answer for all the questions. All the students completed the test within 60 minutes.

3.5. Data Analysis

There are various CDMs such as deterministic, inputs, noisy “and” gate (DINA), deterministic inputs, noisy “or” gate (DINO), noisy-input, deterministic “and” gate (NIDA) and the reduced reparameterized unified model (R-RUM). General CDMs are the log linear cognitive diagnostic model (LCDM), the general diagnostic model (GDM) and generalized DINA (GDINA) model (Ravand, 2016; Ravand & Robitzch, 2015). This study focuses on the GDINA model to analyse the collected data. Thus, a brief description of CDMs and GDINA model is given below.

3.5.1. General Properties of CDMs

According to Gulliksen (1950), one of the most important points for test development is to focus on the relationship between the observed score and skills aimed to be measured. In the field of education, these skills cannot be measured directly since they are latent or unobservable skills or attributes (Demir & Koç, 2018). Therefore, the importance of the latent models has increased and accelerated.

There is different naming for CDMs in the literature. They can be expressed as restricted latent class models (Haertel, 1989), cognitive psychometric models (Rupp, 2007), multiple classification models and structured item response theory models (Rupp & Mislevy, 2007). According to Rupp and Templin (2008a), the reason for this diversity comes from the expression of aspects of the models regarding certain features. Some definitions reflect the theoretical background of the models, some reflect the purpose of the models, and some reflect the statistical properties of the models.

Restricted Latent Class Models are mentioned by Haertel (1989). Depending on these models, participants are divided into latent classes. However, there is a restriction for the number of identified latent class models (Haertel, 1989). Cognitive Psychometric Models focus on two important points. First, it is essential to create multi variable profiles based on the participants' responses and second, the level of the participants according to degree of carrying the characteristics of their profile (Rupp &Templin, 2010). Structured Item Response Theory Models (SIRT) (Rupp & Mislevy, 2007) demonstrates that the theoretical foundations that underlie the creation and use of a particular set of psychometric models. SIRT models depends on evidential arguments. Cognitive Diagnostic Models aim to show the relationship between items and skills like Item Response Theory models. In that sense, they have a common point in essence. Thus, the root of CDMs comes from IRT models.

3.5.2. CDM Types

CDMs are categorized into two groups as specific and general based on whether they have inter-skill relationships or not. This categorization is summarized by Ravand and Robitzch (2015) like in Table 3.7. General CDMs can include both compensatory and non-compensatory types in the same test. On the other hand, specific models can include only compensatory or only non-compensatory types within the same test. In that case, the advantage of the general models is they provide an opportunity to choose the best fit model instead of forcing to pick one single model for all items Therefore, all specific models may be covered by general models (Ravand, 2016).

For compensatory models, mastery of a skill needed to answer the item successfully might compensate non-mastery attributes. It means that even though the test takers fail to answer corresponding item correctly, they may compensate this by answering another item to achieve the attribute (Ravand & Robitzch, 2015; Ravand, 2016). However, in non-compensatory / conjunctive models, lack of mastery of any skill cannot be totally compensated by other skills based on item performance. To sum up, if a learner did not master any of the required attributes s/he cannot answer the question correct. To give the correct answer the learners must master all of the required attributes.

Table 3.7. CDM Types and Examples.

Source: [Ravand & Robitzch, 2015; Ravand, 2016].

	CDM Type	Examples	Author(s)
Specific	Compensatory	1. Deterministic-input, noisy-or-gate model (DINO)	Templin and Henson (2006)
		2. Compensatory reparameterized unified model (C-RUM)	S. M. Hartz (2002)
		3. Additive CDM (ACDM)	de la Torre (2011)
	Non-Compensatory	1. Deterministic-input, noisy-and-gate model (DINA)	Junker and Sijtsma (2001)
		2. Non-compensatory reparametrized unified model (NC-RUM)	DiBello, Stout, and Roussos (1995); S. M. Hartz (2002)
General	Both compensatory and non-compensatory	1. General Diagnostic Model (GDM)	von Davier (2005)
		2. Log-linear CDM (LCDM)	Henson, Templin, and Willse (2009)
		3. Generalized deterministic-input, noisy-and-gate model (GDINA)	de la Torre (2011)

<u>3.5.2.1. DINA Model.</u> First studies on DINA Model were administered by Macready and Dayton (1977), Haertel (1989), and Tatsuoka (2002). After these studies, de la Torre and Douglas (2004), and Junker and Sijtsma (2001) contributed to the literature by revising the DINA model. DINA model creates a basis for cognitive diagnostic assessment to explain the relationship between properties of cognitive items and the qualification of the individuals. (de La Torre, 2008).

DINA consists of the first letters of "Deterministic Input Noisy and Gate". The DINA model requires just two parameters for each object, regardless of how many qualities are necessary. The term of "deterministic input" indicates "1" if a participant has proper latent skills that is required by the item; it indicates "0" in opposite condition. "and" expression means that the model is non-compensatory (Toker & Green, 2012). In non-compensatory models, participants need to have all required and related skills to be able to answer the item correct (de La Torre, 2009b). Since DINA model is a non-compensatory model, to be able to answer an item correctly, participants must have all attained attributes. Otherwise, both a participant does not have just one attribute and a participant does not have all attributes will be considered as non-mastery. Although DINA model is considered as an extension for IRT models, it identifies students' attributes by dividing them into latent class models instead of measuring as a continuous variable (Haertel, 1989).

DINA model shows the relationship between latent variables and observable variable based on probability. It also produces *"slipping (s)"* and *"guessing (g)"* parameters as item parameters. Slipping and guessing parameters can be shown as:

$$s_j = P\ Y_{ij} = 0 \mid \eta_{ij} = 1 \tag{3.1}$$

$$g_j = P\ Y_{ij} = 1 \mid \eta_{ij} = 0. \tag{3.2}$$

S parameter represents that participant answers the item *(j)* incorrectly even though s/he has required attributes. This is known as the false positive probability. The lower the value of the s parameter for the item, the higher the probability that individuals with the desired characteristics will answer the item correctly (Zhang, 2006). The *g parameter,* on the other hand, indicates that the individual answers the item correctly even though he or she does not master the necessary attributes. This is known as true positive probability. The higher the value of the g parameter, the higher the probability of answering the item correctly for individuals who do not have the necessary qualifications to answer the item correctly. Similarly, the lower the value of the g parameter, the higher the probability that the item will be answered correctly only by individuals with the necessary characteristics (Zhang, 2006). Mathematical expression for DINA Model is given below:

$$P(X_{ij}) = 1\ \eta_{ij}, s_j, g_j = 1 - s_j\ {}^{\eta_{ij}}\ g_j^{1-\eta_{ij}} \tag{3.3}$$

where P represents the probability of the student who has all the required skills or attributes to answer the item correctly. Tatsuoka (1983) stated $\alpha_i = (\alpha_{i1} \ldots\ldots \alpha_{ik})$ as “knowledge states” with $\alpha_{ik} = 0$ or 1 based on having attribute k; $\eta_i = (\eta_{i1} \ldots\ldots\ldots \eta_{ij})$, j =the total number of items, as a sign of if all necessary attributes for each item mastered by the participant I, and y_{ij} as the observed score.

<u>3.5.2.2. GDINA Model</u>. GDINA (Generalized DINA) model is one of the compensatory CDM. Like most CDMs, GDINA model is also constructed using J×K Q-matrix and it also has $L = 2^{K_j^*}$ latent classes. Each latent class can be demonstrated by an attribute vector ($\propto_{lj}^*$) and each latent class has a success probability which is calculated by P($\propto_{lj}^*$) (De La Torre, 2011). In DINA model, the possibility of answering an item correctly is possible only if student has all attributes or skills required by the item (de La Torre, 2011). For any other cases, the possibility of answering an item always remains at minimum level. On the other hand, In GDINA model each attribute has an individual effect on the possibility of answering an item correctly (De La Torre & Rudgers, 2011).

The GDINA model also calculates the probability of each $P(\propto_{lj}^*)$ that participant may have for each item. In GDINA model, if a student has one or more attributes or skills required by the item, the probability of answering the item correctly changes based on the weighting of the attribute (De La Torre & Rudgers, 2011).

The item response function of GDINA model can be defined by using one of many *link functions* to relate the likelihood of a correct answer to the model requirements, as the identity, logit, or log link. (de La Torre, & Chiu, 2016; McCullough & Nelder, 1999). The identity link is used to create the GDINA model's canonical form, and its response function (De LaTorre & Rudgers, 2011) is expressed as:

$$P_j(\alpha_{ij}^{*}) = \delta_{j0} + \sum_{k=0}^{K_j^*} \delta_{jk}\, \alpha_{1j} + \sum_{k'>1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\, \alpha_{lk}\, \alpha_{lk'} + \ldots + \delta_{j(12\ldots K_j^*)} \prod_{k=1}^{K_j^*} \alpha_{lk} \tag{3.4}$$

where

δ_{j0} is the intercept for item *j;*

δ_{jk} is the main effect due to α_k;

$\delta_{jkk'}$ is the interaction effect due α_k and $\alpha_{k'}$; and

$\delta_{jk\ldots\ldots K_j^*}$ is the interaction effect due to α_1 ……… $a_{k_j^*}$.

As above the original $P\ (\propto_{lj}^{*})$ based formula for GDINA can be divided into parts according to the total effects of each specific skill and the interaction of skills with each other. The probability formula for the GDINA model (Ma, & de LaTorre, 2020) is given below:

$$P_j(\alpha_{ij}^{*}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\, \alpha_{1k} + \sum_{k=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\, \alpha_{lk}\, \alpha_{lk'} + \ldots + \delta_{j(12\ldots K_j^*)} \prod_{k=1}^{K_j^*} \alpha_{lk} \tag{3.5}$$

where δ_{j0} is the intercept, δ_{jk} is the main effect due to a single attribute α_k, $\delta_{jkk'}$ is a first-order interaction effect between α_k and $\alpha_{k'}$, and $\delta_{j12\ldots k*j}$ is the highest-order interaction effect due to $\alpha_1, \ldots, \alpha_{k*j}$.

<u>3.5.2.3.</u> <u>DINO Model.</u> The DINA model's compensating equivalent is the DINO (Ravand & Robitzsch, 2018). The DINO model states that mastery of any single attribute enhances the likelihood of successfully answering any given question in the same way as mastery of all necessary characteristics would (de La Torre, 2011). The formula of DINO model is given as

$$P\,(X_j = 1|\alpha_1, \alpha_2) = g_j^{(1-\alpha_1)(1-\alpha_2)}\,\left(1 - s_j\right)^{1-(1-\alpha_1)(1-\alpha_2)} \tag{3.6}$$

In the given formula as similar to DINA model, $1 - s_j$ represents the probability of not slipping estimate for item j, and g_j represents the probability of guessing for an item j. Compared to GDINA model parameters, it stands for $\delta_{j0} = g_j$ and $s_j = 1 - \delta_{j0} - \delta_{j1} = 1 - \delta_{j0} - \delta_{j1} - \delta_{j2} = \delta_{j1} - \delta_{j2} - \delta_{j12}$.

3.5.2.4. ACDM. The GDINA model's interaction effects are all adjusted to zero to produce the ACDM (Ravand & Robitzsch, 2018). According to the ACDM, the likelihood that a student would correctly answer item j, which calls for the two attributes a_1 and a_2, is as follows

$$P\ (X_j = 1|\alpha_1, \alpha_2) = \delta_{j0} + \delta_{j1}\alpha_1 + \delta_{j2}\alpha_2 \tag{3.7}$$

Each attribute in the ACDM adds incrementally to the increase in the likelihood of a successful response, and the absence of one attribute can be made up for by a mastered attribute since it is a compensatory model (de la Torre, 2011).

3.5.2.5. NC-RUM. By setting all interaction effects to zero, NC-RUM may be produced from the GDINA similarly to the ACDM (Ravand & Robitzsch, 2018). However, in contrast to the ACDM, the NC-RUM is estimated using a log link function rather than an identity link function (de la Torre, 2011). The following information is provided for a two-attribute item's item response probability

$$\log\ (P\ (X_j = 1|\alpha_1, \alpha_2) = \delta_{j0} + \delta_{j1}\alpha_1 + \delta_{j2}\alpha_2 \tag{3.8}$$

The NC-RUM is often reparameterized as follows

$$P\ (X_j = 1|\alpha_1, \alpha_2) = \pi_j\ r_{j1}^{1-a_1}\ r_{j2}^{1-a_2} \tag{3.9}$$

As a result, some researchers contend that the NC-RUM is the non-compensatory equivalent of the ACDM (Roussos et al., 2007).

3.5.2.6. C-RUM. By setting all interaction effects to zero, the C-RUM may also be produced from the G-DINA, much as the ACDM and NC-RUM. However, the C-RUM utilizes a logit link function as opposed to the ACDM and NC-RUM (de la Torre, 2011). For a two-attribute item, the following is the item response probability

$$\text{logit P}\ (X_j = 1|\alpha_1, \alpha_2) = \delta_{j0} + \delta_{j1}\alpha_1 + \delta_{j2}\alpha_2 \tag{3.10}$$

3.5.3. Data Analysis

After the data collection process, the scores and the responses of the participants first checked manually. Then, all data is entered on an excel file. Each participant is entered to the dataset considering if they answered the item correctly or not**.** In order to check the internal consistency, Cronbach's alpha (α) (Cronbach, 1951) was calculated in SPSS (IBM, 2017). Then, students' response dataset and designed Q-matrix is used to analyse the data. Relative fit indices were calculated between GDINA, DINA, DINO, ACDM, C-RUM and R-RUM. Therefore, it is decided for this study to analyse the designed CDA by using GDINA model based on the test design, the model fit analysis results and literature.

The data of the study is analysed with GDINA model using GDINA package, version 2.8.8. (Ma & de La Torre, 2022) and CDM package (Robitzsch, Kiefer, George, Ünlü, 2022) in R software (4.2.0) (R Core Team, 2022). The syntax for the GDINA model is written in R program. The responses of the students were coded as 0-1 As a result of the analysis, GDINA gives the model and item fit indices, the item parameters, and the standard errors of these parameters, the profile distributions of the attributes of the model and the latent classes of the students. On the other hand, classification consistency values are calculated by using CDM package.

As a result of the analysis, the values of guessing (*g*), slipping (*s*) and GDINA discrimination (GDI), which are the GDINA model parameters that determine the quality of the developed test and the Q-matrix, are obtained. Model data fit indices, item parameters and latent class information related to the final test were examined. Additionally, 2^6 attribute profiles were constructed to identify strengths and weaknesses. Considering the students' responses, students were assigned to attribute profiles. Detailed information provided in the results section.

3.5.4. Analyses of the Study

In this part, detailed information on the analyses used in the study presented as follows; reliability, assessment of model relative and absolute fit to the data, item parameters, identify students' skill mastery profiles and cross-validation.

<u>3.5.4.1. Reliability</u>. Cronbach's alpha (α) (Cronbach, 1951) is stated as a value between 0 and 1 to quantify the internal consistency of a test or scale. Internal consistency refers to how closely each test item measures the same notion or construct, and it is related to how closely one item inside the test is related to the other (Tavakol, & Dennick, 2011). Before a test is used for study or testing, its internal consistency should be established to assure validity. Therefore, to make sure about the internal consistency of the study Cronbach's alpha value was calculated. The acceptable value for alpha, according to several studies (Cronbach, 1951; DeVellis, & Thorpe, 2021; Tavakol, & Dennick, 2011), is 0.70.

<u>3.5.4.2. Model Fit Statistics</u>. Evaluating model fit to the data allows for testing of fundamental coherency between the estimated model and observed data to offer model modifications (DiBella, Roussos & Stout, 2007; Sinharay, Almond & Yan, 2004). Fit indices are often computed at two levels at the test and item levels: relative fit indices and absolute fit indices. While deciding the best fitting model out of many other models, data from the relative fit statistics are beneficial (Chen, de La Torre & Zhang, 2013). On the other hand, the absolute fit of the model indicates whether the models appropriately match the data or not (Duong Thi, & Loye, 2019).

<u>3.5.4.3. Relative Fit Statistics</u>. The purpose of the analysis of models' relative fit to the data is to check the accuracy of the model. In order to choose the most appropriate model to analyse the data among competing models, relative fit statistics are evaluated. (Chen, Torre & Zhang, 2013). The following three statistics:

1. Deviance statistic (i.e., -2log-likelihood (-2LL)): 2ln (ML)
2. Akaike Information Criterion (AIC): -2LL + 2P
3. Schwartz Bayesian Information Criterion (BIC): -2LL + P ln(N)

where ML represents the maximum likelihood of the item parameters, P represents the number of model parameters, L represents the total number of attribute patterns, and N represents the sample size. To choose the best fit model relative fit statistics were applied and the relative fit indices of GDINA, DINA, DINO, ACDM, C-RUM and R-RUM were compared. Regarding the results GDINA was decided to be used to analyse the data. The model with the lowest value for each statistic will be preferred over competing models (Chen, de La Torre & Zhang, 2013).

3.5.4.4. Absolute Fit Statistics. M_2, $RMSEA_2$ (the root means square error of approximation fit index for M_2) with 90% CI (confidence interval), SRMSR (the standardized root mean squared residual) and proportion correct are provided for absolute fit statistics. Therefore, details of absolute fit statistics for GDINA model are presented below in detail.

M_2 is a sensitive value that is appropriate to identify model misspecifications and model-data fit (Chen, Liu, Xin, & Cui, 2018; Henson et al., 2009). Some researchers (Maydeu-Olivares & Joe, 2014) suggest using $RMSEA_2$ instead of M_2 to evaluate the accuracy of approximation for CDMs and to define the level of model error. $RMSEA_2$ is an indicator that resembles effect sizes and helps to compare different models (Chen et al., 2018). The $RMSEA_2$ scale runs from 0 to 1. According to Hooper et al (2008), the values less than .06 represents a good fit for $RMSEA_2$. On the other hand, Liu et al. (2016) states that the values less than 0.05 indicate good fit and the values less than 0.03 indicate excellence fit. Also, non-significant value of p which means $p > 0.05$ shows a good fit (Ravand, 2016). SRMSR is a measure of the degree of the mean of the standardized residuals between the expected and observed covariance matrices (Chen, 2007). SRMSR scores between 0.00 and 0.08 are considered as acceptable (Hu & Bentler, 1999).

3.5.4.5. Accuracy and Consistency. Classification accuracy (P_a) and classification consistency (P_c) values were calculated to refer to the validity and reliability of the classification of the students into the latent classes (Ravand, 2016). Classification consistency measures how consistently a student is placed in the same latent class or how clearly, s/he will be classified as a master or non-master of the same attribute, when the test is administered again (re-test) using the same or a similar format (Ravand & Robitzsch, 2018). Additionally, classification accuracy measures how closely his/her classification corresponds to his/her actual latent class or how clearly s/he is identified as a master or non-master of any given attribute (Ravand & Robitzsch, 2018).

In the literature, there are no clear thresholds for classification accuracy and classification consistency values (Cui, Gierl, & Chang, 2012). However, Cui et al. (2012) indicates the accuracy (0.68), and consistency (0.52) values of Tatsuoka (2002) are acceptable. On the other hand, Ravand and Robitzsch (2018) accept the range of accuracy and consistency categorization rates between 0.70 and 0.80.

3.5.4.6. Cross-Validation. Cross-validation is a resampling technique that tests and trains a model on multiple iterations using different portions of the data (Browne, 2000). It is most frequently employed in situations when the aim is prediction, and the researcher wants to assess how well a predictive model will perform in practice (Browne, 2000). In the present study, since most of the participants had chosen the distractor and the results found unexpected, a cross-validation process conducted by eliminating item 1. The results of this part were presented in the results section.

4. RESULTS

In the result section, model fit statistics, item parameters, classification accuracy and consistency values, attribute prevalence, latent class profile and individual level feedbacks are provided.

4.1. Reliability

In the current study, Cronbach's alpha (α) was calculated for internal consistency. According to the results of the reliability test, the Cronbach's alpha was calculated as 0.76. Since the value of the Cronbach alpha is above 0.70, it is acceptable (Cronbach, 1951; DeVellis, & Thorpe, 2021; Tavakol, & Dennick, 2011).

4.2. Model Fit Statistics

Fit indices are calculated, and relative fit indices and absolute fit indices are reported below.

4.2.1. Relative Fit Statistics

Relative fit indices reported to check the model fit and choose the best fit model. Deviance (-2LL), AIC, BIC and SABIC values of the data are reported in Table 4.1.

As it can be compared regarding the values in Table 4.1, deviance, and AIC values of GDINA model were the lowest. Since BIC statistic was not the lowest for GDINA, SABIC values were also checked. Regarding all relative fit indices GDINA model was chosen to analyse the data since it has lowest deviance and AIC with the highest number of parameters. Also, BIC and SABIC indices were relatively close the lowest value. The attribute-level GDINA model converged, with the following estimates: –2 log likelihood (–2LL) = 11126.19, Akaike information criterion (AIC) = 11460.19, Bayesian information criterion (BIC) = 12167.66, sample-size adjusted BIC (SABIC) =11637.58; 167 parameters. Therefore, these fit values supported the use of GDINA model for the current study (see Table 4.1).

Table 4.1. Relative Fit Indices.

CDM	**-2LL**	**AIC**	**BIC**	**SABIC**	**P**
DINA	11384.04	11590.04	12026.38	11698.51	103
GDINA	**11126.19**	**11460.19**	12167.66	11637.58	**167**
ACDM	11238.59	11496.59	12043.08	11633.61	129
C-RUM	11208.10	11466.10	**12012.59**	**11603.12**	129
R-RUM	11222.85	11480.85	12027.34	11617.87	129
DINO	11494.47	11700.47	12136.82	11809.88	103

Note: P is the number of model parameters

4.2.2. Absolute Fit Statistics

M_2, $RMSEA_2$ (the root means square error of approximation fit index for M_2) with 90% CI (confidence interval), SRMSR (the standardized root mean squared residual) and proportion correct are provided for absolute fit statistics.

$RMSEA_2$ value for the presented data is calculated as 0.02 which indicates all models fit the data well. Also, p value found as non-significant (0.70>0.05) which is an indicator for good fit. The absolute fit values of the GDINA model are recorded as M_2= 37.78, $RMSEA_2$ = 0.02 df=43, p=0.70, and SRMSR= 0.04. According to the results of absolute fit statistics, the model had a good fit to the data.

4.3. Classification Accuracy (P_a) and Classification Consistency (P_c)

Classification accuracy (P_a) and classification consistency (P_c) values are calculated to refer to the validity and reliability of the classification of the students into the latent classes (Ravand, 2016). Classification consistency and accuracy values are reported in the Table 4.2 below. Table 4.2 shows that how well the participants are accurately and consistently categorized as masters and non-masters of each attribute. Both accuracy and consistency values could be regarded as relatively high since the values were above 0.68 and 0.52 (Cui et al., 2012). On the other hand, overall values for the test level accuracy and consistency values were calculated as 0.67 and 0.52. They are considered as acceptable (Cui et al., 2012). Regarding all, classification consistency and classification accuracy values of the presented study were acceptable.

Table 4.2. Classification accuracy and consistency.

Attributes	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**
P_a	.87	.87	.88	.91	.93	.93
P_c	.79	.81	.74	.75	.84	.89

4.4. Item Parameters

Guessing and slipping parameters with standard errors and GDINA parameters are calculated and reported below.

4.4.1. Guessing and Slipping Parameters

The mean of guessing parameter was 0.16 (see Table 4.3) which means that a participant had 16.13% chance of giving the correct answer to the questions even if they have not mastered all the needed attributes. Regarding the guessing parameters, all items were below 0.50 which showed a good fit (Ravand, Barati, & Widhiarso, 2013). Item 17

(0.28) had the highest guessing parameter whereas item 16 (0.00) had the lowest guessing parameter value among all items in the test.

The mean of slipping parameter was 0.16 (see Table 4.3) which means that participants had 15.83% of possibility to give the incorrect answer, although they had mastery in all needed attributes. When the slipping parameters are checked (see Table 4.3), it can be said that item 14 (0.62) and item 19 (0.55) have higher values than 0.50. (Ravand, Barati, & Widhiarso, 2013). Additionally, even though item 5 (0.47) and item 20 (0.39) are below 0.50, they are relatively high regarding the rest of the items in the test. On the other hand, item 16 (0.00) and item 3 (0.00) are the items with the lowest slipping parameter. Also rest of the slipping parameters of the items are less than 0.20 which also indicates good fit (de La Torre, Hong, & Deng, 2010).

Low guessing and slipping parameter estimations imply that participants who have mastered the measured attributes can demonstrate these abilities in the test appropriately (Sen & Arican, 2015). As demonstrated in Table 4.3, items 1, 3, 4, 6, 7, 10, 11, 12, 13, 16, and 18 are the ones which provide the best informative data since all had low slipping and guessing parameter estimates. Especially item 16 is the most informative item because it has the lowest slipping (0.00) and the lowest guessing (0.00) parameter estimate. Therefore, for a participant who has mastered attributes required by item 16 (A4 and A5), the probability of giving the wrong answer is 0.01% since slipping parameter estimate is 0.00. Also, for a participant who has not mastered attributes required by item 16 (A4 and A5), the probability of giving the correct answer is again 0.01% since guessing parameter estimate is 0.00. According to Rupp et al. (2010), to have good model-data fit it is important to have low slipping and guessing parameter estimates. Therefore, high slipping and guessing parameters might indicate weak model-data fit. To sum up, regarding the averages of slipping (0.16) and guessing (0.16) parameter estimates for the data, the model-data fit for the presented study might be considered as good (de La Torre, Hong, & Deng, 2010; Ravand, Barati, & Widhiarso, 2013). On the other hand, high slipping parameter estimates may indicate possible misfits for the items 14 and 19 in the data set.

Table 4.3. Item Parameter Estimates.

Items	**Guessing (g)**	**SE [g]**	**Slipping (s)**	**SE [s]**	**Item Discrimination Index (IDI)**	**Problem Type**
1	.19	.06	.12	.07	.69	Routine
2	.20	.07	.04	.04	.75	Routine
3	.16	.07	.00	.02	.84	Routine
4	.16	.03	.02	.04	.82	Non-Routine
5	.07	.03	.47	.10	.46	Non-Routine
6	.13	.06	.05	.02	.82	Routine
7	.16	.06	.06	.02	.78	Routine
8	.20	.07	.05	.02	.74	Routine
9	.25	.05	.16	.07	.59	Non-Routine
10	.20	.04	.10	.07	.70	Non-Routine
11	.16	.06	.03	.02	.81	Routine
12	.15	.09	.03	.03	.82	Routine
13	.10	.10	.03	.04	.87	Routine
14	.20	.03	.62	.11	.18	Non-Routine
15	.23	.03	.16	.09	.61	Non-Routine
16	.00	.08	.00	.04	.99	Routine
17	.28	.09	.15	.07	.57	Routine
18	.11	.05	.14	.05	.75	Routine
19	.12	.02	.55	.09	.34	Non-Routine
20	.14	.03	.39	.19	.47	Non-Routine
Mean	.16		.16		.68	

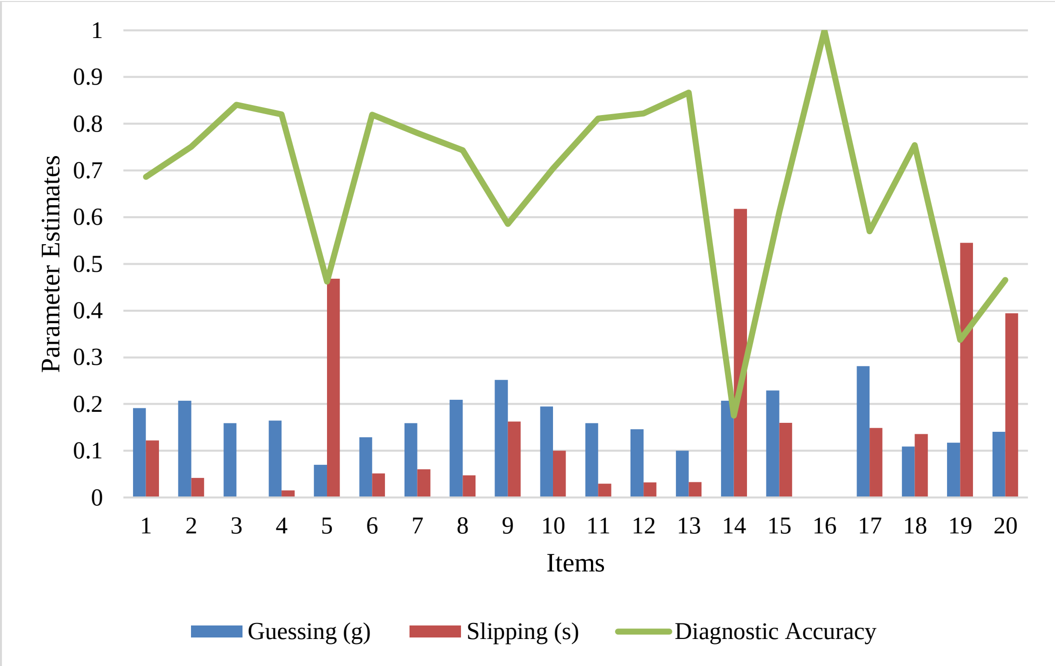


Figure 4.1. Diagnostic Accuracy.

Figure 4.1 shows the diagnostic accuracy for each item. Diagnostic accuracy has a reverse relationship with guessing and slipping parameters. According to the Figure 4.1, item 5, item 14, and item 19 have low diagnostic accuracy. On the other hand, items 1, 3, 4, 6, 7, 11, 12, 13, 16 and 18 have high diagnostic accuracy. Therefore, these items provide more diagnostic information on the learners.

4.4.2. Attribute Combinations

GDINA parameter estimates for the items of the study are shown in Table 4.4. The first column displays the item number, the second one (required attribute) includes the required attributes for the item, the possible patterns of attribute mastery (attribute combination) is shown in the third and fifth column, fourth and sixth columns shows the success probability (p) of the item based on participants' mastery of attribute that is required by the item. The number of estimated item parameters for the study (104) is calculated by the sum of number of parameters for each item in the study. However, number of parameters is not equal for each item. It depends on how many attributes are needed to master the item. All the main effects and interactions are evaluated since GDINA is a saturated cognitive diagnostic model. For instance, items 1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 15, 16, 18, and 19 require two attributes. For these items four item parameters are estimated for each item as on intercept (00), two main effects (10, 01) and one interaction effect (11). Mathematically, since the attribute level of the study is designed as dichotomous level, the number of the item parameters for any item can be calculated by the base of 2. Therefore, the number of the item parameters of each item can be calculated by using “2^a” where a is the number of the required attributes for the item. Accordingly, for items 5, 12, 13, 14, 17, and 20, three attributes are required. Eight parameters (2^3) are calculated for each item as follows, one intercept (000), three main effects (100, 010, 001) and four interaction effects (110, 011, 101, 111).

The intercept parameters display the chance of giving the right answer even if none of the needed attributes are mastered (Ravand, 2016). The intercept parameter measures the probability of “*guessing*” the correct answer without mastering any of the attributes. The main effects demonstrate the change in the likelihood of successfully answering each item when only one of the required attributes is mastered. Interaction effects demonstrate the change in the likelihood of successfully answering each item when more than one attributes are mastered. All item parameter estimates for the study and all intercept, main effects and interaction effects are shown in Table 4.4. Items 1, 3, 12, 14, and 16 are explained below by using the data from Table 4.4. In the Table 4.4, attributes are coded as A1 (addition), A2 (subtraction), A3 (multiplication), A4 (division), A5 (routine problem-solving), A6 (non-routine problem-solving).

Table 4.4. Item Parameter Estimates.

Item Number	**Required attribute**	**Attribute combination**	**p**	**Attribute combination**	**p**
1	A1-A5	P (00)	.19	P (01)	.27
		P (10)	.00	P (11)	.88
2	A1-A5	P (00)	.21	P (01)	.74
		P (10)	.90	P (11)	.96

Table 4.4. Item Parameter Estimates (cont.).

3	A1-A5	P (00)	.16	P (01)	.94
		P (10)	.95	P (11)	.99
4	A1-A6	P (00)	.16	P (01)	.44
		P (10)	.36	P (11)	.98
5	A1-A2-A6	P (000)	.07	P (110)	.27
		P (100)	.00	P (101)	.99
		P (010)	.40	P (011)	.00
		P (001)	.20	P (111)	.53
6	A2-A5	P (00)	.13	P (01)	.54
		P (10)	.61	P (11)	.95
7	A2-A5	P (00)	.16	P (01)	.64
		P (10)	.35	P (11)	.94
8	A2-A5	P (00)	.21	P (01)	.66
		P (10)	.99	P (11)	.95
9	A2-A6	P (00)	.25	P (01)	.00
		P (10)	.53	P (11)	.84
10	A2-A6	P (00)	.20	P (01)	.00
		P (10)	.29	P (11)	.90
11	A3-A5	P (00)	.16	P (01)	.57
		P (10)	.48	P (11)	.97
12	A2-A3-A5	P (000)	.15	P (110)	.46
		P (100)	.00	P (101)	.48
		P (010)	.68	P (011)	.99
		P (001)	.46	P (111)	.97
13	A1-A3-A5	P (000)	.10	P (110)	.0001
		P (100)	.00	P (101)	.37
		P (010)	.73	P (011)	.81
		P (001)	.49	P (111)	.97

Table 4.4. Item Parameter Estimates (cont.).

14	A3-A4-A6	P (000)	.21	P (110)	.21
		P (100)	.00	P (101)	.34
		P (010)	.10	P (011)	.14
		P (001)	.00	P (111)	.38
15	A3-A6	P (00)	.23	P (01)	.10
		P (10)	.09	P (11)	.84
16	A4-A5	P (00)	.00	P (01)	.41
		P (10)	.79	P (11)	.99
17	A1-A4-A5	P (000)	.28	P (110)	.64
		P (100)	.00	P (101)	.48
		P (010)	.25	P (011)	.69
		P (001)	.32	P (111)	.85
18	A4-A5	P (00)	.11	P (01)	.31
		P (10)	.30	P (11)	.86
19	A4-A6	P (00)	.12	P (01)	.50
		P (10)	.11	P (11)	.45
20	A3-A4-A6	P (000)	.14	P (110)	.18
		P (100)	.06	P (101)	.00
		P (010)	.19	P (011)	.00
		P (001)	.79	P (111)	.61

Below, some items with high, middle, and low diagnostic accuracy were chosen and presented (see Figure 4.2, Figure 4.3, Figure 4.4, Figure 4.5, and Figure 4.6). For example, item 1 is accepted as a diagnostically informative item. According to Table 4.4, the intercept effect of item 1 is 0.16. It means that the probability for someone guessing the item 1 correctly is 15.93%. Item 1 requires attributes A1 (addition) and A5 (routine problem-solving) to answer correctly. The main effects of these attributes are .00 and .27. This means that the probability of answering correctly for the participants who have mastered at only addition is very low (0.00%) which means that mastering only A1 may mislead the participants. The probability of answering the item correctly is 26.63% for the participants who have mastered only routine problem-solving skills. The probability of giving the correct answer for a participant who has mastered both addition and routine problem-solving skills is 87.80%. Figure 4.2 shows the distribution of the probabilities along attribute combinations for item 1.

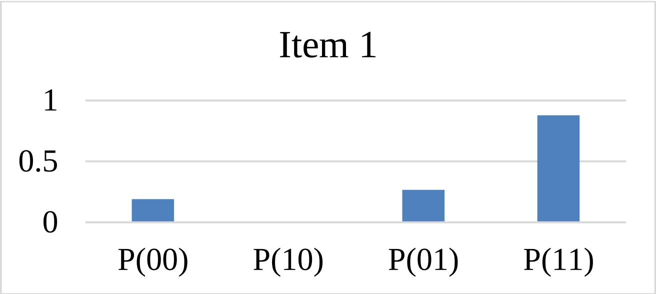


Figure 4.2. Parameter distribution of Item 1.

To exemplify an item that is not diagnostic, item 3 is examined. Item 3 has low slipping (.00) and guessing (.16) parameters. Item 3 requires two attributes as A1 (addition) and A5 (routine problem-solving). The probability of guessing the correct answer to the item without mastering any of the attributes is 15.91%. Giving the correct answer by mastering only addition is 95.41% and mastering only routine problem-solving is 93.88%. The probability of giving the current answer for the participants who have mastered both attributes is 99.99%. It can be stated that mastering at least one of the required attributes increases the probability significantly (see Figure 4.3).

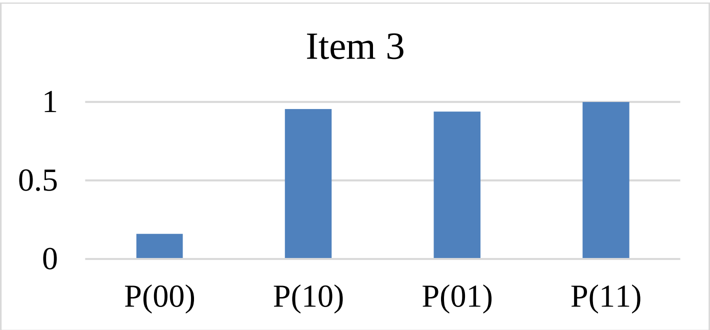


Figure 4.3. Parameter distribution of Item 3.

As another example, item 12 requires three attributes to master as A2 (subtraction), A3 (multiplication), A5 (routine problem-solving). The probability of guessing the correct answer without mastering any of the required attributes is 14.58%. The main effects of each attribute that show the probabilities of answering the item correctly by mastering only one attribute are 0.01% for subtraction, 67.59% for multiplication, 46.38% for routine problem-solving. The probability of answering the item correctly for a participant who have mastered subtraction and multiplication is 46.48%, subtraction and routine problem solving is 48.47% and multiplication and routine problem-solving is 99.99%. Additionally, the probability for the ones who have mastered all three attributes is 96.80%. For item 12, P (010) has higher probability than P (110) and P (101). Also, P (011) has higher probability than P (111) (see Figure 4.4).

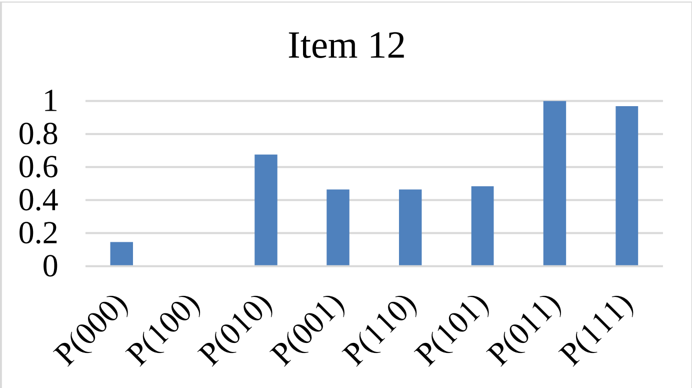


Figure 4.4. Parameter distribution of Item 12.

Item 14 is classified as one of the lowest informative items regarding Figure 3. Item 14 requires three attributes to master as A3 (multiplication), A4 (division) and A6 (non-routine problem-solving). The probability of giving the correct answer by guessing is 20.69%. The main effects are multiplication 0.01%, division 9.79% and non-routine problem-solving 0.01%. interactions for multiplication and division 21.46%, multiplication and non-routine problem-solving 34.37%, and division and non-routine problem-solving 13.98%. Also, for a participant who has mastered all three attributes the probability of answering the item correctly is 38.22%. Figure 4.5 shows the distribution of the probabilities along attribute combinations for item 14.

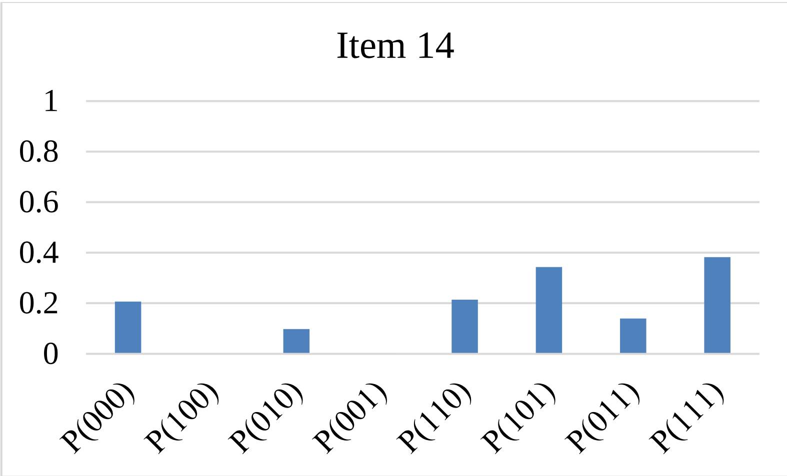


Figure 4.5. Parameter distribution of Item 14.

Item 16 is diagnostically the most informative item based on Figure 4.6. Item 16 requires two attributes which are A3 (multiplication) and A5 (routine problem-solving). The participants who have not mastered any of these two attributes have almost no chance of answering the item correctly (0.01%). On the other hand, the main effects of the items are as .77 and .41. The results show that a participant who has mastered only multiplication has 76.63% probability of answering correctly and a participant who has mastered only routine problem-solving has 41.32% probability to answer right. For anyone who has proficiency in all required attributes, the probability of giving the right answer is 99.99% (see Figure 4.6).

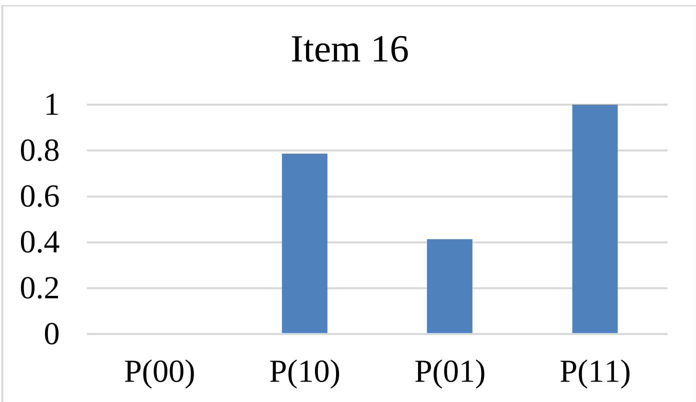


Figure 4.6. Parameter distribution of Item 16.

4.5. Attribute Prevalence

The GDINA model also provides an opportunity to measure attribute prevalence for the sample group. The attribute prevalence is calculated by adding the probability for each latent class that requires the particular attribute. Table 4.5 shows the attribute prevalence for the current study. In Figure 4.7, the frequency of attribute mastery of the six attributes are provided. The attributes in the study are named as A1 (adding), A2 (subtracting), A3 (multiplying), A4 (dividing), A5 (routine problem solving) and A6 (non-routine problem solving). Therefore, while A1, A2, A3, and A4 are related to the four operations and operational skills, A5 and A6 are related to problem-solving skills.

Table 4.5. Attribute Prevalence.

Attributes	A1	A2	A3	A4	A5	A6
Mastery of attribute (Level 1)	.31	.57	.33	.33	.75	.17

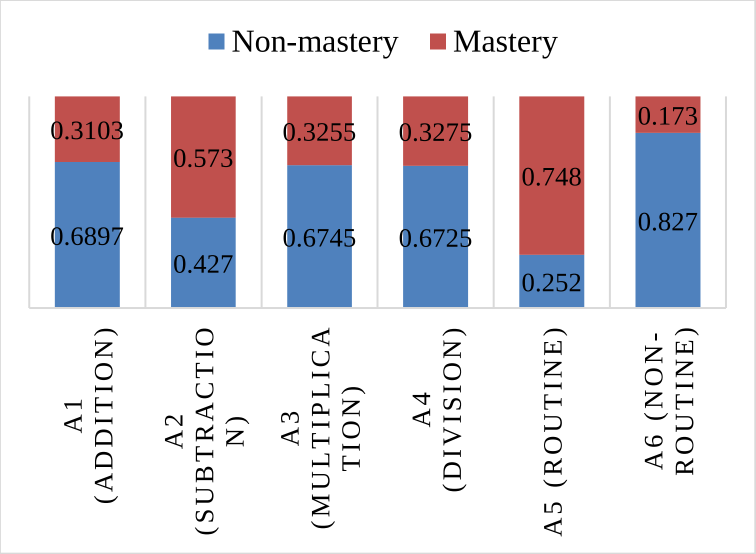


Figure 4.7. Attribute Probabilities.

Among six attributes, the participants have the highest attribute prevalence for A5 (routine problem-solving). On the other hand, A6 (non-routine problem-solving) has the lowest attribute prevalence than any other attribute in the study. Therefore, the results in Table 4.5 indicate that attribute prevalence of the participants is relatively higher for A5. Furthermore, as the attribute prevalence estimates (see Table 4.5) are examined, it shows that A2 (subtracting) is higher than A1 (adding), A3 (multiplying) and A4 (dividing). The results indicate that participants are more likely to have proficiency in subtracting rather than any other operational skill. However, regarding the attribute prevalence for the operational skills of the participants, it can be said that the mastery probability for subtracting skills (.57) highest. Division (.33), multiplication (.33) and addition (.31) skills follow the subtraction. However, while A2 (subtracting) is relatively higher than other operational skills, there is not a high difference between A1 (addition), A3 (multiplication) and A4 (division). These results shows that students have difficulty in mastering operational skills and especially non-routine problem-solving skills. However, they have mastered at routine problem-solving skills. The strongest mathematical operation for the given group is subtraction.

4.6. Attribute Matery Profiles

Attribute class patterns and the attribute probabilities are shown in Table 4.6. The class patterns can be named as latent class profile or attribute mastery profile. There are 64 different latent attribute classes depending on the number of attributes in the study. The percentage of participants for each profile is shown in Table 4.6 along with the 64 profiles. The attribute patterns in the study vary from completely non-mastery profile (000000) to completely mastery profile (111111) and it includes all possible outcome profiles. Since the total of the probabilities for the 64 distinct latent class profiles is equal to one whole (1.00), the probability estimates presented in Table 4.6 are expressed as percentages (Sen & Arican, 2015).

Figure 4.8 shows the distribution of attribute mastery profile probabilities for the largest 21 attribute mastery profiles because rest of the attribute mastery profiles have a probability less than 1%. Since the rest of the attribute mastery profiles were so low or zero, they were not presented below. According to Figure 4.8, it can be said that there is a quite high difference between the probability distributions for the most common two attribute mastery profiles, 000010 and 010010, and all other attribute mastery profiles. Also, after the largest sixth attribute mastery profile, the probability of the rest is below 5%.

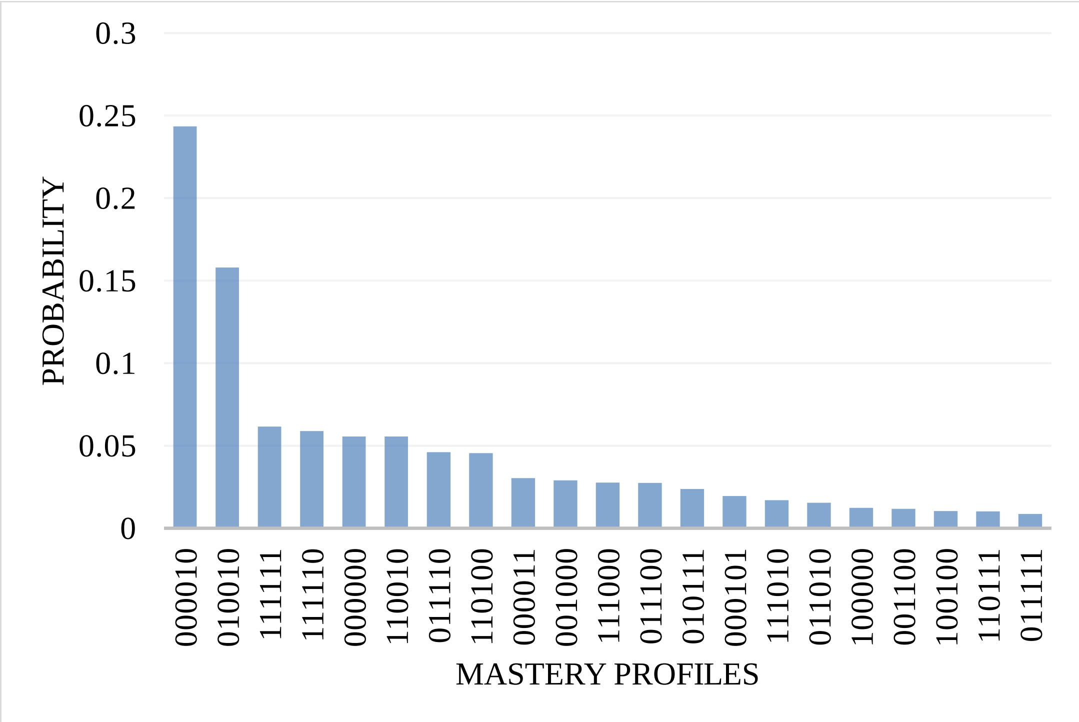


Figure 2.8. Latent Class Probability Distribution.

In the line of Table 4.6, the probability of non-mastery profile (000000) is .06. Therefore, 5.56% of the students have not mastered any of the attributes. 000000 attribute mastery profile is the fifth largest profile. On the other hand, the probability of profile with all mastered attributes is .06. It means that 6.16% of the participant in the research group have mastered all attributes. 111111 attribute mastery profile is the third largest profile. Therefore, it is important to indicate that both non-mastery and all mastery attribute profiles are relatively common for the participants. Additionally, according to the Table 4.6, the most popular attribute class profile is 000010. This profile indicates that participants have highest probability of mastering routine problem solving skills. 24.35% of the participants are matched with mastering only routine problem-solving attribute. However, as it can be seen in the Table 4.6, the probability of having the profile with mastery of non-routine problem-solving attribute only (000001) is equal to 0. According to the data, there is no one who is matched with the 000001-attribute profile.

Table 4.6. Attribute Class Mastery Probabilities.

Class Profile	**Probability**	**Class Profile**	**Probability**
000000	.05	011100	.02
100000	.01	011010	.01
010000	.00	011001	.00
001000	.02	010110	.00
000100	.00	010101	.00
000010	.24	010011	.00
000001	.00	001110	.00
110000	.00	001101	.00
101000	.00	001011	.00
100100	.01	000111	.00
100010	.00	111100	.00
100001	.00	111010	.02
011000	.00	111001	.00
010100	.00	110110	.00
010010	.15	110101	.00
010001	.00	110011	.00
001100	.01	101110	.00
001010	.00	101101	.00
001001	.00	101011	.00
000110	.00	100111	.00

Table 4.6. Attribute Class Mastery Probabilities. (cont.).

000101	.01	011110	.05
000011	.03	011101	.00
111000	.02	011011	.00
110100	.04	010111	.02
110010	.05	001111	.00
110001	.00	111110	.06
101100	.00	111101	.00
101010	.00	111011	.00
101001	.00	110111	.01
100110	.00	101111	.00
100101	.00	011111	.01
100011	.00	111111	.06

The most common attribute profiles among the participants are 000010 (24.35%), 010010 (15.79%), 111111 (6.16%), 111110 (5.89%), 000000 (5.56%), and 110010 (5.56%). Regarding the most common attribute profiles, the attribute patterns of these profiles are explained. For example, 000010 (24.35%) is the most popular attribute profile. A student who is assigned to 000010 profile has mastered only routine problem-solving skills. The next popular attribute profile is 010010 (15.79%). A student who is assigned to 010010 has mastered both subtracting and routine problem-solving attributes. These results are expected based on attribute prevalence estimates in the Table 4.5 since routine problem-solving skills (A5) has the highest attribute prevalence estimate (.75) and it is followed by the subtraction attribute (A2) prevalence estimate (.57). Additionally, the most popular third attribute profile is 111111 (6.16%) which shows the probability of assigned participants to the latent class. Attribute profile 111110 follows the line with 5.89% probability of participants assigned to the class. Attribute profile 111110 means that the participants in that class have mastered all attributes except A6 (non-routine problem-solving). According to the Table 4.6, non-mastered attribute profile (000000) follows closely behind with 5.56%. This probability of non-mastery profile is among the high probability profiles. Furthermore, 110010 attribute profile shares the same probability with the non-mastery profile which is 5.56%. The students who are matched with 110010 attribute profile have mastered addition (A1), subtraction (A2) and routine-problem-solving (A5) attributes.

Based on the most common attribute profile patterns (000010, 010010, 111111, 111110, 000000, 110010) as we mentioned, it is observed that adding (A1), subtracting (A2), and routine problem-solving (A5) attributes have mastered mostly. However, multiplying (A3), dividing (A4), and non-routine problem-solving (A6) attributes have not mastered mostly compared to adding, subtracting and routine problem-solving attributes. Consequently, it can be said that multiplying, dividing and non-routine problem-solving attributes are found relatively difficult to master by the most participants. Especially, non-routine problem-solving attribute (A6) has not been mastered by any of the most frequent attribute profiles except 111111 (6.16%). Similarly, multiplying (A3) and dividing (A4) attributes have not been mastered except attribute profiles 111111 (6.16%) and 111110 (5.89%). The attribute mastery profiles that no one belongs are 000100, 000001, 110000, 101000, 100010, 100001, 011000, 010001, 000110, 110001, 101100, 101010, 101001, 100110, 100101, 011001, 010110, 010101, 001101, 001011, 000111, 111100, 110110, 110101, 110011, 101110, 101101, 100111, 011011, 001111, and 111101.

4.7. Estimated Attribute Profiles of Individuals

The estimates of attribute profiles for each participant with the mastery probability for each attribute are reported below. Table 4.7 shows the estimated attribute profile for the participant. The participants in the table were chosen regarding their laten class, success rate and the number of correct answers to be able to compare their learner profiles. On Table 4.7, the estimates of attribute profiles (EAP), individual mastery probabilities for each attribute, average success percentage of each participant depending on their mastery levels and the number of correct answers in the test are provided. EAP is coded by using attribute mastery probabilities as a base. If the participant has mastered the attribute, then the attribute is coded as “1”, if s/he have not mastered the attribute, it is coded as “0”. The threshold for the mastery level is accepted as 0.50 by GDINA package (Ma & de La Torre, 2022). The participants who have a mastery probability above .5 is coded as “1”, if it is below .5, it is coded as “0”.

Table 4.7. Estimated Attribute Probabilities.

Participant	EAP	A1	A2	A3	A4	A5	A6	Succ. prob	Correct
ID11	1 1 0 1 0 0	.86	.89	.003	.86	.14	.0003	45%	6
ID23	1 1 0 1 0 0	.98	.98	.001	.98	.019	.0004	49%	8
ID35	0 1 0 0 1 0	.09	.50	.0013	.012	.98	.0008	26%	8
ID40	0 0 0 0 1 0	.07	.47	.0033	.012	.98	.041	26%	8
ID53	1 1 1 1 1 0	.73	.98	.93	.89	.99	.004	75%	10
ID72	0 1 1 1 1 1	.04	.94	.87	.86	.99	.63	72%	10
ID74	1 1 1 1 1 0	.94	.99	.95	.960	1.00	.11	82%	14
ID83	1 1 1 1 1 1	.72	.99	.96	.86	.99	.94	92%	15
ID171	1 1 1 1 1 1	.97	1.0	.84	.97	1.00	.57	89%	16
ID192	1 1 1 0 1 1	.99	.99	.97	.001	1.00	.97	82%	16
ID475	1 1 1 1 1 1	1.0	1.0	1.0	1.0	1.0	.99	99%	20

4.7.1. Student-Level Feedback

Table 4.7 provides diagnostic information on each participants’ weaknesses and strengths in the light of six attributes. For example, ID11 has answered six items in the test correctly. s/he is assigned to EAP (1 1 0 1 0 0) since she/ has mastered A1 (86.21%), A2 (89.07%) and A4 (86.34%) but s/he has not mastered A3 (0.03%), A5 (13.67%) and A6 (0.03%). According to Table 4.7 ID11 answered 6 items out of 20 items correctly. Even though the number of the correct answers is below the fifty per cent, ID11 can answer the items that requires A1, A2 and A4. However, ID11 has not mastered A3, A5 and A6. Especially A3 and A6 are the lowest probabilities (see Figure 4.9). Therefore, the weakness of the participant are multiplication, routine, and non-routine problem-solving attributes. Although s/he has quite high probabilities for addition, subtraction, and division, they may also improve.

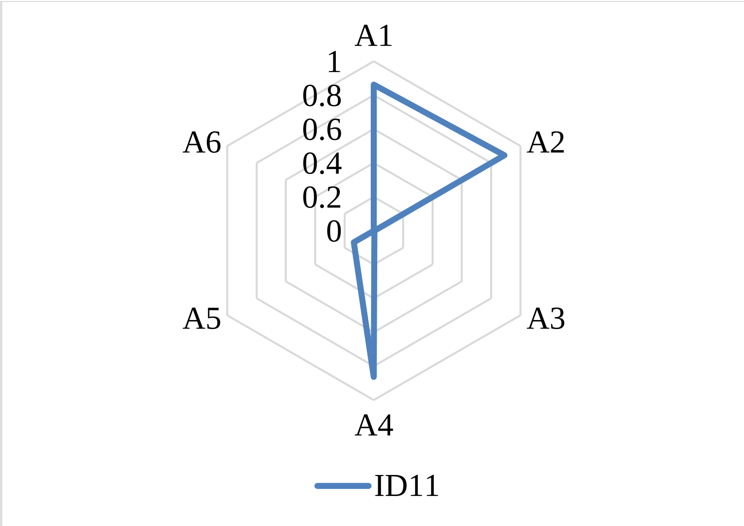


Figure 3.9. Estimated Attribute Probability Distribution of ID11.

As another example, ID192 has answered 16 items correctly. S/he is assigned the EAP (1 1 1 0 1 1) because s/he has mastered A1 (99.05%), A2 (99.98%), A3 (97.74%), A5(100%), A6 (97.74%). However, the mastery probability for A4 is equal to 0.09% which is almost impossible. Even though the general probabilities of the participant high, s/he has a fundamental problem with division (see Figure 4.11). Therefore, the weakness of the participant is division, and the strengths of the student are addition, subtraction, multiplication, routine, and non-routine problem-solving attributes.

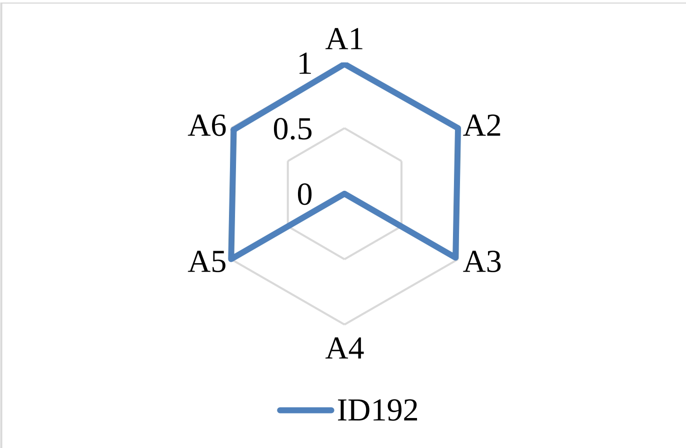


Figure 4.10. Estimated Attribute Probability Distribution of ID192.

Figure 4.11 includes attribute mastery probabilities of two participants who have the same number of the correct answers. Both ID53 and ID72 has answered 10 items out of 20 items correctly. In a standard classroom assessment with equal score distribution per item, these participants would share the same score and be assumed in the same level. However, while ID53 is assigned to EAP (1 1 1 1 1 0) with 75% success probability, ID72 is assigned to EAP (0 1 1 1 1 1) with 72% success probability (see Figure 4.11). When the distributions are analysed, ID53 has mastered A1 (72.98%), A2 (93.83%), A3 (89.18%), A4 (89.18) and A5 (0.04%). On the other hand, ID72 has mastered A2 (94.86%), A3 (94.86%), A4 (87.60%), A5 (99.96%) and A6 (63.46%). Additionally, ID53 has failed to master A6 (0.4%) but ID72 has failed to master A1 (4.3%). To sum up, their common strengths are subtraction, multiplication, division and routine problem-solving. The weakness of ID53 is non-routine problem-solving but the mastery probability of addition is lower than the other mastered attributes. The weakness of ID72 is addition but the mastery probability of non-routine problem-solving attribute is lower than the other mastered attributes.

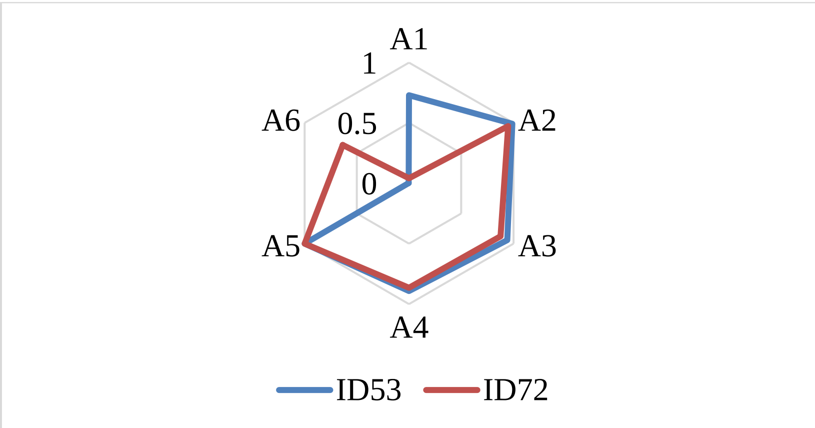


Figure 4.11. Estimated Attribute Probability Distribution of ID53 and ID72.

Figure 4.12 involves attribute mastery probabilities of two participants who are assigned to the attribute profile. Both ID11 and ID23 are assigned to EAP (1 1 0 1 0 0). ID11 has answered 6 and ID23 has answered 8 out of 20 items correctly. Both ID11 and ID23 have mastered A1 (86.21%; 98.16%), A2 (89.07%; 98.52%) and A4 (86.34; 98.20%), respectively. They also both have failed to master A3 (0.3%; 0.1%), A5 (13.67%; 1.86%) and A6 (0.03%; 0.004%), respectively. Thus, the strengths of both ID11 and ID23 are addition, subtraction and division and their weaknesses are multiplication, routine and non-routine problem-solving. Regarding the data in Table 4.7 and the pattern in Figure 4.12, the similarity between two participants who share the same attribute profile can be seen.

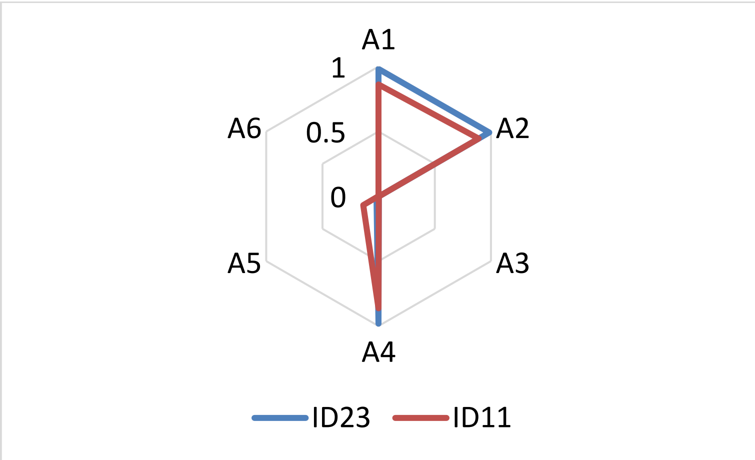


Figure 4.12. Estimated Attribute Probability Distribution of ID11 and ID23.

Figure 4.13 displays that attribute mastery probabilities of three participants who are assigned to the attribute profile. ID83, ID171 and ID475 are assigned to EAP (111111) which is the attribute profile for the participants who have mastered all required attributes in the test. ID83 has answered 15 items correctly with 92% average mastery probability, ID171 has answered 16 items correct with 89% average mastery probability, and ID475 has answered 20 items out of 20 items correctly with 99% average mastery probability. All participants have mastered all required attributes. ID83 has mastered A1 (71.79%), A2 (99.64%), A3 (96.02%), A4 (99.98%), A5 (94.73%) and A6 (94.73%). ID171 has mastered A1 (96.82%), A2 (100%), A3 (84.16%), A4 (97.56%), A5 (100%) and A6 (57.50%). ID475 has mastered A1 (100%), A2 (100%), A3 (100%), A4 (100%), A5 (100%) and A6 (99.98%). Regarding these mastery attribute probabilities, the lowest probability of ID83 is addition (72.79) and ID171 non-routine problem-solving (57.50%) might be improved.

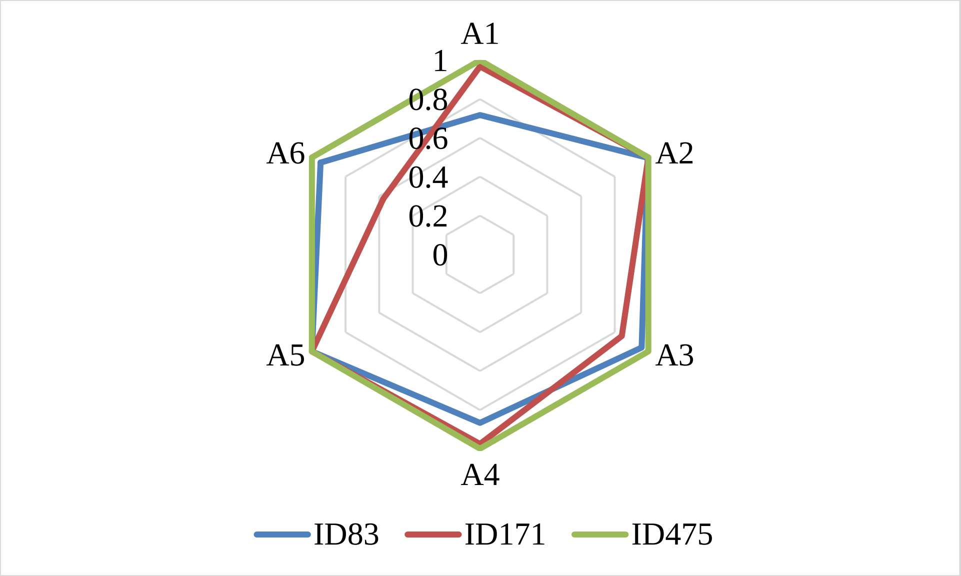


Figure 4.13. Estimated Attribute Probability Distribution of ID83, ID171 and ID475.

4.8. Cross Validation Analysis

The results in the distribution of the attributes created a question mark since addition attribute was lower than expected. It is realized that in Item 1, the students repeated the same mistake. It was realized that most participants selected the same distractor in item 1. Item 1 requires mastery in addition and routine problem-solving attributes. For that reason, it is thought that this item may affect the results. To evaluate this effect, the attribute prevalence is recalculated by excluding item 1. The attribute prevalence values of the data except item 1 is reported in Table 4.8.

Table 4.8. Attribute Prevalence for Cross Validation Analysis.

Attributes	A1	A2	A3	A4	A5	A6
Level 1	0.54	0.48	0.62	0.74	0.19	0.74
Level 2	0.46	0.52	0.38	0.26	0.81	0.26

Level 0 = non mastery, Level 1 = mastery

According to Figure 4.14, the probability of addition attribute is higher than the prevalence with item 1. Eliminating item 1 cause to an increase in the probability of mastering addition. However, still subtraction has the highest attribute prevalence among all operations. On the other hand, routine problem-solving attribute also increased but not as much as addition attribute.

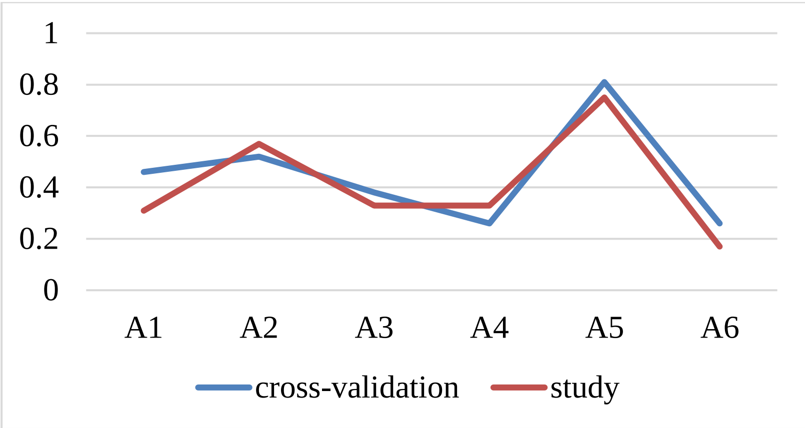


Figure 4.14. Comparison of Attribute Prevalence.

5. DISCUSSION

The purpose of this study is to determine the weaknesses and strengths of fourth-grade students' abilities to solve routine and non-routine problems involving four operations by using CDMs. A CDA test is particularly developed based on *fine-grained attributes* to offer diagnostic and educational feedback to the students. In order to achieve this aim, 2021-2022 fourth grade math curricula were examined, and six attributes are identified as addition, subtraction, multiplication, division, routine problem-solving and non-routine problem-solving. An instrument with 20 items was designed to collect data. First form of the Q-matrix was developed by the researcher and one academic from the measurement and evaluation field. Then, Q-matrix was confirmed by one homeroom and two math teachers to assure the validation of Q-matrix. After the necessary modifications, the form of the Q-matrix was finalized. Data collection and data entrance processes were completed, and data were analysed.

The collected data were analysed with DINO, DINA, ACDM, C-RUM, R-RUM and GDINA to evaluate the model fit to the data. The results of the model-fit tests showed that the GDINA model more closely matches the data than the other models. According to the results of the model fit studies (Basokcu, 2014; Ma, Iaconangelo & de La Torre, 2016), GDINA is a more advantageous model. Basokcu (2014) conducted a study with a similar test design to the presented study including multiple-choice math items. He evaluated the model-fit of different CDMs for different Q-matrices. The results showed that changes in Q-matrix have a smaller effect on latent class classifications for GDINA models (Basokcu, 2014) compared to other models. Ma, Iaconangelo & de La Torre (2016) also investigated how to identify the best fit model, they also come up with a similar conclusion. GDINA is a non-compensatory model which has flexibility for the assumption of equal likelihood of correct responses when learners do not fully master the necessary attributes. This creates a variety for the probability of giving the correct answer, even so students do not have proficiency for all the attributes (Duong Thi & Loye, 2019; Loye, 2010). All in all, the collected data of the study were analysed using GDINA to define and interpret attribute mastery profiles of the students and the diagnostic validity of the test items.

Validity and reliability values of the test were also analysed, and classification accuracy and classification consistency values are calculated to check validity and reliability. Attribute reliability is an important factor in CDA since it contributes to diagnostic score quality (Gierl et al., 2009). There is no definite base value for both classification consistency and accuracy values in the literature (Cui, Gierl, & Chang, 2012). Therefore, the validity and reliability values were checked by comparing the values of other researchers in the literature (Cui et al., 2012; Tatsuoka, 2002; Ravand & Robitzsch, 2018). Attribute-level and test level accuracy (0.67) and consistency values (0.52) were found acceptable (Cui et al., 2012; Tatsuoka, 2002; Ravand & Robitzsch, 2018).

Item parameters were calculated and reported in the results section. In the study, guessing and slipping parameters and GDINA parameters were presented. Item parameters were used to check how informative and diagnostic items are. For the average of guessing parameters all items are in an acceptable rank which is below both 0.5 (Ravand, Barati, & Widhiarso, 2013) and 0.20-0.30 (de La Torre, Hong, & Deng, 2010). It means that it is hard for the participants to give the correct answers, if they have not mastered the required attributes in the test. However, out of 20 items 18 items were acceptable for the slipping parameter. Items 14 and item 19 were relatively high regarding overall slipping estimates in the study. This may indicate that the items were found hard by the participants. The level of the items and the clarity can be checked. Items with low slipping and guessing parameters were regarded as diagnostically informative by Rupp et al. (2010). When the slipping parameters were ranked it can be seen that the highest six items are related to non-routine problems and two of them was not acceptable. When also latent class profile probabilities were checked there is a relationship between the items and latent class profiles. The probability of mastering only non-routine problem-solving attribute is impossible according to analysis of the test. When all mastery profiles that require non-routine problem-solving ability are checked over, the probability of mastery for these profiles is so low except the full mastery attribute profile. The attribute profile who involves the students who have mastered all the attributes is among the most common attribute classes. Therefore, the probability of mastering non-routine problems is higher for the students who have mastered all addition, subtraction, multiplication, division and routine problem-solving skills. According to these results, mastering four operations and routine problem-solving skills may increase the probability of mastering non-routine problem-solving attribute. Additionally, when the GDINA parameters are examined, it can be seen that there is a pattern in the distribution of success probability of the routine problem-solving items except item 1. Except for the non-mastery profile, the success probabilities of the profiles of the routine problem-solving items are quite high compared to the distribution of all non-routine items. This shows that students do not need to master all required attributes for the routine problem-solving items. It is enough for them to master only operational skills or problem-solving skills. The reason for why item 1 is an exception might be a distractor. When the answers of the participants were checked, there is a repeated mistake that led them to the distractor. The success probability distribution of the item supports that claim because it shows that students who have mastered only addition could not answer the item correctly. Even though having only routine problem-solving attribute contribute more to give the correct answer, the best probability belongs to having all required attributes. All in all, the items in the test were considered as diagnostically informative regarding the item parameter values.

Attribute prevalence values were also measured to demonstrate the frequency of attribute mastery. The distribution of the attribute mastery for the study from the highest to the lowest probability as follows; routine problem-solving, subtraction, multiplication, division, addition, and non-routine problem solving. The results of the attribute prevalence shows while the most common attribute is routine problem solving, the least common attribute is non-routine problem-solving. When we combine the data from the item parameters, the attribute prevalence is an expected result. According to these results, mastering non-routine problem is difficult than mastering routine problem-solving and any other attributes in the study. As it is mentioned above, GDINA parameters showed that the probability of being able to solve a routine problem is high, even though the learner has mastered one of the attributes. Therefore, the students who have mastered operational skills have higher chance to be able to solve routine problems rather than non-routine problems. For the non-routine problems students need to have also non-routine problem-solving skills. The second common attribute is subtraction. It is followed by division, multiplication, and addition. However, the probabilities for these three operations are too close. Therefore, addition, multiplication and division attributes were found harder than subtraction attribute to master. Prieto (2016) stated that issues on addition and subtraction with natural numbers may cause many problems since these operations are taught as procedures to be follow with the same kind of examples. Kılıç (2013) also indicated that students are better at subtraction and addition rather than multiplication and division.

According to the results of presented study subtraction has a higher probability of success but addition is quite like other operational skills. The reason of this might be the distractors in the addition problems. The results of the cross over study showed that eliminating the item 1 increases the attribute prevalence of addition and routine problem-solving skills. Regarding the results of attribute prevalence for cross over analysis, the highest attribute is routine problem-solving skills with a quite high probability of success, then subtraction, addition, multiplication, division, and non-routine problem-solving skills. Therefore, the participants of the study are better at routine problem-solving, subtraction and addition rather than multiplication, division, and non-routine problem-solving skills according to the cross over study. It is not surprising that students have better probability of addition and subtraction rather than multiplication and division since they are learning these skills for a longer time and practice them more (Kılıç, 2013; Brandt, Bassoi, & Baccon, 2016). However, the difference between the probabilities for routine and non-routine problem solving is quite high. However, the difference between the probabilities for routine and non-routine problem solving is quite high. Passolunghi and Pazzaglia (2005) and Tertemiz (2017) also indicated that students' success is getting lower, when researchers change the way, they ask the questions. Even though they are asking adding or subtracting questions, students show lower success when they see a question out of the box. One of the reasons might be question types they regularly solve in the classroom environment (Kılıç, 2013) and also the perception of the teachers toward mathematical problem-solving (Stoyanova, 2003).

When attribute mastery profiles of the participants were analysed, the most frequent latent classes can be identified. Using the analysis of the attribute mastery profiles, classroom level diagnostic feedbacks can provide to the teachers and educators. In this study, the most common latent class is 000010 which includes the participants who have mastered only routine problem-solving skills. It is followed by 010010 which involves the participants who have mastered subtraction and routine problem-solving skills. These two results are expected since they match with also attribute prevalence results. According to the Kılıç (2013), students tend to pose questions like simple numeric expressions or simple problems, and it reflects the perspective of the students to the math problems. Since drill and practice method is a common method for the instruction it affects students' problem-solving skills also. There is a huge difference between the routine and non-routine problem-solving attributes. The attribute profiles require routine problem-solving skills contain larger percentage of the participants. It means that mastering routine problem-solving skills is more common among participants rather than any other attributes. On the other hand, those which include non-routine problem-solving attribute are not that popular except the profile with mastery of all attributes. Stoyanova (2003) indicated that problem-solving skills of the students are related to the problem-solving skills of the teacher and the type of classroom works that they complete in the class while learning. Therefore, the classroom instruction should be improved and enhanced with new educational approaches in order to increase students' non-routine problem-solving skills. The following largest skill profiles are students who have mastered all attributes (111111), who have mastered all except non-routine problem-solving attribute (111110), who have mastered none (000000) and who have mastered addition, subtraction and routine problem-solving (110010). Therefore, the largest part of the class has mastered routine problems or subtraction. Rest of the students mostly belong to 111111, 111110 or 110010. Also, while it is possible for the participant to master only routine problem-solving or subtraction attributes, the possibility of mastering only non-routine problems is found as zero which is impossible. It means that students who have not mastered all four operations and routine problem-solving have difficulty to be able to master non-routine problem-solving. Also, there are latent classes with no assigned student in the test. The reason might be the high difference between the probabilities mastering the attributes. Since there is a high difference between routine and non-routine problem-solving attributes, it may limit the variety. Also, sample size of the study may be the reason. With a larger sample size, the attribute master profile of the students may differentiate more.

One of the most significant purposes of the study is to provide diagnostic feedback for each learner based on the cognitive diagnostic assessment test which is developed for this research. As a result of the analysis, each participant is given an attribute profile showing which skills they have mastered and which they have not. In addition, the analysis shows the percentage of participants who mastered a certain skill or attribute, as well as the percentage of participants in each latent class (de La Torre, 2019). Summative classroom assessments or large-scale exams are mostly designed to provide one single score to the participants to show their success or conceptual knowledge (de Ayala, 2009; Hambleton, Yao & Boughton, 2007). These one-dimensional summative tests, rank or categorize the students according to one total score that depends on number of the correct answers (Wang, 2009) instead of providing in-depth feedback to students to improve and enhance their learning (Choi, 2010; de La Torre & Karelitz, 2009). Therefore, the study may help teachers and students to diagnose weaknesses and strengths of the individuals on the domain. In that way, teachers may help students to improve their learning by arranging the instruction. Also, students can change their studying habits based on these diagnostic feedbacks.

To show the benefits of the cognitive diagnostic assessment, individual assessments were examples of two students who share the same score were compared and two students who share the same latent class were compared. In a regular test ID53 and ID72 may be evaluated in the same level, share the same grade in their report card and considered equally successful since they have the same number of correct answers. However, cognitive diagnostic assessments provide to see the differences in individuals’ learning, the weaknesses they can improve and also the strengths that they have mastered. According to the results, ID53 and ID72 has quite different attribute profiles. While ID53 found non-routine problems difficult to master, ID72 had difficulty to master addition. Since they answer half of the items correctly, in a standardized test they may found average. However, according to the results both students have mastered subtraction, multiplication, division, and non-routine problem solving. ID53 needs to practice non-routine problems and also s/he can improve also addition percentage. ID72, on the other hand, needs to study mainly addition and s/he can also improve non-routine problem-solving skills. As it can be seen they have different strengths and weaknesses with different probability of success. The feedback for these individuals needs to be differentiated and detailed regarding these variety.

As another example, ID11 and ID23 share the same latent class but they have different number of correct answers. However, their learner profiles seem quite similar. They both have mastered addition, subtraction, and multiplication. However, they found difficult to master division, routine, and non-routine problem-solving attributes. Even though they have small differences considering the probabilities of mastering attributes, their weaknesses and strengths are common. Therefore, cognitive diagnostic assessments provide both class level and individual level diagnostic feedbacks that might be helpful for both educators and learners. As last, the results of ID83, ID171 and ID475 are evaluated. These students have different number of correct answers, but all are assigned to the profile 111111 which means that they have mastered all attributes. Interestingly, students do not have to answer all items correctly to be able to master all attributes. They can make mistakes and see whether these mistakes have a pattern that shows weaknesses in their learning or not. For example, ID475 has mastered all, and s/he has high probability of success for all attributes. However, ID183 may practice addition and ID171 is suggested to practice non-routine problem-solving. Therefore, even for the participant who have mastered all, diagnostic feedbacks can be beneficial to see the weaknesses and strengths of the students.

5.1. Implications

The current study is designed to identify weaknesses and strengths of the students in problem-solving skills with four operations by using cognitive diagnostic models. The test that is used in the study is designed for the study specifically for cognitive diagnostic purposes. Therefore, the first major feature of the study is using a cognitive diagnostic test and analysing it by using cognitive diagnostic models unlike most of the studies in the literature (Toker & Green, 2012; Sen & Arican, 2015; Ravand, 2016; Dogan & Tatsuoka, 2008; Im & Park, 2010). There are several ways to demonstrate this diagnostic focus in instruction. The study suggests that cognitive diagnostic assessments provide informative and in-depth feedbacks to educators, teachers and students in both class level and individual level. Class level feedbacks might be useful for the educators to improve the curriculum in the case of common or repeated patterns.

The participant of the study did not perform well in mastering non-routine problems compared to any other attributes. Also, they did not perform good at multiplication, division, and addition attributes, as well. The students' ability of problem solving depends on the reflections of the teachers' perception toward problem-solving. Therefore, to increase the probability of non-routine problem solving or other attributes, the classroom instructions and teachers' perceptions should be changed. However, change in the classroom and student level requires different feedbacks. As parallel to that, the strengths and the weaknesses of the latent classes and individuals are examined. The results of latent class profiles may lead the way for the teachers to revise their instructions regarding the general needs of the students. They can easily analyze which attributes or abilities are missing and which ones can be improved. They can also evaluate their instruction, while regarding the distribution of probabilities for the latent classes. In that way, they can easily identify the misconceptions or the strong sides of the students. Based on the individual level results of the study, teacher may differentiate the need of the students effectively. The study showed that sharing the same number of the correct items does not mean that the learners have the same content knowledge or cognitive abilities. The learning profiles of the students are differentiated. Their strong and weak sides are different. Therefore, they cannot be evaluated and graded in the same way. This study offers to give diagnostic and informative feedback in also individual level to improve learners' content knowledge and abilities. Since the needs of each learner is different, the study contributes to the field to corresponds these needs.

All in all, the study was administered in a group of 511 4th grade students with a multiple-choice instrument. However, the results of the study provided qualitative and in-depth information on the learner profile. This implicates that CDAs are convenient to design and analyse with CDMs as also large-scale assessment and provide feedbacks for the curricula, instruction, and the learner profiles.

5.2. Limitations

Finally, the current study has some limitations. It is important to remember that the limitations of the study affect how broadly applicable the results of the study may be.

Firstly, the Q-matrix is designed for the study and confirmed by the educators. The Q-matrix in the study is assumed as correct. Regarding that all analysis are completed. The items of the test are designed as problems that require at least one of the operations. In the test, all items measure at least two attributes. It may be the reason for the routine problem-solving profile has larger probability value. Therefore, some items might be added to study to measure only one operational attribute to create more inductive test structure. Also, the study measures mastering routine and non-routine problem-solving abilities with four operations. However, verbal problems or reading issues may mislead the results of the participants.

The data were collected from the fourth graders in Istanbul. However, to provide generalizability the same test may administer in various cities in Turkey. The number of the students in a class, the perception of the homeroom teacher towards math and problem-solving who teaches in the class may affect the results.

5.3. Suggestions

The results of the research showed that the attribute that the participants had the most difficulty in mastering were non-routine problems. On the other hand, the attribute that participants were most likely to master was routine problems. The probability of the participants to master in addition, multiplication and division attributes is similar, although not very high. However, subtraction skills were found to be the highest in terms of probability of mastering. For this reason, it is necessary to increase the weight given to non-routine questions in classroom studies. Teachers' readiness is very important at this point, as teachers and instruction are very effective factors in the development of problem-solving skills. For this reason, further studies can be developed to provide feedback on the problem-solving skills of in-service teachers or pre-service teacher.

In addition, the presented study focuses on routine and non-routine problems. According to the results, research can be conducted on students' non-routine problem-solving skills based on cognitive diagnostic models. Thus, effective feedback can be provided to the participants and the training. In addition, the presented study draws a perspective on the problem-solving skills of 4th grade students. However, studies can be conducted with multigroup models to understand whether these skills depend on variables such as gender, achievement status or economic status.

As another suggestion, the instrument of the current study might be revised by eliminating or editing the items with high slipping ang guessing parameters. Then, the study might be repeated to check the validity of the data, distribution of the attribute profiles and the attribute prevalence.

REFERENCES

Altun, M., 2001, *Matematik Öğretimi*. Alfa Basım Yayın Dağıtımı, İstanbul.

Arabeyyat, T. H., 2017, *A Comparative Study Examining Preservice Teachers' Beliefs on Mathematical Problem Solving and Their Performances in Solving Real-Life Mathematical Problems in Jordan and the United States*, Ph.D. Thesis, State University of Newyork.

Ardıç, E. Ö., 2020, *Bilişsel Tanı ve Çok Boyutlu Madde Tepki Modellerinin Sınıflama Doğruluğu ve Parametrelerinin Karşılaştırılması*, Ph.D. Thesis, Hacettepe University.

Arikan, S., E. Erktin, & M. Pesen, 2020, "Development And Validation Of A STEM Competencies Assessment Framework", *International Journal Of Science And Mathematics Education*, Vol. 20, No. 3, pp. 1-24.

Ayala, R. J., 2009, *The Theory and Practice of Item Response Theory*, Guilford Publications, New York.

Başokcu, T. O., 2014, "Classification Accuracy Effects Of Q-Matrix Validation And Sample Size In DINA And G-DINA Models", *Journal of Education and Practice*, Vol. 5 No. 6, pp. 220-230.

Black, P., D. Wiliam, 2009, "Developing The Theory Of Formative Assessment", *Educational Assessment, Evaluation and Accountability*, Vol. 21, No. 5, pp. 5-31.

Bolt, D., 2007, "The Present and Future of IRT-Based Cognitive Diagnostic Models (ICDMs) and Related Methods", *Journal of Educational Measurement*, Vol. 44, No. 4, pp. 377–383.

Brandt, C. F., T. S. Bassoi, & A. L. P. Baccon, 2016, "Difficulties Of 6th Grade Elementary School Students in Solving the Four Basic Fundamental Operations: Addition, Subtraction, Multiplication And Division Of Natural Numbers", *Creative Education*, Vol. 7, No. 13, pp. 18-20.

Brown, N. M., 2003, *A Study Of Elementary Teachers' Abilities, Attitudes, And Beliefs About Problem- Solving*. Ph.D. Thesis, Georgia Southern University.

Brown, S., 2004, “Assessment For Learning”, *Learning and Teaching in Higher Education,* Vol. 1, No. 1, pp. 81-89.

Browne, M. W., 2000, “Cross-Validation Methods”, *Journal of Mathematical Psychology*, Vol. 44, No. 1, pp. 108-132.

Chen, J., J. de La Torre, & Z. Zhang, 2013, “Relative and Absolute Fit Evaluation in Cognitive Diagnosis Modelling”, *Journal of Educational Measurement*, Vol. 50, No. 2, pp. 123-140.

Chen, F., Y. Liu, T. Xin, & Y. Cui, 2018, “Applying The M2 Statistic To Evaluate The Fit Of Diagnostic Classification Models In The Presence Of Attribute Hierarchies”, *Frontiers in psychology*, Vol. 9, No. 5, pp.18-75.

Cheng, Y., 2010, “Improving Cognitive Diagnostic Computerized Adaptive Testing by Balancing Attribute Coverage: The Modified Maximum Global Discrimination Index Method”, *Educational and Psychological Measurement*, Vol. 70, No. 6, pp. 902-913.

Cui, Y., M. J. Gierl, & H. H. Chang, 2012, “Estimating Classification Consistency And Accuracy For Cognitive Diagnostic Assessment”, *Journal of Educational Measurement*, Vol. 49, No. 1, pp. 19-38.

Chiu, C. Y., 2013, “Statistical Refinement Of The Q-Matrix In Cognitive Diagnosis”, *Applied Psychological Measurement*, Vol. 37, No. 8, pp. 598-618.

Choi, H. J., 2010, *A Model That Combines Diagnostic Classification Assessment With Mixture Item Response Theory Models,* Ph.D. Thesis, University of Georgia.

Crocker, L., & J. Algina, 1986, *Introduction To Classical And Modern Test Theory,* Holt Rinehart and Winston, Orlando.

Cronbach, L. J., 1951, “Coefficient Alpha And The Internal Structure Of Tests”, *Psychometrika*, Vol. 16, No. 3, pp. 297-334.

De Ayala, R. J., 2009, *The Theory And Practice Of Item Response Theory,* Guilford Press, New York.

Demir, E. K., & N. Koç, 2018, "DINA Model Ile Geliştirilen Bir Testin Psikometrik Özelliklerinin Belirlenmesi", *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, Vol. 18, No. 1, pp. 130-156.

De La Torre, J., 2008, "An Empirically Based Method Of Q-Matrix Validation For The DINA Model: Development And Applications*", Journal of Educational Measurement*, Vol. 45, No. 4, pp. 343–362.

De La Torre, J., 2009a, "DINA Model and Parameter Estimation: A Didactic", *Journal Of Educational and Behavioral Statistics*, Vol. 34, No. 1, pp. 115-130.

De La Torre, J., 2009b, "A Cognitive Diagnosis Model For Cognitively Based Multiple-Choice Options", *Applied Psychological Measurement*, Vol. 33, No. 3, pp. 163-183.

De La Torre, J., 2011, "The Generalized DINA Model Framework", *Psychometrika,* Vol. 76, No. 7, pp. 179-199.

De La Torre, J., 2019, "Implementation Of Cognitive Diagnosis Modeling Using The GDINA R Package", *Eurasian Journal of Educational Research*, Vol. 19, No. 80, pp. 171-192.

De La Torre, J., & C. Y. Chiu, 2016, "A General Method of Empirical Q-matrix Validation" *Psychometrika*, Vol. 81, No. 2, pp. 253–273.

De La Torre, J., J.A. Douglas, 2004, "Higher-Order Latent Trait Models For Cognitive Diagnosis", *Psychometrika,* Vol. 69, No. 1, pp. 333–353

De La Torre, J., Y. Hong, & W. Deng, 2010, "Factors Affecting The Item Parameter Estimation And Classification Accuracy Of The DINA Model", *Journal of Educational Measurement*, Vol. 47, No. 2, pp. 227-249.

De La Torre, J., & N. Minchen, 2014, "Cognitively Diagnostic Assessments and The Cognitive Diagnosis Model Framework", *Psicologia Educativa*, Vol. 20, No. 2, pp. 89–97.

De La Torre, J., & T. M. Karelitz, 2009, "Impact Of Diagnosticity On The Adequacy Of Models For Cognitive Diagnosis Under A Linear Attribute Structure: A Simulation Study", *Journal of Educational Measurement*, Vol. 46, No. 4, pp. 450-469.

De La Torre, J., & T. Rutgers, 2011, "The Generalized Dina Model Framework", *Psychometrika*, Vol. 76, No. 2, pp. 179–199.

Deng, N., & R. K. Hambleton, 2013, "Evaluating CTT-And IRT-Based Single-Administration Estimates Of Classification Consistency And Accuracy", Paper Presented At The *New Developments In Quantitative Psychology*, Millsap, R.E., van der Ark, L.A., Bolt, D.M., Woods, C.M., Vol. 66, pp. 235-250, New York.

DeVellis, R. F., & C. T. Thorpe, 2021, *Scale Development: Theory And Applications*, Sage publications, San Fransisco.

DiBello, L., L. Roussos, & W. Stout, 2007, "Review Of Cognitively Diagnostic Assessment And A Summary Of Psychometric Models", *Handbook of Statistics*, Vol. 26, No. 1, pp. 979-1030.

DiBello, L. V., W. F. Stout, & L. Roussos, 1995, "Unified Cognitive Psychometric Assessment Likelihood-Based Classification Techniques", In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 361-390). Hillsdale, NJ: Lawrence Erlbaum.

Dogan, E., & K. Tatsuoka, 2008, "An International Comparison Using A Diagnostic Testing Model: Turkish Students' Profile Of Mathematical Skills on TIMSS-R", *Educational Studies in Mathematics*, Vol. 68, No. 3, pp. 263-272.

Duong Thi, D., & N. Loye, 2019, "Cognitive Diagnostic Analyses of the Progress in International Reading Literacy Study (PIRLS) 2011 Results", *Mesure et évaluation en éducation*, Vol. 42, No.1, pp. 127-165.

Embretson, S.E., 1983, “Construct Validity: Construct Representation vs. Nomothetic Span”, *Psychological Bulletin*, Vol. 93, No.1, pp.179–197.

Embretson, S. E., 1994, *Application of Cognitive Design Systems to Test Development*. In C. R. Reynolds (Ed.), Cognitive Assessment: A Multidisciplinary Perspective, pp. 107-135, Plenum Press, New York.

Gail, M., 1996, “Problem Solving about Problem Solving: Framing a Research Agenda”, paper is presented at *Proceedings of the Annual National Educational Computing Conference*, Minnesota, Vol. 17, pp. 255-261.

Gao, M., M. D. Miller, R. Liu, 2017, “The Impact of Q-matrix Misspecification and Model Misuse on Classification Accuracy in the Generalized DINA Model”, *Journal of Measurement and Evaluation in Education and Psychology*, Vol. 8, No. 4, pp. 391-403.

George, A. C., A. Robitzsch, T. Kiefer, J. Gross, & A. Uenlue, 2016, “The R Package CDM for Cognitive Diagnosis Models”, *Journal of Statistical Software*, Vol. 74, No. 2, pp. 1-24.

Gierl, M. J., Y. Cui, & J. Zhou, 2009, “Reliability and Attribute-Based Scoring in Cognitive Diagnostic Assessment”, *Journal of Educational Measurement*, Vol. 46, No. 3, pp. 293-313.

Gorin, J. S., 2007, “Test Construction and Diagnostic Testing”. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive Diagnostic Assessment: Theory and Applications*, pp. 173-201. Cambridge University Press, New York.

Gulliksen, H., 1950, *Theory of Mental Tests*, Mcgrawhill Book, New York.

Haagenars, J., & A. McCutcheon, 2002, *Applied Latent Class Analysis*, Cambridge University Press, New York.

Haberman, S., S. Sinharay, & G. Puhan, 2009, “Reporting Subscores for Institutions”, *British Journal of Mathematical and Statistical Psychology*, Vol. 62, No. 1, pp. 79-95.

Haertel, E. H., 1989, "Using Restricted Latent Class Models To Map The Skill Structure Of Achievement Items", *Journal of Educational Measurement,* Vol. 26, No. 1, pp. 333-352.

Hambleton, R. K., H. Swaminathan, & H. J. Rogers, 1991, *Fundamentals of Item Response Theory,* Second Edition, Sage Publications, California.

Huff, K., & D. P. Goodman, 2007, "The Demand For Cognitive Diagnostic Assessment". In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive Diagnostic Assessment: Theory and Applications,* pp. 19-60, Cambridge University Press, New York.

Harlen, W., C. Gipps, P. Broadfoot, & D. Nuttall, 1992, "Assessment and the Improvement of Education", *The Curriculum Journal*, Vol. 3, No. 3, pp. 215–230.

Hartz, S. M., 2002, *A Bayesian Framework For The Unified Model For Assessing Cognitive Abilities: Blending Theory With Practicality,* Ph.D. Thesis, University of Illinois.

Henson, R. A., J. L. Templin, & J. T. Willse, 2009, "Defining a Family Of Cognitive Diagnosis Models Using Log-Linear Models With Latent Variables", *Psychometrika*, Vol. 74, No.1, pp. 191-210.

IBM Corp., 2017, *IBM SPSS Statistics for Windows*, Version 25.0, IBM Corp., New York.

Im, S., & H. J. Park, 2010, "A Comparison Of US And Korean Students' Mathematics Skills Using A Cognitive Diagnostic Testing Method: Linkage To Instruction", *Educational Research and Evaluation*, Vol. 16, No. 3, pp. 287-301.

Fischer, G. H., 1973, "The Linear Logistic Model As An Instrument In Educational Research", *Acta Psychologica,* Vol. 37, No. 1, pp. 359–374.

Junker, B. W., & K. Sijtsma, 2001, "Cognitive Assessment Models With Few Assumptions, And Connections With Nonparametric Item Response Theory", *Applied Psychological Measurement,* Vol. 25, No. 1, pp. 258-272.

Junker, B. W. & K. Sijtsma, 2001, "Cognitive Assessment Models with few Assumptions, and Connections with Nonparametric Item Response Theory", *Applied Psychological Measurement*, Vol. 25, No. 3, pp. 258-272.

Kaplan, R. M., & D. P. Saccuzzo, 1997, *Psychological Testing: Principles, Applications, and Issues,* Fourth Edition, Pacific Grove, California.

Kayan, F., 2007, *A Study On Preservice Elementary Mathematics Teachers' Mathematical Problem Solving Beliefs,* M.S. Thesis, Middle East Technical University.

Kayan, F. & E. Cakiroglu, 2008, "Ilkogretim Matematik Ogretmen Adaylarinin Matematiksel Problem Cozmeye Yonelik Inanclari", *Hacettepe Universitesi Egitim Fakultesi Dergisi*, Vol. 35, No. 1, pp. 218-226.

Ketterlin-Geller, L., & P. Yovanoff, 2019, "Diagnostic Assessments in Mathematics to Support Instructional Decision Making", *Practical Assessment, Research, and Evaluation*, Vol. 14, No. 1, pp. 1-16.

Kılıç, Ç., 2013, "İlköğretim Öğrencilerinin Doğal Sayilarla İlgili Dört İşlem Gerektiren Problem Kurma Etkinliklerindeki Performanslarinin Belirlenmesi", *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, Vol. 20, No. 1, pp. 256-274.

Koyuncu, M. S., 2020, "Bilişsel Tanı Modellerinde Yapısal Eşitlik Modeli ile Q-Matris Doğruluğunun Belirlenmesi", *Angewandte Chemie International Edition*, Vol. 6, No. 11, pp. 951–952.

Lee, Y. S., Y. S. Park, & D. Taylan, 2011, "A Cognitive Diagnostic Modelling of Attribute Mastery In Massachusetts, Minnesota, And The US National Sample Using the TIMSS 2007", *International Journal of Testing*, Vol. 11, No. 2, pp. 144-177.

Leighton, J. P., & M. J. Gierl, 2007, *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge University Press, Cambridge.

Leighton, J. P., & M. J. Gierl, M. J., 2007, "Why Cognitive Diagnostic Assessment?", In Leighton, J., & Gierl, M. (Eds.). *Cognitive Diagnostic Assessment for Education: Theory and Applications*, pp. 3-19, Cambridge University Press, Cambridge.

Lester, F. K., 1994, “Musings about Mathematical Problem-Solving Research: 1970-1994”, *Journal for Research in Mathematics Education*, Vol. 25, No. 6, pp. 660-675.

Li, H., 2011, “A Cognitive Diagnostic Analysis of the MELAB Reading Test”, *Spaan Fellow*, Vol. 9, No, 1, pp. 17-46.

Li, L., X. Zhou, J. Huang, D. Tu, X. Gao, Z. Yang, & M. Li, 2020, “Assessing Kindergarteners’ Mathematics Problem Solving: The Development Of Cognitive Diagnostic Test”, *Studies in Educational Evaluation*, Vol. 66, No. 1, pp. 1-10.

Lin, Q., K. Xing, & Y. S. Park, 2020, “Measuring Skill Growth And Evaluating Change: Unconditional And Conditional Approaches To Latent Growth Cognitive Diagnostic Models”, *Frontiers in Psychology*, Vol. 11, No. 1, pp. 1-13.

Liu, R., A. C. Huggins-Manley, & L. Bradshaw, 2017, “The Impact Of Q-Matrix Designs On Diagnostic Classification Accuracy In The Presence Of Attribute Hierarchies”, *Educational And Psychological Measurement*, Vol. 77, No. 2, pp. 220-240.

London, R., 1993, *A Curriculum of Nonroutine Problems*, paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Loye, N., 2010, “Odyssée Des Modèles De Classification Diagnostique (MCD)”, *Mesure Et Évaluation En Éducation*, Vol. 33, No. 3, pp. 75-98.

Ma, W. & J. de la Torre, J., 2020, “GDINA: An R Package for Cognitive Diagnosis Modeling”, *Journal of Statistical Software*, Vol. 93, No. 14, pp. 1-26.

Ma, W. & J. de la Torre, 2022, “GDINA: The Generalized DINA Model Framework. R Package Version 2.8.8”, retrived from https://CRAN.R-project.org/package=GDINA

Ma, W., C. Iaconangelo, C., & J. de la Torre, 2016, “Model Similarity, Model Selection, And Attribute Classification”, *Applied Psychological Measurement*, Vol. 40, No. 3, pp. 200-217.

Macready, G.B., C. M. Dayton, 1977, "The Use Of Probabilistic Models In The Assessment Of Mastery", *Journal of Educational Statistics,* Vol 2, No. 1., pp. 99–120.

Magno, C., 2009, "Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data", *The International Journal of Educational and Psychological Assessment*, Vol. 1, No. 1, pp. 1–11.

Mason, L., 2003, "High School Students' Beliefs About Maths, Mathematical Problem Solving, And Their Achievement In Maths: A Cross-Sectional Study", *Educational Psychology,* Vol. 23, No. 1, pp. 73-85.

Maydeu-Olivares, A., & H. Joe, 2014, "Assessing Approximate Fit In Categorical Data Analysis", *Multivariate Behavioral Research*, Vol. 49, No. 4, pp. 305-328.

Millî Eğitim Bakanlığı (MEB), 2022, *İlköğretim Kurumları Matematik Dersi Öğretim Program Taslağı,* İstanbul, Turkey.

Mislevy, R. J., 1996, "Test Theory Reconceived", *Journal of Educational Measurement,* Vol. 33., No. 4, pp. 379-416.

Mullis, I. V. S., & M. O. Martin, 2017, *TIMSS 2019 Assessment Frameworks*, http://timssandpirls.bc.edu/timss2019/frameworks/, accessed on June 2, 2022.

National Council of Teachers of Mathematics, 2000, *Principles and Standards for School Mathematics*, NCTM, Reston, VA.

National Research Council, 2001, *Knowing what students know: The Science and Design of Educational Assessment.* Committee on the Foundations of Assessment, Reston, VA.

Nathan, M. J., & K. R. Koedinger, 2000, "An Investigation Of Teachers' Beliefs of Students' Algebra Development", *Cognition and Instruction,* Vol. 18, No. 2, pp. 209-237.

Nichols, P. D., 1994, "A Framework For Developing Cognitively Diagnostic Assessments" *Review of Educational Research*, Vol. 64, No. 4, pp. 575-603.

National Research Council, 2001, *Knowing What Students Know: The Science And Design Of Educational Assessment*, National Academies Press, Reston, VA.

National Research Council, 2003, *Improved Operational Testing and Evaluation and Methods of Combining Test Information for the Stryker Family of Vehicles and Related Army Systems: Phase II Report*, National Academies Press, Reston, VA.

Olkun, S., & Z. T. Uçar, 2009, *İlköğretimde Etkinlik Temelli Matematik Öğretimi,* Eighth Editon, Ayrıntı Publication, Ankara, Turkey.

Passolunghi, M.C., & F. Pazzaglia, 2005, "A Comparison of Updating Processes In Children Good Or Poor In Arithmetic Word Problem Solving", *Learning and Individual Differences,* Vol. 15, No. 1, pp. 257–269.

Parlak, B., 2017, *Investigation Of TIMSS 2015 Mathematics Achievement For Eighth Grade Students With The Cognitive Diagnostic Model*, Ph.D. Thesis, Hacettepe University.

Pellegrino, J, N. Chudowsky, & R. Galser, 2001, *Advances in the Sciences of Thinking and Learning,* Committee on the Foundations of Assessment, National Academy Press, Washington, DC.

Prieto, A. C. S., 2016, "Go One? Lend One? What Does That Mean Exactly?", http://www.planetaeducacao.com.br/portal/artigo.asp?artigo=590, accessed on June 15, 2022.

Polya, G., 1962, *Mathematical Discovery: On Understanding, Teaching, and Learning Problem Solving.* John Wiley, New York.

Polya, G., 1966, *On Teaching Problem Solving the Role of Axioms and Problem Solving in Mathematics,* The Conference Board of the Mathematics Sciences, Washington, DC.

Polya, G., 2004, *How to Solve it: A New Aspect of Mathematical Method*, Princeton university press, Princeton.

Ravand, H., 2016, "Application of a Cognitive Diagnostic Model to a High-Stakes

Reading Comprehension Test”, *Journal of Psychoeducational Assessment*, Vol. 34, No. 8, pp. 782–799.

Ravand, H., H. Barati, & W. Widhiarso, 2013, “Exploring Diagnostic Capacity of a High Stakes Reading Comprehension Test: A Pedagogical Demonstration”, *International Journal of Language Testing*, Vol. 3, No. 1, pp. 11-37.

Ravand, H., & A. Robitzsch, 2015, “Cognitive Diagnostic Modeling Using R”, *Practical Assessment, Research, and Evaluation*, Vol. 20, No. 1, pp. 11-24.

Ravand, H., & A. Robitzsch, 2018, “Cognitive Diagnostic Model of Best Choice: A Study of Reading Comprehension”, *Educational Psychology*, Vol. 38, No. 10, pp. 1255-1277.

R Core Team, 2022, “R: A Language And Environment For Statistical Computing. R Foundation For Statistical Computing”, https://www.R-project.org/, accessed on 2 Feb, 2022.

Reigeluth, C. M., & J. R. Karnopp, 2013, *Reinventing Schools: It’s Time To Break The Mold*, R&L Education, New York.

Resnick, L. B. & D. P. Resnick, 1992, *Assessing The Thinking Curriculum: New Tools For Educational Reform.* Kluwer Press, Norwell, MA.

Robitzsch A, Kiefer T, George AC, Ünlü A, 2022, “CDM: Cognitive Diagnosis Modeling”, R package version 8.1-12, https://CRAN.R-roject.org/package=CDM, accessed on 10 Feb, 2022.

Roussos, L. A., L. V. DiBello, W. Stout, S. M. Hartz, R. A. Henson & J. L. Templin, 2007, “The Fusion Model Skills Diagnosis System”, In Leighton, J., & Gierl, M. (Eds.). *Cognitive diagnostic assessment for education: Theory and applications*, pp. 275-318, Cambridge University Press, Cambridge.

Rupp, A. A., 2007, “The Answer Is In The Question: A Guide for Describing and Investigating the Conceptual Foundations and Statistical Properties of Cognitive Psychometric Models”, *International Journal of Testing*, Vol. 7, No. 2, pp. 95- 125.

Rupp, A. A. & R. J. Mislevy, R. J., 2007, *Cognitive Psychology as it Applies to Diagnostic Assessment. Theory and Practice,* Cambridge University Press, Cambridge, UK.

Rupp, A. A. & J. Templin, J., 2008a, "The Effects of Q-Matrix Misspecification on Parameter Estimates and Classification Accuracy in the DINA Model", *Educational and Psychological Measurement,* Vol. 68, No. 1, pp. 78-96.

Rupp, A. A., & J. L. Templin, 2008b, "Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review Of The Current State-Of-The-Art", *Measurement: Interdisciplinary Research and Perspectives,* Vol. 6, No. 4, pp. 219-262.

Rupp, A. A., J. Templin, & R. Henson, 2010, *Diagnostic Measurement: Theory, Methods,and Applications*, Guilford Press, New York, US.

Shafipoor, M., H. Ravand, & P. Maftoon, 2021, "Test-Level and Item-Level Model Fit Comparison of General vs. Specific Diagnostic Classification Models: A Case of True DCM", *Language Testing in Asia*, Vol. 11, No. 1, pp. 1-20.

Santos-Trigo, M., 1998, "Instructional Qualities of a Successful Mathematical Problem-Solving Class", *International Journal of Mathematical Education in Science and Technology*, Vol. 29, No. 5, pp. 631-646.

Saygılı, S., 2017, "Examining the Problem-Solving Skills and the Strategies Used by High School Students in Solving Non-Routine Problems", *E-International Journal of Educational Research*, Vol. 8, No. 2, pp. 91-114.

Schmidt, M. E., & P. A. Brosnan, 1996, "Mathematics Assessment: Practices and Reporting Methods", *School Science and Mathematics,* Vol. 96, No. 1, pp. 17-20.

Schroeder, T. L., & F. K. Lester, 1989, "Developing Understanding in Mathematics via Problem Solving", *New Directions for Elementary School Mathematics,* Vol. 31, No. 3, pp. 31-42.

Sen, S. & M. Arican, 2015, "A Diagnostic Comparison of Turkish And Korean Students' Mathematics Performances On The TIMSS 2011 Assessment", *Journal of*

Measurement and Evaluation in Education and Psychology, Vol. *6, No.* 2, pp. 238-253.

Sen, S., & A. S. Cohen, 2021, “Sample Size Requirements For Applying Diagnostic Classification Models”, *Frontiers in Psychology*, Vol. 11, No. 6, pp. 1-16.

Shute, V. J., & S. Rahimi, 2017, “Review of Computer-Based Assessment for Learning in Elementary and Secondary Education”, *Journal of Computer Assisted Learning*, Vol. 33, No. 1, pp. 1-19.

Shute, V. J., & D. Zapata-Rivera, 2010, *Educational Measurement and Intelligent System of International Encyclopedia of Education.* Elsevier Publishers, Oxford, UK.

Sinharay, S., R. Almond, & D. Yan, 2004, “Assessing Fit of Models with Discrete Proficiency Variables in Educational Assessment”, *ETS Research Report Series*, Vol. 2004, No. 1, pp. 1-49.

Sinharay, S., S. Haberman, & G. Puhan, 2007, “Subscores Based on Classical Test Theory: To Report or not to Report”, *Educational Measurement: Issues and Practice*, Vol. 26, No. 4, pp. 21-28.

Shinharay, S., G. Puhan, & S. J. Haberman, 2010, “Reporting Diagnostic Scores in Educational Testing: Temptations, Pitfalls, and Some Solutions”, *Multivariate Behavioral Research*, Vol. 45, No. 3, pp. 553-573.

Steele, D. F., & T. F. Widman, 1997, “Practitioner's Research: A Study in Changing Preservice Teachers’ Conceptions about Mathematics and Mathematics Teaching and Learning”, *School Science and Mathematics*, Vol. 97, No. 4, pp. 184-191.

Stiggins, R. J., 2002, “Assessment Crisis: The Absence of Assessment for Learning”, *Phi Delta Kappan*, Vol. 83, No. 10, pp. 758-765.

Stoyanova, E., 2003, “Extending Students' Understanding of Mathematics via Problem-Posing”, *Australian Mathematics Teacher*, Vol. 59, No. 2, pp. 32-40.

Schoenfeld, A. H., 2016, “Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics”, *Journal of education,* Vol. 196, No. 2, pp. 1-38.

Sun, J., T. Xin, S. Zhang, & J. De La Torre, 2013, “A Polytomous Extension of the Generalized Distance Discriminating Method”, *Article Applied Psychological Measurement,* Vol. 37, No. 7, pp. 503–521.

Su, Y. L., K. M. Choi, W. C. Lee, T. Choi, & M. McAninch, 2013, “Hierarchical Cognitive Diagnostic Analysis for TIMSS 2003 Mathematics”, *Centre for Advanced Studies in Measurement and Assessment*,Vol. 35, No. 1, pp. 1-71.

Tatsuoka, K. K., 1983, “Rule-Space. An Approach For Dealing With Misconceptions Based On Item Response Theory”, *Journal of Educational Measurement*, Vol. 20, No. 1, pp. 345-354.

Tatsuoka, K. K., 1985, “A Probabilistic Model For Diagnosing Misconceptions By The Pattern Classification Approach”, *Journal of Educational and Behavioral Statistics* Vol. 10, No. 1, pp. 55-73.

Tatsuoka, K., 1990, *Toward An Integration Of Item-Response Theory And Cognitive Error Diagnosis*, Lawrence Erlbaum Inc, Champaign, IL.

Tatsuoka, K., 1991, “Boolean Algebra Applied To Determination Of Universal Set Of Knowledge States”, *ETS Research Report Series* ONR-1, Vol. 1991, No. 2, pp. 1-36.

Tatsuoka, C., 2002, “Data Analytic Methods For Latent Partially Ordered Classification Models”, *Journal of the Royal Statistical Society: Series C (Applied Statistics),* Vol. 51, No. 3, pp. 337-350.

Tavakol, M., & R. Dennick, 2011, “Making sense of Cronbach's alpha”, *International journal of medical education,* Vol. 2, No. 1, pp. 53-55.

Templin, J. L., & R. A. Henson, 2006, “Measurement of Psychological Disorders Using Cognitive Diagnosis Models”, *Psychological Methods,* Vol. 11, No. 3, pp. 287-305.

Tertemiz, N. I., 2017, “İlkokul Öğrencilerinin Dört Işlem Becerisine Dayalı Kurdukları Problemlerin Incelenmesi”, *Türk Eğitim Bilimleri Dergisi*, Vol. 15, No. 1, pp. 1-25.

Toker, T., & K. Green, 2012, “An Application of Cognitive Diagnostic Assessment on TIMMS-2007 8th Grade Mathematics Items”, *paper presented at the Annual Meeting of the American Educational Research Association*, Vancouver, British Columbia, Canada.

Van De Walle, J., 2014, *Teaching Through Problem-Solving-Elementary & Middle School Mathematics: Teaching*, Pearson Education, Boston, MA.

Van de Walle, J. A., K. S. Karp, & J. M. Bay-Williams, 2016, *Elementary and Middle School Mathematics*. Pearson Education, London, UK.

Von Davier, M., 2005, *A General Diagnostic Model Applied To Language Testing Data*, Educational Testing Service, Princeton, US.

Yamaguchi, K., & K. Okada, 2018, “Comparison Among Cognitive Diagnostic Models For The TIMSS 2007 Fourth Grade Mathematics Assessment”, *Plos One*, Vol. 13, No. 2, pp. 1-17.

Wang, Y. C., 2009, *Factor Analytic Models And Cognitive Diagnostic Models: How Comparable Are They? —A Comparison Of R-RUM And Compensatory MIRT Model With Respect To Cognitive Feedback*. Ph.D. Thesis, The University of North Carolina.

Wang, W., L. Song, P. Chen, Y. Meng, & S. Ding, 2015, “Attribute-Level and Pattern-Level Classification Consistency and Accuracy Indices for Cognitive Diagnostic Assessment”, *Journal of Educational Measurement*, Vol. 52, No. 4, pp. 457-476.

Wiggins, G., 1998, *Educative Assessment. Designing Assessments To Inform and Improve Student Performance*. Jossey-Bass Publishers, San Francisco, US.

Xin, Z., C. Lin, L. Zhang, & R. Yan, 2007, “The Performance of Chinese Primary School Students on Realistic Arithmetic Word Problems”, *Educational Psychology in Practice*, Vol. 23, No. 2, pp. 145-159.

Yao, L., & K. A. Boughton, 2007, “A Multidimensional Item Response Modeling Approach For Improving Subscale Proficiency Estimation And Classification”, *Applied Psychological Measurement*, Vol. 31, No. 2, pp. 83-105.

Zhang, W., 2006, Detecting Differential Item Functioning Using the DINA Model, Ph.D. Thesis, The University Of North Carolina.

APPENDIX 1-TEST ITEMS

Sınav Süresi 60 dakikadır.

OKULU:	
OKUL TİPİ:	DEVLET ☐ ÖZEL ☐

1. Bir fırıncı sabah 825 ekmek üretiyor. Öğlen ise sabah çıkarttığı ekmekten 349 daha fazla ekmek üretiyor. Fırıncı bugün toplam kaç ekmek üretmiştir?

A) 698
B) 1174
C) 1650
D) 1999

2. Boş olarak kütlesi 1250 kg gelen kamyonete 480 kg domates ve 825 kg patates yükleniyor. Bu kamyonetin toplam kütlesi kaç kg olmuştur?

A) 1250
B) 1305
C) 2555
D) 2655

3. Karavanla tatile çıkan bir aile tatilin,

• birinci haftasında 2230 km,

• ikinci haftasında 1985 km,

• üçüncü haftasında 2368 km yol gitmiştir.

Buna göre üç haftanın sonunda gidilen yol toplam kaç kilometredir?

A) 2368
B) 4215
C) 4353
D) 6583

4. Ayşegül her gün o günün tarihini yazıyor. Sonra, yanyana yazdığı bu rakamların arasındaki noktayı silerek 4 basamaklı bir sayı oluşturuyor.

Örneğin 20 Ekim için 20.10 yazıyor ve 2010 sayısını elde ediyor.

Bu şekilde Ayşegül'ün bir yıl boyunca her gün yaptığı bu işlemlerden bulabileceği en büyük 4 basamaklı sayı ile en küçük 4 basamaklı sayının toplamı kaç olur?

A) 4002
B) 4106
C) 4113
D) 4124

5. Bir bardağın yüksekliği 230 milimetre (mm). İki bardak iç içe konulduğunda bardakların yüksekliği 400 mm oluyor.

Bardaklar iç içe konulara oluşturulan bir kulenin uzunluğunun 800 mm'den uzun ve 1200 mm'den kısa olduğu biliniyor.

Buna göre kulenin inşası için kullanılan bardak sayısı aşağıdakilerden hangisi olabilir?

A) 4
B) 6
C) 8
D) 10

6. Aylık geliri 5500 lira olan bir ailenin kira masrafı 2000 TL, fatura masrafı 500 TL ve 1000 TL mutfak masrafı vardır. Bu ailenin kalan giderleri için kaç TL si kalacaktır?

A) 1500
B) 2000
C) 3000
D) 9000

7. Yaprak, 662 sayfalık bir kitabın 228 sayfasını okudu. Yaprak, 155 sayfa daha okursa kitabı bitirmesi için okuması gereken kaç sayfası kalır?

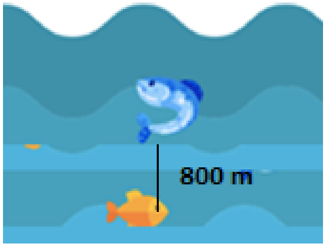
A) 73
B) 279
C) 383
D) 1045

8. Bir manav bir hafta boyunca 847 kg meyve ve sebze satmıştır. Süper market ise manavda yapılan satıştan 275 kg daha az meyve ve sebze satılmıştır. Buna göre, süper markette bir hafta boyunca kaç kg meyve ve sebze satılmıştır?

A) 472
B) 482
C) 572
D) 582

9. Suya dışarıdan bakan bir gözlemci, suyun altında bulunan cisimlerin kendisine olan mesafesini, gerçek mesafeden daha yakın olarak algılar. Gözlemci tarafından algılanan bu mesafe ‘’görünür derinlik’’ olarak isimlendirilir.

Örneğin, gerçekte 20 metre derinlikte bulunan bir balığa dışarıdan bakan bir gözlemci için görünür derinlik 20 metreden daha azdır.



Ali, suya dışarıdan bakmaktadır.

- Sarı küçük balık ve mavi büyük balık arasındaki gerçek mesafe: 800 mm
- Ali için sarı küçük balığın görünür derinliği: 1500 mm

Yukarıda verilen bilgilere göre mavi büyük balığın gerçek derinliği kaç mm olabilir?

B) 560   B) 601   C) 699   D) 701

10. Bir dart oyununda şekildeki gibi dart tahtasının kırmızı, beyaz ve siyah bölgelerine atışlar yapılmaktadır. Aşağıda bu dart oyunun kuralları verilmiştir: Atılan dart oku,

- En içteki daireye gelirse oyuncu 1000 puan alır.
- Beyaz bölgelere gelirse oyuncudan 250 puan silinir.
- Siyah bölgelere gelirse oyuncudan 300 puan silinir.

Daha önceden üç atış yapan Sezgin 1700 puan toplamıştır. İki yeni atış daha yaptıktan sonra Sezgin’in puanı aşağıdakilerden hangisi olabilir?

A) 1250
B) 2450
C) 2950
D) 3000

11. Bir apartmanda toplam 12 daire, her dairede 5 pencere vardır. Aynı özellikteki 7 apartmanda toplam kaç pencere vardır?

A) 35

B) 60

C) 420

D) 520

12. 30 kuruşa alınan kalem 50 kuruşa satılıyor. 75 kalem satışından toplam kaç kuruş kar elde edilir?

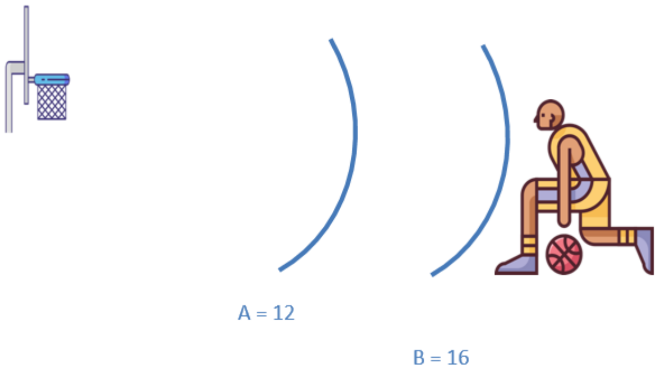
A) 75

B) 100

C) 120

D) 1500

13.



Bir okulda basketbol atış turnuvası düzenleniyor. Puanlar, A ve B noktasının değerleri ile o noktadan yapılan isabetli atış sayısı çarpılarak hesaplanıyor. Deniz A noktasından 23 ve B noktasından 18 isabetli atış yapıyor.

Bu durumda Deniz'in puanı kaç olur ?

A) 528

B) 564

C) 628

D) 668

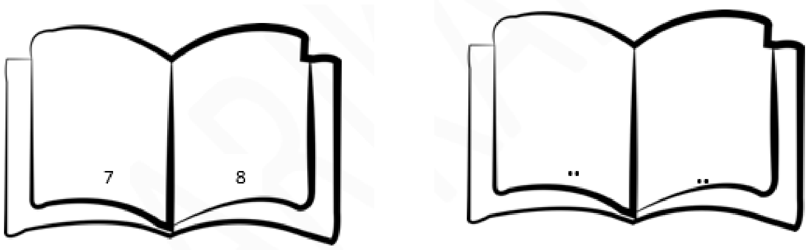
14. Bir market aynı çikolatayı farklı sayılarda paketleyerek bir satıyor. Çikolataların tek tek satışının yasak olduğu biliniyor.

	1 paket fiyatı
10'lu paket	6 TL
6'lı paket	4 TL

Ayşe cüzdanındaki parayla bu çikolatadan en fazla 120 tane alabiliyor. Buna göre Ayşe'nin cüzdanında kaç lirası olabilir?

A) 70
B) 75
C) 80
D) 85

15.



Bir kitabın sayfaları ardışık doğal sayılar kullanılarak 1,2,3,4 … şeklinde yukarıda görüldüğü gibi numaralandırılmıştır. Can kitabı eline alıp rastgele bir sayfayı açtığında sayfa numaralarının çarpımının 380 olduğunu bulmuştur.
Can'ın bir sonraki sayfayı çevirdiğinde bulacağı sonuç kaç olur?

A) 440
B) 460
C) 462
D) 483

16. 810 soruluk bir test kitabını 18 günde bitirmeyi hedefleyen Caner günde kaç soru çözmelidir?

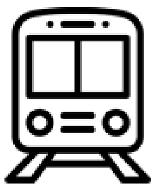
A) 25
B) 45
C) 65
D) 85

17. Şevval apartmanındaki 864 basamağın yarısını 6'şar 6'şar diğer yarısını ise 8'er 8'er çıkmıştır. Buna göre Şevval toplam kaç adım atarak bu basamakları çıkmıştır?

A) 48
B) 72
C) 126
D) 162

18. Bir tren hakkında aşağıdaki bilgiler verilmiştir.

- 2610 yolcu taşınmıştır.
- Yolcular 45'er kişilik vagonlarda seyahat etmiştir.
- Her vagonda eşit sayısa yolcu taşınmıştır.



Buna göre bu trenin kaç vagonu vardır?

A) 54
B) 56
C) 58
D) 60

19. Şehirler arası yük taşıyan bir firmada çalışan Emre Bey kamyoneti ile her biri 12 kg olan kutulardan taşımaktadır. Şirket kurallarına göre bir kamyonet <u>en az</u> 800 kg yük taşımalıdır.

Buna göre Emre Bey'in <u>en az</u> kaç kutu taşıması gerekmektedir?

A) 65
B) 66
C) 67
D) 68

20. Hatice, gittiği bir manavdan her biri 150-gram olan 6 tane elma ve her biri 10 gram olan çileklerden bir miktar alacaktır. Manav; elmaları terazinin sağ kefesine, çilekleri terazinin sol kefesine koyduğunda yandaki gibi bir görüntü elde etmiştir.

Buna göre Hatice <u>en az</u> kaç tane çilek almıştır?

E) 89
F) 90
G) 91
H) 92

APPENDIX 2-INTERVIEW QUESTIONS

1. Anlamakta zorluk çektiğiniz sorular var mıydı? Varsa hangisi ya da hangileriydi?1
2. Testte anlamını bilemediğiniz kelimeler var mıydı?
3. Testin size göre en zor soruları hangileriydi? Neden?
4. Test için size verilen süre yeterli miydi?

APPENDIX 3-CONSENT FORMS

T.C.

BOĞAZİÇİ ÜNİVERSİTESİ

FEN BİLİMLERİ VE MÜHENDİSLİK ALANLARI

İNSAN ARAŞTIRMALARI ETIK KURULU

KATILIMCI BİLGİ ve ONAM FORMU

Araştırmayı destekleyen kurum: Boğaziçi Üniversitesi

Araştırmanın adı: 4. Sınıf Öğrencilerinin Rutin ve Rutin Olmayan Problem Çözme Yeteneklerinin Bilişsel Tanılama Yöntemleri Kullanılarak İncelenmesi

Proje Yürütücüsü/Araştırmacının adı: Züleyha TAŞTAN

Adresi: Boğaziçi Üniversitesi, Kuzey Kampüs, Eta-B Blok, 34342 Bebek, İstanbul

E-mail adresi: zuleyha.tastan@boun.edu.tr

Telefonu: 0546 292 36 97

Sayın veli,

Aşağıda detayları açıklanan araştırmamız Boğaziçi Üniversitesi Matematik ve Fen Bilimleri Bölümü'nde tez araştırması olup Doç. Dr. Serkan Arıkan danışmanlığında yürütülmektedir. Müdürünüz okulun bu çalışmaya katılması için izin verdi. Bu araştırmada bize yardımcı olmanız için öğrencilerimizi de projemize davet ediyoruz. Kararınızdan önce araştırma hakkında sizi bilgilendirmek istiyoruz. Bu bilgileri okuduktan sonra velisi bulunduğunuz öğrencinin araştırmaya katılmasını isterseniz lütfen bu formu imzalayıp kapalı bir zarf içinde bize ulaştırınız.

I. Amaç: Çalışmanın amacı, 4. sınıf öğrencilerinin dört işlem sorularında (toplama, çıkarma, çarpma ve bölme) rutin ve rutin olmayan problemleri çözme becerilerini bilişsel tanılama modellerini kullanarak araştırmaktır. Bu çalışmada, öğrenciler için bir bilişsel tanılama ölçeği geliştirilmiştir. Öğrencilerin cevaplarını göz önünde bulundurarak her öğrencinin güçlü ve zayıf becerileri tanımlanacaktır. Çalışmaya katılmak için 4. sınıf öğrencisi olmak yeterlidir. Öğrencilerimiz geliştirdiğimiz 20 soruluk testi çözerek bu çalışmanın bir parçası olmaya davetlidir.

II. Prosedürler: Öğrencimiz geliştirdiğimiz 20 soruluk testi çözerek bu çalışmanın bir parçası olmaya davetlidir. Proje kapsamında hazırlanan envanterdeki soruların 2022 bahar döneminde yanıtlanması istenecektir. Çalışmaya katılmak için 4. sınıf öğrencisi olmak yeterlidir. Çalışmaya katılmaya karar verirseniz, çocuğunuz 20 soruluk matematik testini sınıfta cevaplandıracaktır. Veri toplanacak, gözden geçirilecek, analiz edilecek ve araştırmada öğrencimizin güçlü ve zayıf yönlerini belirlemek üzere kullanılacaktır. Çalışma boyunca 600 öğrenciden veri toplamayı planlıyoruz. Katılımcıların kimlik bilgileri istenmeyecek ve her türlü kişisel bilgi gizli tutulacaktır. Veri toplama sürecinin başında gönüllü 10 öğrenci ile yarı yapılandırılmış bir görüşme gerçekleştirilecektir. Görüşmede envarterin maddelerine ve zamanlamaya dair sorular sorulacak ve herhangi bir kişisel bilgi kaydedilemeyecektir. Görüşmeler sesli ya da görüntülü olarak kayıt altına alınmayacaktır. Bu çalışma test çözüm süresinin dışında fazladan zaman gerektirmeyecektir. Toplanan veri ileride başka çalışmalar için de kullanılabilir.

III. Riskler: Çalışmanın herhangi bir riski bulunmamakla birlikte normal bir günden daha fazla risk içermemektedir.

IV. Kazanımlar: Çalışma öğrencilerimize kişisel anlamda bir katkı sağlamayacaktır, okul notlarına herhangi bir etkisi bulunmamaktadır. Fakat soru çözerek akademik gelişimlerine küçük bir destekte bulunduğunu söyleyebiliriz. Diğer yandan, çalışmaya sağlanan veriler kullanılarak alan yazına katkı sağlanacaktır.

V. Gönüllü Katılım ve Çekilme: Araştırmaya katılım isteğe bağlıdır. Bu çalışmada olmak zorunda değilsiniz. Çalışmada olmaya karar verirseniz ve fikrinizi değiştirirseniz, istediğiniz zaman vazgeçme hakkınız vardır. Çalışmaya katılmaktan vazgeçmeniz halinde tüm verileriniz imha edilecektir

VI. Gizlilik: Bu araştırma bilimsel bir amaçla yapılmaktadır ve katılımcı bilgilerinin gizliliği esas tutulmaktadır. Kayıtlarınızı araştırmacının izin formlarını yönetmesi, toplaması ve saklaması için izin verilen ölçüde gizli tutulacaktır. Katılımcıların sonuçları ve toplanan tüm verileri dosyalanacaktır. Notlar alındıktan sonra, yalnızca araştırmacılar kimin gönüllü olduğunu belirlemek için rıza formlarını gözden geçirecektir. Araştırma amacıyla yalnızca izin vermiş olan katılımcılardan gelen veriler kullanılacaktır. Tüm tanımlayıcı bilgiler katılımcı verilerinden kaldırılacak; katılımcıların yansımalardaki isimleri kaldırılacak ve araştırmacı tarafından takma adlarla değiştirilecektir. Çalışmanın sonunda, yalnızca araştırmacı verdiğiniz bilgilere erişebilecektir. Veriler, çalışma tamamlandıktan sonra süresiz olarak ileriki tarihlerde araştırmalarda kullanılabilmek adına araştırmacının bilgisayarında depolanabilir. Bilgiler, çalışmanın doğru yapıldığından emin olmak adına veriler Boğaziçi Üniversitesi öğretim üyeleri ile paylaşılabilir. Bu çalışmanın sonuçları araştırma ve eğitim toplulukları ile paylaşılacaktır (konferanslarda, öğretmen mesleki gelişimi ve yayınlarda), ancak hiçbir tanımlayıcı bilgi paylaşılmayacaktır.

VII. İrtibat Kişileri: Bu formu imzalamadan önce, çalışmayla ilgili sorularınız varsa lütfen sorun. Eğer çalışma ile ilgili sorularınız, endişeleriniz veya şikayetleriniz varsa Doç. Dr. Serkan Arıkan (serkan.arikan1@boun.edu.tr) veya Züleyha TAŞTAN (0546 292 36 97-zuleyha.tastan@boun.edu.tr) ile irtibata geçin. Ayrıca, araştırmanın zarar gördüğünü düşünüyorsanız arayabilir, araştırmalar hakkında sorularınız, endişeleriniz, girdi sunmanız, bilgi edinmeniz veya önerileriniz hakkında konuşabilirsiniz. Araştırmayla ilgili katılımcı hakları konusundaki tüm sorularınızı Boğaziçi Üniversitesi Fen Bilimleri ve Mühendislik Alanları İnsan Araştırmaları Etik Kurulu'na (fminarek@boun.edu.tr) danışabilirsiniz.

VIII. Rıza Formunun Konuyu Kopyası: Size saklamak için bu rıza formunun bir kopyasını vereceğiz. Bu araştırma için gönüllü olmaya istekli iseniz, lütfen aşağıdan imzalayın.

Bana anlatılanları ve yukarıda yazılanları anladım. Bu formun bir kopyasını aldım.

Çalışmaya katılmayı kabul ediyorum.

Katılımcının VELİSİNİN	Araştırmacının
Adı-Soyadı	Adı-Soyadı:.....................
İmzası	İmzası
Tarih (gün/ay/yıl):....../..../..........	Tarih (gün/ay/yıl):....../..../..........

T.C.

BOĞAZİÇİ ÜNİVERSİTESİ

FEN BİLİMLERİ VE MÜHENDİSLİK ALANLARI

İNSAN ARAŞTIRMALARI ETIK KURULU

KATILIMCI BİLGİ ve ONAM FORMU

Araştırmayı destekleyen kurum: Boğaziçi Üniversitesi

Araştırmanın adı: 4. Sınıf Öğrencilerinin Rutin ve Rutin Olmayan Problem Çözme Yeteneklerinin Bilişsel Tanılama Yöntemleri Kullanılarak İncelenmesi

Proje Yürütücüsü/Araştırmacının adı: Züleyha TAŞTAN

Adresi: Boğaziçi Üniversitesi, Kuzey Kampüs, Eta-B Blok, 34342 Bebek, İstanbul

E-mail adresi: zuleyha.tastan@boun.edu.tr

Telefonu: 0546 292 36 97

Sayın katılımcı,

Aşağıda detayları açıklanan araştırmamız Boğaziçi Üniversitesi Matematik ve Fen Bilimleri Bölümü'nde tez araştırması olup Doç. Dr. Serkan Arıkan danışmanlığında yürütülmektedir. Müdürünüz okulun bu çalışmaya katılması için izin verdi. Bu araştırmada bize yardımcı olmanız için sizi de projemize davet ediyoruz. Kararınızdan önce araştırma hakkında sizi bilgilendirmek istiyoruz. Bu bilgileri okuduktan sonra araştırmaya katılmak isterseniz lütfen bu formu imzalayıp kapalı bir zarf içinde bize ulaştırınız.

I. Amaç: Çalışmanın amacı, 4. sınıf öğrencilerinin dört işlem sorularında (toplama, çıkarma, çarpma ve bölme) rutin ve rutin olmayan problemleri çözme becerilerini bilişsel tanılama modellerini kullanarak araştırmaktır. Bu çalışmada, öğrenciler için bir bilişsel tanılama ölçeği geliştirilmiştir. Öğrencilerin cevaplarını göz önünde bulundurarak her öğrencinin güçlü ve zayıf becerileri tanımlanacaktır. Çalışmaya katılmak için 4. sınıf öğrencisi olmak yeterlidir. Öğrencilerimiz geliştirdiğimiz 20 soruluk testi çözerek bu çalışmanın bir parçası olmaya davetlidir.

II. Prosedürler: Sizler geliştirdiğimiz 20 soruluk testi çözerek bu çalışmanın bir parçası olmaya davetlidir. Proje kapsamında hazırlanan envanterdeki soruların 2022 bahar döneminde yanıtlanması istenecektir. Çalışmaya katılmak için 4. sınıf öğrencisi olmak yeterlidir. Çalışmaya katılmaya karar verirseniz, 20 soruluk matematik testini sınıfta cevaplandırmanız gerekmektedir. Veri toplanacak, gözden geçirilecek, analiz edilecek ve araştırmada katılımcılarımızın güçlü ve zayıf yönlerini belirlemek üzere kullanılacaktır. Çalışma boyunca 600 öğrenciden veri toplamayı planlıyoruz. Katılımcıların kimlik bilgileri istenmeyecek ve her türlü kişisel bilgi gizli tutulacaktır. Veri toplama sürecinin başında gönüllü 10 öğrenci ile yarı yapılandırılmış bir görüşme gerçekleştirilecektir. Görüşmede envarterin maddelerine ve zamanlamaya dair sorular sorulacak ve herhangi bir kişisel bilgi kaydedilemeyecektir. Görüşmeler sesli ya da görüntülü olarak kayıt altına alınmayacaktır. Bu çalışma test çözüm süresinin dışında fazladan zaman gerektirmeyecektir. Toplanan veri ileride başka çalışmalar için de kullanılabilir.

III. Riskler: Çalışmanın herhangi bir riski bulunmamakla birlikte normal bir günden daha fazla risk içermemektedir.

IV. Kazanımlar: Çalışma sizlere kişisel anlamda bir katkı sağlamayacaktır, okul notlarına herhangi bir etkisi bulunmamaktadır. Fakat soru çözerek akademik gelişimlerine küçük bir destekte bulunduğunu söyleyebiliriz. Diğer yandan, çalışmaya sağlanan veriler kullanılarak alan yazına katkı sağlanacaktır.

V. Gönüllü Katılım ve Çekilme: Araştırmaya katılım isteğe bağlıdır. Bu çalışmada olmak zorunda değilsiniz. Çalışmada olmaya karar verirseniz ve fikrinizi değiştirirseniz, istediğiniz zaman vazgeçme hakkınız vardır. Çalışmaya katılmaktan vazgeçmeniz halinde tüm verileriniz imha edilecektir

VI. Gizlilik: Bu araştırma bilimsel bir amaçla yapılmaktadır ve katılımcı bilgilerinin gizliliği esas tutulmaktadır. Kayıtlarınızı araştırmacının izin formlarını yönetmesi, toplaması ve saklaması için izin verilen ölçüde gizli tutulacaktır. Katılımcıların sonuçları ve toplanan tüm verileri dosyalanacaktır. Notlar alındıktan sonra, yalnızca araştırmacılar kimin gönüllü olduğunu belirlemek için rıza formlarını gözden geçirecektir. Araştırma amacıyla yalnızca izin vermiş olan katılımcılardan gelen veriler kullanılacaktır. Tüm tanımlayıcı bilgiler katılımcı verilerinden kaldırılacak; katılımcıların yansımalardaki isimleri kaldırılacak ve araştırmacı tarafından takma adlarla değiştirilecektir. Çalışmanın sonunda, yalnızca araştırmacı verdiğiniz bilgilere erişebilecektir. Veriler, çalışma tamamlandıktan sonra süresiz olarak ileriki tarihlerde araştırmalarda kullanılabilmek adına araştırmacının bilgisayarında depolanabilir. Bilgiler, çalışmanın doğru yapıldığından emin olmak adına veriler Boğaziçi Üniversitesi öğretim üyeleri ile paylaşılabilir. Bu çalışmanın sonuçları araştırma ve eğitim toplulukları ile paylaşılacaktır (konferanslarda, öğretmen mesleki gelişimi ve yayınlarda), ancak hiçbir tanımlayıcı bilgi paylaşılmayacaktır.

VII. İrtibat Kişileri: Bu formu imzalamadan önce, çalışmayla ilgili sorularınız varsa lütfen sorun. Eğer çalışma ile ilgili sorularınız, endişeleriniz veya şikayetleriniz varsa Doç. Dr. Serkan ARIKAN (serkan.arikan1@boun.edu.tr) veya Züleyha TAŞTAN (0546 292 36 97- zuleyha.tastan@boun.edu.tr) ile irtibata geçin. Ayrıca, araştırmanın zarar gördüğünü düşünüyorsanız arayabilir, araştırmalar hakkında sorularınız, endişeleriniz, girdi sunmanız, bilgi edinmeniz veya önerileriniz hakkında konuşabilirsiniz. Araştırmayla ilgili katılımcı hakları konusundaki tüm sorularınızı Boğaziçi Üniversitesi Fen Bilimleri ve Mühendislik Alanları İnsan Araştırmaları Etik Kurulu'na (fminarek@boun.edu.tr) danışabilirsiniz.

VIII. Rıza Formunun Konuyu Kopyası: Size saklamak için bu rıza formunun bir kopyasını vereceğiz. Bu araştırma için gönüllü olmaya istekli iseniz, lütfen aşağıdan imzalayın.

Bana anlatılanları ve yukarıda yazılanları anladım. Bu formun bir kopyasını aldım.

Çalışmaya katılmayı kabul ediyorum.

Katılımcı Adı-Soyadı:…………………

İmzası: …………………………………

Tarih (gün/ay/yıl):........./.........../.............

Araştırmacının Adı-Soyadı:......................

İmzası:...

Tarih (gün/ay/yıl):...../......./.............

APPENDIX 4-MINISTERY of EDUCATION PERMISSION

44661070/1

İstanbul İl Milli Eğitim Müdürlüğü ANKET ARAŞTIRMA KOMİSYONU DEĞERLENDİRME FORMU

ARAŞTIRMA SAHİBİNİN	
Adı Soyadı	Züleyha TAŞTAN
Kurumu / Üniversitesi	Boğaziçi Üniversitesi
Araştırma Yapılacak İller	İstanbul
Araştırma Yapılacak Eğitim Kurumu ve Kademesi	......
Araştırmanın Konusu	Dördüncü Sınıf Öğrencilerinin Rutin ve Rutin Olmayan Problem Çözme Yeteneklerinin Bilişsel Tanılama Yöntemleri Kullanılarak İncelenmesi
Üniversite / Kurum Onayı	Var
Veri Toplama Araçları	Anket

MEB 21/01/2020 tarih ve 1563890 sayılı 2020/2 Genelge Kapsamında Araştırma, Yarışma ve Sosyal Etkinlik İzinlerinde Dikkat Edilecek Hususlar

Maddeler	Uygun	Uygun Değil	Maddeler	Uygun	Uygun Değil	Maddeler	Uygun	Uygun Değil
2020/2 Genelgenin 1. Maddesi	☐	☐	2020/2 Genelgenin 2. Maddesi	☐	☐	2020/2 Genelgenin 3. Maddesi	☐	☐
2020/2 Genelgenin 4. Maddesi	☐	☐	2020/2 Genelgenin 5. Maddesi	☐	☐	2020/2 Genelgenin 6. Maddesi	☐	☐
2020/2 Genelgenin 7. Maddesi	☐	☐	2020/2 Genelgenin 8. Maddesi	☐	☐	2020/2 Genelgenin 9. Maddesi	☐	☐
2020/2 Genelgenin 10. Maddesi	☐	☐	2020/2 Genelgenin 11. Maddesi	☐	☐	2020/2 Genelgenin 12. Maddesi	☐	☐
2020/2 Genelgenin 13. Maddesi	☐	☐	2020/2 Genelgenin 14. Maddesi	☐	☐	2020/2 Genelgenin 15. Maddesi	☐	☐
2020/2 Genelgenin 16. Maddesi	☐	☐	2020/2 Genelgenin 17. Maddesi	☐	☐	2020/2 Genelgenin 18. Maddesi	☐	☐
2020/2 Genelgenin 19. Maddesi	☐	☐	2020/2 Genelgenin 20. Maddesi	☐	☐	2020/2 Genelgenin 21. Maddesi	☐	☐
2020/2 Genelgenin 22. Maddesi	☐	☐	2020/2 Genelgenin 23. Maddesi	☐	☐	2020/2 Genelgenin 24. Maddesi	☐	☐
2020/2 Genelgenin 25. Maddesi	☐	☐	2020/2 Genelgenin 26. Maddesi	☐	☐	2020/2 Genelgenin 27. Maddesi	☐	☐
2020/2 Genelgenin 28. Maddesi	☐	☐	2020/2 Genelgenin 20. Maddesi	☐	☐	2020/2 Genelgenin 30. Maddesi	☐	☐
2020/2 Genelgenin 31. Maddesi	☐	☐	2020/2 Genelgenin 32. Maddesi	☐	☐	2020/2 Genelgenin 33. Maddesi	☐	☐
2020/2 Genelgenin 34. Maddesi	☐	☐						

KOMİSYON GÖRÜŞÜ

Veri toplama araçlarının eğitim-öğretimi aksatmayacak şekilde gönüllülük esasına dayalı olarak uygulanması ve araştırma sonuç raporunun müdürlüğümüzden izin alınmadan kamuoyuyla paylaşılmaması koşuluyla yürütülmesinde bir sakınca bulunmaktadır ☐/ bulunmamaktadır ☒

Komisyon Kararı	Uygundur	Oybirliğiyle Alınmıştır.

KOMİSYON

Açıklama:	Ünvan-Adı Soyadı-İmza
	(2.03.2022) Üye: Dr. Ercan KAZEL
	(2.03.2022) Üye: Hale GEYİK
	(2.03.2022) Üye: Nur Büşra Kasu

APPENDIX 5-ETHICAL COMMITTEE PERMISSION

Evrak Tarih ve Sayısı: 18.02.2022-53875

T.C.
BOĞAZİÇİ ÜNİVERSİTESİ REKTÖRLÜĞÜ
Fen Bilimleri ve Mühendislik Alanları İnsan Araştırmaları Etik Kurulu
(FMİNAREK)

Sayı : E-84391427-050.01.04-53875 17.02.2022
Konu : 2022/07 Kayıt no'lu başvurunuz hakkında

Sayın Doç. Dr. Serkan ARIKAN
Matematik ve Fen Bilimleri Eğitimi Bölüm Başkanlığı - Öğretim Üyesi

"4. Sınıf Öğrencilerinin Rutin ve Rutin Olmayan Problem Çözme Yeteneklerinin Bilişsel Tanılama Yöntemleri Kullanılarak İncelenmesi – An Analysis of 4th Grade Students' Routine and Non-Routine Problem-Solving Skills Using Cognitive Diagnostic Models" başlıklı projeniz ile Boğaziçi Üniversitesi Fen Bilimleri ve Mühendislik Alanları İnsan Araştırmaları Etik Kurulu (FMİNAREK)'e yaptığınız 2022/07 kayıt numaralı başvuru 07.02.2022 tarihli ve 2022/02 No.lu kurul toplantısında incelenerek etik onay verilmesi uygun bulunmuştur. Bu karar tüm üyelerin toplantıya on-line olarak katılımıyla ve oybirliği ile alınmıştır.

COVID-19 önlemleri nedeniyle üyelerden ıslak imza alınamadığından bu onam mektubu tüm üyeler adına Komisyon Başkanı tarafından e-imzalanmıştır.

Saygılarımızla bilginize sunarız.

Prof. Dr. Tınaz EKİM AŞICI
Başkan

INDEX 1-R CODES IN GDINA AND CDM PACKAGES

#--------GDINA model--------#

- install.packages("GDINA")
- library(GDINA)
- dat <- DATA
- Q <- Q_MATRIX1
- mod1 <- GDINA (DATA, Q_MATRIX1)
- mod1
- summary(mod1)

- CA(mod3, what = "MAP")

- CA(GDINA.obj, what = "MAP")

- modelfit(mod3, CI = 0.9, ItemOnly = FALSE)

- options(max.print=1000000)

- dat <- data
- Q <- Q
- mod1 <- GDINA(dat = dat, Q = Q, model = "GDINA")
- mod1

summary information

- summary(mod3)
- AIC(mod3)
- BIC(mod3)

- logLik(mod3)
- deviance(mod3) # deviance:-2 log-likelihood
- npar(mod3) # number of parameters
- head(indlogLik(mod3)) # individual log-likelihood
- head(indlogPost(mod3)) # individual log-posterior

structural parameters

see ?coef

- coef(mod3) # item probabilities of success for each latent group
- coef(mod3, withSE = TRUE) # item probabilities of success & standard errors
- coef(mod3, what = "delta") # delta parameters
- coef(mod3, what = "delta",withSE=TRUE) # delta parameters
- coef(mod3, what = "gs") # guessing and slip parameters
- coef(mod3, what = "gs",withSE = TRUE) # guessing and slip parameters & standard errors

person parameters

see ?personparm

- personparm(mod3) # EAP estimates of attribute profiles
- personparm(mod3, what = "MAP") # MAP estimates of attribute profiles
- personparm(mod3, what = "MLE") # MLE estimates of attribute profiles

#plot item response functions for item 10

- plot(mod3,item = 1)
- plot(mod3,item = 10,withSE = TRUE)
- # with error bars

#plot mastery probability for individuals 1, 20 and 50

- plot(mod3,what = "mp", person =c(1,20,50))

Use extract function to extract more components

See ?extract

- plot(mod3,what = "mp", person =c(59,181))
- plot(mod3,what = "mp", person =c(56, 158, 188))
- coef(mod3,"lambda", digits=6) mastery probability for each attribute

EAP estimates of attribute profiles

- coef(object, what = c("catprob", "delta", "gs", "itemprob", "LCprob", "rrum", "lambda"), withSE = FALSE, SE.type = 2, digits = 4, ...)

- coef(mod3, what = c("catprob", withSE = FALSE, SE.type = 2, digits = 4))

- coef(mod3, what = c("catprob", "itemprob", "LCprob", "lambda"), withSE = FALSE, SE.type = 2, digits = 4)

S3 method for class 'GDINA'

- coef(object, what = c("catprob", "delta", "gs", "itemprob", "LCprob", "rrum", "lambda"), withSE = FALSE, SE.type = 2, digits = 4, ...)
- coef(mod3, what = c("itemprob"), withSE = TRUE, SE.type = 2, digits = 4)
- coef(mod3, what = c("catprob"), withSE = TRUE, SE.type = 2, digits = 4)
- coef(mod3, what = c("LCprob"), withSE = TRUE, SE.type = 2, digits = 4)
- coef(mod3, what = c("delta"), withSE = TRUE, SE.type = 2, digits = 4)

S3 method for class 'GDINA'

- extract(object, what, SE.type = 2, ...)

S3 method for class 'GDINA'

- personparm(mod3, what = c("EAP", "MAP", "MLE", "mp", "HO"), digits = 4)

EAP estimates of attribute profiles

- personparm(mod3, what = c("mp"), digits = 4)
- personparm(mod3, what = c("EAP"), digits = 4)

#-------- CDM --------#

- install.packages("CDM")
- library(CDM)
- dat <- data
- Q <- Q
- mod3 <- CDM (data, Q, rule =GDINA)
- mod3
- summary(mod3)
- data(DATA, package="CDM")
- data(Q_MATRIX, package="CDM")
- mod4 <- CDM::gdina(DATA, q.matrix=Q_MATRIX), rule="GDINA")

estimate classification reliability

- cdm.est.class.accuracy(mod4)
- d2 <- CDM::gdina(DATA, q.matrix=Q_MATRIX)
- coef(d2)

INDEX 2- EAP

Student ID	Latent Class	A1	A2	A3	A4	A5	A6	Success Percentage
ID1	0 1 1 0 1 0	0.464	0.957	0.599	0.001	1	0.364	56
ID2	0 0 0 0 1 0	0.000	0.018	0.072	0.073	0.929	0.002	18
ID3	0 0 0 0 1 0	0.072	0.479	0	0.03	0.952	0	25
ID4	1 1 0 0 1 0	0.538	0.933	0.115	0.002	0.998	0	43
ID5	0 1 0 0 1 0	0.068	0.967	0.008	0	1	0	34
ID6	0 0 0 0 1 0	0.203	0.413	0.054	0.007	0.989	0	27
ID7	0 1 0 0 1 0	0.068	0.608	0.383	0	0.999	0.008	34
ID8	0 0 0 0 1 0	0.000	0.072	0.012	0.003	0.981	0.023	18
ID9	1 1 1 0 1 0	0.883	0.948	0.762	0.002	1	0.005	59
ID10	1 1 1 0 1 0	0.581	0.657	0.578	0	0.999	0.017	47
ID11	1 1 0 1 0 0	0.862	0.891	0.003	0.863	0.137	0	45
ID12	0 1 0 0 1 0	0.209	0.969	0	0.046	0.954	0	36
ID13	1 1 0 0 1 0	0.669	0.954	0.104	0	1	0.006	45
ID14	0 1 0 0 1 0	0.492	0.791	0.003	0	1	0.002	38
ID15	0 1 0 0 1 0	0.175	0.880	0.054	0.001	0.999	0.066	36
ID16	0 0 1 1 0 0	0.000	0.001	0.649	0.649	0.35	0	27
ID17	1 1 1 1 1 1	0.938	0.971	0.918	0.918	1	0.788	92
ID18	0 1 0 0 1 0	0.399	0.666	0.004	0.001	0.997	0	34
ID19	0 1 0 0 1 0	0.121	0.63	0.408	0	1	0.007	36
ID20	1 1 1 1 1 1	0.884	0.999	0.721	0.980	1	0.784	89
ID21	1 1 0 1 1 1	0.548	0.998	0	0.955	1	0.961	74
ID22	0 0 0 0 1 1	0.006	0.233	0.012	0.012	0.989	0.535	29
ID23	1 1 0 1 0 0	0.982	0.985	0.001	0.982	0.019	0	49
ID24	0 0 0 0 1 0	0.012	0.117	0.015	0.023	0.988	0.009	19
ID25	0 0 0 0 0 0	0.000	0.001	0.341	0.016	0.206	0	9
ID26	1 0 0 0 0 0	0.942	0	0.001	0.05	0.023	0	16
ID27	1 1 0 0 1 0	0.769	0.944	0.038	0	1	0.017	46

ID28	0 1 0 0 1 0	0.312	0.736	0.498	0.335	1	0.059	48
ID29	1 1 1 1 1 0	0.920	0.964	0.867	0.820	1	0.478	84
ID30	0 1 0 0 1 0	0.451	0.691	0.084	0.075	1	0.055	39
ID31	1 1 0 0 1 0	0.589	0.836	0.006	0.024	0.999	0.027	41
ID32	0 1 0 0 1 0	0.068	0.987	0.498	0.001	0.999	0.002	42
ID33	0 1 1 1 1 0	0.235	0.871	0.815	0.682	0.996	0.004	60
ID34	0 0 0 0 1 0	0.073	0.129	0.005	0.006	0.994	0.312	25
ID35	0 1 0 0 1 0	0.099	0.501	0.001	0.012	0.981	0.001	26
ID36	0 1 0 0 1 0	0.471	0.909	0.365	0.074	0.556	0.011	39
ID37	0 0 0 0 1 0	0.000	0.075	0.23	0.062	0.742	0	18
ID38	1 1 0 0 1 0	0.532	0.921	0	0	0.999	0.009	41
ID39	0 1 0 0 1 0	0.333	0.621	0.252	0.052	0.684	0.009	32
ID40	0 0 0 0 1 0	0.072	0.476	0.003	0.012	0.990	0.041	26
ID41	0 1 0 0 1 0	0.106	0.891	0.001	0	0.999	0	33
ID42	0 0 0 0 1 0	0.000	0.016	0.125	0.014	0.84	0.04	17
ID43	0 0 0 0 1 0	0.477	0.371	0.001	0.483	0.517	0.024	31
ID44	1 1 0 0 1 0	0.832	0.970	0.173	0.204	1	0.038	53
ID45	0 1 1 1 1 0	0.264	0.946	0.945	0.911	1	0.004	67
ID46	0 1 0 0 1 0	0.202	0.803	0.002	0.002	0.997	0	33
ID47	1 1 1 1 1 0	0.881	0.997	0.986	0.955	1	0.002	80
ID48	1 1 1 1 1 0	0.642	0.978	0.597	0.781	1	0.208	70
ID49	0 1 1 0 1 0	0.220	0.923	0.735	0.491	0.999	0.007	56
ID50	0 1 0 0 1 0	0.406	0.964	0.128	0	1	0	41
ID51	1 1 0 1 1 1	0.900	0.997	0.048	0.572	1	0.528	67
ID52	0 1 0 0 1 0	0.401	0.554	0.001	0.001	0.999	0	32
ID53	1 1 1 1 1 0	0.730	0.988	0.94	0.892	1	0.004	75
ID54	0 1 0 0 1 0	0.200	0.794	0.164	0.037	0.757	0.001	32
ID55	0 0 0 0 1 0	0.015	0.182	0.006	0.068	0.993	0.057	22
ID56	1 1 1 1 1 1	0.999	1	0.992	0.985	1	0.981	99
ID57	1 1 0 0 1 0	0.671	0.920	0.077	0.029	0.999	0.029	45
ID58	0 0 0 0 1 0	0.176	0.392	0.196	0.017	0.796	0.005	26
ID59	1 1 1 1 1 1	1	1	1	1	1	1	99

ID60	0 1 0 0 1 0	0.004	0.741	0.152	0.333	0.960	0.202	39
ID61	0 1 1 1 0 0	0	0.961	0.797	0.865	0.202	0.07	48
ID62	1 1 1 0 1 0	0.702	0.995	0.690	0.226	1	0.398	66
ID63	1 1 0 1 1 1	0.994	1	0.011	0.974	1	0.970	82
ID64	0 1 0 0 1 0	0.092	0.835	0.005	0.036	0.964	0.063	33
ID65	0 0 0 0 1 0	0.077	0.140	0.076	0.008	0.874	0.137	21
ID66	0 0 0 0 1 0	0.001	0.098	0.229	0.068	0.637	0	17
ID67	0 1 0 0 1 0	0.313	0.818	0.152	0.122	0.718	0.013	35
ID68	0 0 0 0 0 0	0.206	0	0.044	0.013	0.042	0	5
ID69	0 1 0 1 1 0	0.093	0.88	0.28	0.683	0.999	0.416	55
ID70	1 1 0 1 0 0	0.609	0.779	0	0.587	0.414	0.003	39
ID71	0 0 0 0 1 0	0.022	0.464	0.016	0.001	0.999	0.013	25
ID72	0 1 1 1 1 1	0.043	0.949	0.876	0.865	1	0.635	72
ID73	0 1 0 0 1 0	0.440	0.704	0.003	0.002	1	0.001	35
ID74	1 1 1 1 1 0	0.945	0.998	0.954	0.960	1	0.106	82
ID75	1 1 1 1 1 0	0.725	0.893	0.657	0.645	1	0.012	65
ID76	1 1 1 1 1 1	1	1	1	1	1	1	99
ID77	1 1 1 0 0 0	0.855	0.887	0.848	0.001	0.146	0	45
ID78	1 1 0 0 1 0	0.683	0.951	0.101	0	1	0.036	46
ID79	0 1 0 0 1 0	0.005	0.674	0.171	0.175	0.983	0.085	34
ID80	1 1 0 1 0 0	0.867	0.801	0.001	0.868	0.132	0	44
ID81	1 1 0 1 0 0	0.821	0.928	0.007	0.819	0.182	0	45
ID82	1 1 1 1 1 1	0.980	1	1	1	1	0.999	99
ID83	1 1 1 1 1 1	0.718	0.996	0.96	0.867	1	0.947	91
ID84	0 1 0 0 0 0	0.020	0.912	0.103	0.089	0.008	0	18
ID85	0 1 0 0 1 0	0.169	0.795	0.174	0.004	0.83	0.001	32
ID86	0 0 0 0 1 0	0.091	0.156	0.001	0.001	0.988	0.021	20
ID87	0 0 0 0 1 0	0.036	0.470	0.005	0.008	0.997	0.016	25
ID88	0 0 0 0 0 0	0.000	0	0.047	0.216	0.001	0.216	8
ID89	0 0 0 0 1 0	0.200	0.275	0.351	0.063	0.998	0.014	31
ID90	0 1 0 0 1 0	0.102	0.532	0.014	0.057	0.999	0.046	29
ID91	1 1 0 0 0 0	0.629	0.986	0.491	0.099	0.409	0.004	43

ID92	0 1 0 0 1 0	0.065	0.909	0.014	0.357	0.998	0.358	45
ID93	0 1 0 0 1 0	0.082	0.783	0.469	0.188	0.998	0.078	43
ID94	1 1 0 1 0 0	0.664	0.912	0.001	0.624	0.384	0.009	43
ID95	1 1 1 1 1 0	0.810	0.986	0.960	0.917	1	0.036	78
ID96	1 1 0 0 1 0	0.591	0.984	0	0	1	0	42
ID97	0 0 0 0 0 0	0.001	0	0.056	0.044	0.001	0.042	2
ID98	0 1 1 1 0 1	0	0.985	0.897	0.888	0.328	0.752	64
ID99	0 1 0 0 1 0	0.083	0.656	0.468	0.414	1	0.267	48
ID100	0 1 0 0 1 0	0.020	0.848	0.024	0.11	0.965	0.073	33
ID101	1 1 0 1 0 0	0.564	0.573	0.002	0.564	0.436	0.005	35
ID102	0 0 0 0 0 0	0.045	0	0.39	0.048	0.022	0.031	8
ID103	0 0 0 0 1 0	0.009	0.037	0.241	0.068	0.993	0.001	22
ID104	0 0 0 0 1 0	0.004	0.044	0.118	0.031	0.911	0	18
ID105	0 0 0 0 1 0	0.018	0.077	0.065	0.002	0.983	0.021	19
ID106	0 0 0 0 1 0	0.015	0.034	0.005	0.007	0.996	0.002	17
ID107	0 1 1 1 1 0	0.002	0.988	0.858	0.858	0.953	0	61
ID108	0 1 0 0 1 1	0.006	0.948	0.068	0.064	0.950	0.626	44
ID109	0 1 0 0 1 0	0.076	0.97	0.192	0.193	0.965	0.001	39
ID110	0 1 0 0 1 0	0.048	0.567	0.228	0.062	0.999	0.005	31
ID111	1 0 1 0 1 1	0.909	0.377	0.855	0.071	1	0.669	64
ID112	1 1 1 1 1 0	0.876	0.991	0.985	0.951	1	0.001	80
ID113	0 1 0 1 1 1	0.112	0.942	0.059	0.918	0.999	0.874	65
ID114	0 0 0 0 1 0	0.001	0.005	0.410	0.013	0.719	0	19
ID115	0 1 1 1 0 0	0	0.625	0.698	0.953	0.047	0.255	42
ID116	0 1 1 1 0 0	0	0.750	0.984	0.984	0.017	0.001	45
ID117	0 0 1 0 0 0	0	0	0.984	0.043	0.003	0.054	18
ID118	0 1 1 1 0 0	0	0.843	0.952	0.954	0.081	0.003	47
ID119	0 0 1 0 0 0	0	0	0.966	0.120	0.019	0.057	19
ID120	1 1 1 1 1 0	0.921	0.997	0.870	0.967	1	0.347	85
ID121	0 0 1 0 0 0	0.413	0.483	0.544	0.041	0.450	0.072	33
ID122	0 0 1 0 1 0	0.115	0.393	0.510	0.371	0.999	0.045	40
ID123	1 1 1 1 1 1	0.878	0.991	0.975	0.963	1	0.811	93

ID124	1 1 1 1 1 1	0.923	0.999	0.978	0.962	1	0.682	92
ID125	0 1 0 0 1 0	0.006	0.818	0.057	0.304	0.961	0.248	39
ID126	0 1 1 1 1 1	0.050	0.991	0.810	0.796	1	0.640	71
ID127	0 1 0 0 1 0	0.394	0.885	0.001	0.001	0.999	0.047	38
ID128	0 0 0 0 0 0	0	0	0.256	0.255	0.239	0	12
ID129	0 1 1 1 0 0	0.004	0.897	0.614	0.606	0.398	0.001	41
ID130	1 1 1 1 1 0	0.682	0.992	0.868	0.791	1	0.079	73
ID131	0 0 1 0 0 0	0	0	0.945	0.046	0.002	0.009	16
ID132	1 0 0 0 0 0	0.849	0	0.006	0.045	0.058	0	15
ID133	0 0 0 0 1 0	0.005	0.006	0.025	0.028	0.977	0.026	17
ID134	0 0 0 0 1 0	0.070	0.490	0	0.015	0.985	0	26
ID135	0 0 1 0 0 1	0	0	0.996	0.004	0.001	0.916	31
ID136	1 1 0 1 0 0	0.865	0.906	0.001	0.865	0.135	0.008	46
ID137	0 0 0 0 1 0	0.019	0.028	0.008	0.008	0.997	0	17
ID138	0 1 1 1 1 0	0.125	0.749	0.758	0.773	1	0.321	62
ID139	0 1 1 1 0 0	0	0.951	0.977	0.971	0.066	0.005	49
ID140	0 0 0 0 1 0	0.005	0.015	0.072	0.068	0.935	0.115	20
ID141	0 0 0 0 1 0	0.120	0.326	0.341	0.009	0.970	0	29
ID142	0 0 1 0 0 0	0.001	0	0.776	0.047	0.006	0.017	14
ID143	0 0 0 0 1 0	0.001	0.057	0.018	0.011	0.974	0.023	18
ID144	0 0 0 1 0 1	0	0	0.002	1	0	0.998	33
ID145	0 1 1 1 0 0	0.001	0.919	0.698	0.7	0.302	0	43
ID146	1 1 0 0 1 0	0.656	0.920	0.039	0	1	0.006	43
ID147	0 1 0 0 1 0	0.001	0.909	0.19	0.155	0.918	0.279	40
ID148	1 1 1 1 1 0	0.866	0.995	0.965	0.935	1	0.018	79
ID149	1 0 0 1 0 0	0.557	0.432	0.001	0.559	0.442	0.005	33
ID150	0 1 0 0 1 0	0.387	0.844	0.006	0.023	0.999	0.018	37
ID151	1 1 1 0 1 0	0.732	0.886	0.692	0	1	0.002	55
ID152	1 1 1 1 1 0	0.572	0.922	0.863	0.818	1	0.074	70
ID153	1 1 0 0 1 0	0.605	0.853	0.096	0.001	1	0.053	43
ID154	1 0 0 0 1 1	0.512	0.200	0.008	0.002	0.997	0.677	39
ID155	1 1 0 1 1 1	0.959	0.999	0.346	0.920	1	0.647	81

ID156	0 1 0 0 1 0	0.116	0.769	0.007	0.002	0.993	0.001	31
ID157	0 1 0 0 1 0	0.211	0.605	0.272	0	0.814	0.009	31
ID158	1 1 1 1 1 1	0.977	1	0.989	0.983	1	0.992	98
ID159	0 0 0 0 1 0	0.020	0.483	0.001	0.001	0.999	0	25
ID160	0 1 0 0 1 0	0.004	0.548	0.145	0.151	0.957	0.006	30
ID161	0 1 1 1 0 0	0.000	0.876	0.849	0.851	0.149	0.001	45
ID162	1 1 1 1 1 0	0.924	1	0.871	0.969	1	0.348	85
ID163	0 1 0 1 1 1	0.137	0.997	0.16	0.948	1	0.814	67
ID164	0 0 0 0 1 0	0.09	0.435	0.096	0	0.934	0.202	29
ID165	0 0 0 0 1 0	0.009	0.031	0.083	0.039	0.968	0.003	18
ID166	0 1 1 1 1 0	0.160	0.995	0.773	0.686	0.999	0.008	60
ID167	0 1 0 0 1 0	0.108	0.902	0.48	0	1	0.033	42
ID168	1 1 0 1 0 0	0.905	0.862	0	0.905	0.098	0.002	46
ID169	1 1 1 0 0 0	0.571	0.985	0.528	0	0.475	0.004	42
ID170	0 0 0 0 1 0	0.255	0.008	0.010	0.014	0.709	0.009	16
ID171	1 1 1 1 1 1	0.968	1	0.842	0.976	1	0.575	89
ID172	0 0 0 0 1 0	0.001	0.478	0.089	0.041	0.928	0.012	25
ID173	1 1 1 1 1 0	0.871	0.995	0.978	0.954	1	0.011	80
ID174	1 1 1 1 1 0	0.832	0.971	0.879	0.590	1	0.007	71
ID175	0 1 0 0 1 0	0.439	0.987	0	0	0.999	0	40
ID176	0 0 0 0 1 0	0.041	0.088	0.003	0.001	0.995	0.022	19
ID177	0 0 0 0 1 1	0.000	0.012	0.008	0.001	0.953	0.787	29
ID178	0 1 0 0 1 0	0.059	0.887	0.052	0.299	0.997	0.25	42
ID179	0 0 0 0 0 0	0.002	0.001	0.043	0.128	0.109	0.091	6
ID180	1 0 0 0 0 0	0.749	0	0.005	0.137	0.165	0.017	17
ID181	1 1 1 1 1 1	1	1	1	1	1	1	99
ID182	0 0 0 0 1 0	0.005	0.074	0.029	0.026	0.998	0.328	24
ID183	0 0 0 0 1 0	0.099	0.170	0.003	0	0.999	0.020	21
ID184	0 0 0 0 1 0	0.000	0.002	0.023	0.009	0.971	0	16
ID185	0 1 0 0 1 0	0.034	0.811	0	0.003	0.998	0	30
ID186	1 1 1 1 1 0	0.814	0.994	0.988	0.988	0.999	0.026	80
ID187	0 0 0 0 1 0	0.126	0.272	0.013	0.02	0.986	0.008	23

ID188	1 1 1 1 1 1	0.980	1	1	1	1	0.999	99
ID189	1 1 0 1 0 0	0.976	0.994	0	0.975	0.025	0	49
ID190	0 0 0 0 0 0	0.410	0.471	0.006	0.412	0.476	0.019	29
ID191	0 1 0 0 1 0	0.043	0.567	0	0.004	0.997	0	26
ID192	1 1 1 0 1 1	0.991	1	0.977	0.001	1	0.977	82
ID193	1 1 1 1 1 0	0.587	0.964	0.890	0.890	0.997	0.001	72
ID194	0 0 0 0 0 0	0.003	0	0.058	0.003	0.001	0	1
ID195	0 0 0 0 1 0	0.071	0.123	0.020	0	1	0.008	20
ID196	0 0 0 0 1 0	0.013	0.231	0.077	0.069	0.887	0.016	21
ID197	0 1 0 0 1 0	0.389	0.845	0.002	0.380	0.640	0.024	37
ID198	0 0 0 0 1 0	0.0150	0.460	0.001	0.009	0.990	0.001	24
ID199	1 1 0 0 1 0	0.538	0.989	0.286	0.192	1	0.001	50
ID200	0 1 0 0 1 0	0.413	0.981	0.130	0	1	0	42
ID201	0 1 0 0 1 0	0.439	0.987	0	0	1	0	40
ID202	0 1 0 1 1 1	0.051	0.983	0.017	0.891	0.962	0.839	62
ID203	0 0 0 0 1 0	0.001	0.159	0.027	0.023	0.972	0.003	19
ID204	0 0 0 1 0 1	0	0.001	0.354	0.895	0.104	0.544	31
ID205	1 0 0 1 0 0	0.694	0	0.003	0.852	0.148	0.155	30
ID206	1 1 0 0 0 0	0.576	0.741	0.435	0.089	0.458	0.022	38
ID207	0 0 0 0 1 1	0.212	0.186	0.104	0.078	0.787	0.738	35
ID208	0 0 0 0 0 0	0.001	0	0.056	0.044	0.001	0.042	2
ID209	0 1 0 0 1 0	0.031	0.903	0.204	0.063	0.998	0.007	36
ID210	0 0 0 0 1 0	0.013	0.155	0.01	0.004	0.995	0.001	19
ID211	0 1 0 0 1 0	0.123	0.685	0.02	0.02	0.981	0.237	34
ID212	0 1 0 0 1 0	0	0.891	0.11	0.111	0.84	0.003	32
ID213	0 1 0 0 1 0	0.022	0.881	0.001	0.001	0.999	0	31
ID214	0 1 0 0 1 0	0.363	0.648	0.001	0.309	0.692	0.002	33
ID215	1 1 0 0 1 0	0.630	0.727	0.001	0.002	0.986	0.003	39
ID216	0 1 1 0 1 0	0.035	0.883	0.506	0.469	0.999	0.036	48
ID217	0 1 1 0 1 0	0.220	0.923	0.735	0.491	0.999	0.007	56
ID218	0 1 0 0 1 0	0.094	0.843	0.464	0.001	0.999	0.001	40
ID219	0 1 0 0 1 0	0.384	0.659	0.003	0	1	0.008	34

ID220	1 1 0 0 1 0	0.507	0.871	0.004	0	1	0	39
ID221	0 1 0 0 1 0	0.038	0.842	0.028	0.001	0.999	0.019	32
ID222	0 1 1 1 1 0	0.109	0.849	0.79	0.663	0.999	0.005	56
ID223	1 1 0 1 0 0	0.609	0.851	0.009	0.590	0.411	0.001	41
ID224	1 1 0 0 1 0	0.508	0.870	0	0	1	0	39
ID225	0 0 0 0 1 0	0.009	0.342	0.007	0.005	0.994	0.001	22
ID226	0 1 1 1 1 0	0.181	0.998	0.97	0.97	1	0	68
ID227	0 0 0 0 1 0	0.115	0.021	0.002	0.109	0.873	0	18
ID228	0 0 0 0 1 0	0.001	0.020	0.02	0.025	0.975	0.028	17
ID229	1 1 1 1 1 0	0.63	0.999	0.925	0.775	1	0.007	72
ID230	0 1 0 0 1 0	0.074	0.967	0.041	0	0.961	0	34
ID231	0 1 0 0 1 0	0.001	0.538	0.052	0.045	0.598	0.012	20
ID232	0 0 0 0 1 0	0.039	0.472	0.033	0.005	0.957	0.005	25
ID233	1 1 0 1 0 0	0.553	0.693	0.014	0.565	0.373	0	36
ID234	0 1 0 0 1 0	0.046	0.790	0.041	0.103	0.998	0.068	34
ID235	0 0 0 0 1 0	0.172	0.302	0.002	0.001	0.982	0	24
ID236	0 0 0 0 1 0	0.195	0.215	0.002	0.145	0.799	0.001	22
ID237	0 1 0 0 1 0	0.226	0.796	0	0.15	0.846	0	33
ID238	0 1 0 0 1 0	0.075	0.762	0.018	0.019	0.983	0.003	30
ID239	0 1 0 1 1 0	0.021	0.960	0.307	0.709	1	0.458	57
ID240	0 1 0 0 1 0	0.001	0.629	0.072	0.064	0.529	0.020	21
ID241	0 1 0 0 1 0	0.016	0.930	0.007	0.018	0.919	0	31
ID242	0 0 0 0 1 0	0.080	0.119	0.002	0.079	0.921	0.022	20
ID243	0 0 1 0 0 0	0	0	0.963	0.134	0.011	0.004	18
ID244	0 0 0 0 1 0	0.013	0.086	0.082	0.038	0.993	0.021	20
ID245	1 1 1 1 1 0	0.863	0.983	0.957	0.925	1	0.004	78
ID246	0 0 0 0 1 1	0.006	0.034	0.013	0.013	0.997	0.585	27
ID247	0 0 0 0 1 0	0.010	0.159	0.129	0.092	0.949	0.078	23
ID248	0 0 0 0 1 0	0.036	0.220	0	0.001	0.976	0	20
ID249	0 0 0 0 0 0	0	0	0.051	0.129	0.001	0.128	5
ID250	0 1 0 0 1 0	0.142	0.621	0.228	0.155	0.999	0.043	36
ID251	0 1 0 0 1 0	0.024	0.876	0.028	0.001	0.999	0.009	32

ID252	0 0 0 0 1 0	0.060	0.182	0.161	0.164	0.965	0.003	25
ID253	1 1 1 1 1 0	0.706	0.960	0.860	0.948	0.926	0.420	80
ID254	1 1 1 0 0 0	0.917	0.939	0.919	0	0.087	0.002	47
ID255	0 0 0 1 0 1	0.017	0.239	0.239	0.880	0.120	0.616	35
ID256	1 1 0 1 0 0	0.985	0.981	0	0.984	0.017	0.001	49
ID257	0 0 0 0 1 0	0.001	0.176	0.038	0.027	0.965	0	20
ID258	0 0 0 1 0 1	0	0.011	0.019	0.852	0.148	0.834	31
ID259	0 0 0 0 1 0	0.011	0.122	0.04	0.029	0.982	0.022	20
ID260	0 1 1 1 1 0	0.077	0.988	0.871	0.749	0.999	0	61
ID261	0 0 0 0 0 0	0.002	0	0.224	0.049	0.004	0.04	5
ID262	0 0 0 0 0 0	0.267	0.442	0.201	0.187	0.343	0.041	24
ID263	1 1 1 1 1 0	0.945	0.998	0.954	0.96	1	0.106	82
ID264	1 1 1 0 0 0	0.820	0.855	0.835	0.004	0.176	0	44
ID265	0 1 0 0 1 0	0.031	0.514	0.098	0.094	0.997	0.021	29
ID266	1 1 1 1 1 0	0.952	0.996	0.988	0.956	1	0.006	81
ID267	0 0 0 0 1 1	0.002	0.010	0.008	0	1	0.931	32
ID268	1 1 1 1 1 0	0.731	0.985	0.925	0.77	1	0.004	73
ID269	1 1 1 1 1 0	0.952	0.996	0.988	0.956	1	0.006	81
ID270	0 1 0 0 1 0	0.082	0.732	0.425	0	0.999	0.002	37
ID271	1 1 1 1 1 1	0.999	1	0.992	0.985	1	0.981	99
ID272	0 1 0 0 1 0	0.000	0.923	0.440	0.430	0.828	0.46	51
ID273	0 1 1 1 0 0	0	0.997	0.993	0.995	0.005	0	49
ID274	0 1 1 1 0 0	0.040	0.973	0.893	0.943	0.058	0	48
ID275	0 1 0 0 1 0	0.015	0.630	0.014	0.043	0.989	0.04	28
ID276	0 1 0 0 1 0	0.004	0.795	0.001	0	0.986	0	29
ID277	1 0 0 1 0 0	0.930	0.002	0.004	0.935	0.066	0	32
ID278	0 0 0 0 0 0	0	0	0.055	0.050	0.008	0.048	2
ID279	0 0 0 0 0 0	0.000	0.026	0.064	0.080	0.029	0.035	3
ID280	0 0 0 0 0 0	0	0.003	0.196	0.016	0.322	0.191	12
ID281	0 0 0 0 0 0	0.211	0.34	0.246	0.051	0.491	0.083	23
ID282	0 0 0 0 1 1	0	0	0.094	0.037	0.884	0.612	27
ID283	0 0 0 0 0 0	0.423	0	0.032	0.024	0.030	0	8

ID284	0 0 0 0 1 0	0.022	0.486	0.013	0.005	0.999	0.005	25
ID285	0 0 0 0 0 0	0	0	0.003	0.003	0.001	0	0
ID286	1 0 0 0 0 0	0.816	0	0.006	0.068	0.066	0.039	16
ID287	0 1 0 1 1 1	0.006	0.956	0.005	0.939	0.998	0.936	63
ID288	0 1 0 0 1 0	0.022	0.875	0.151	0.003	0.998	0.031	34
ID289	0 0 0 0 1 0	0.002	0.069	0.023	0.002	0.956	0.04	18
ID290	0 0 1 0 0 0	0.000	0.03	0.504	0.043	0.363	0.066	16
ID291	1 1 1 1 1 0	0.594	0.815	0.784	0.786	0.998	0.003	66
ID292	0 0 1 1 1 0	0.181	0.444	0.699	0.701	0.900	0.003	48
ID293	0 0 0 0 1 0	0.077	0.005	0.007	0.09	0.910	0.03	18
ID294	1 0 1 0 1 1	0.894	0.138	0.915	0.077	1	0.841	64
ID295	0 1 0 0 1 0	0.024	0.690	0.193	0.188	0.945	0	33
ID296	0 1 1 1 0 0	0	0.845	0.778	0.789	0.214	0.066	44
ID297	1 1 0 1 0 0	0.638	0.717	0.001	0.628	0.362	0.002	39
ID298	0 0 0 0 1 0	0.027	0.017	0.061	0.086	0.914	0	18
ID299	0 0 0 0 1 0	0.414	0.467	0.017	0.424	0.590	0	31
ID300	0 1 0 1 1 1	0.021	0.996	0.140	0.970	1	0.855	66
ID301	0 1 0 1 1 1	0.028	0.987	0.362	0.915	1	0.580	64
ID302	0 0 0 0 1 0	0.131	0.200	0.03	0	0.999	0.019	22
ID303	0 0 0 0 0 0	0.003	0	0.004	0.034	0.006	0.032	1
ID304	1 1 1 0 0 0	0.941	0.946	0.783	0.163	0.044	0.017	48
ID305	0 1 0 0 1 0	0.105	0.762	0.118	0.373	0.999	0.348	45
ID306	0 1 0 0 1 0	0.105	0.762	0.118	0.373	0.999	0.348	45
ID307	0 0 0 0 1 0	0.005	0.031	0.256	0.064	0.599	0.211	19
ID308	0 0 0 0 0 0	0.141	0	0.149	0.040	0.200	0.032	9
ID309	1 1 1 0 0 1	0.997	0.998	0.997	0	0.003	0.997	66
ID310	0 1 0 0 1 0	0.433	0.805	0.042	0.084	1	0.046	40
ID311	1 1 0 1 0 0	0.923	0.929	0.003	0.931	0.069	0.006	47
ID312	1 1 0 1 0 0	0.848	0.787	0.001	0.848	0.126	0	43
ID313	0 0 0 0 1 0	0.438	0.414	0.421	0.002	0.555	0.002	30
ID314	0 1 1 1 1 0	0.038	0.994	0.851	0.851	1	0	62
ID315	1 1 0 0 1 0	0.749	0.938	0	0	1	0	44

ID316	0 0 1 0 0 0	0	0.218	0.665	0.276	0.317	0.001	24
ID317	1 1 0 1 0 0	0.712	0.924	0.001	0.693	0.315	0.007	44
ID318	0 0 0 1 0 1	0	0	0.004	1	0	0.996	33
ID319	0 0 0 0 1 1	0.000	0.018	0.004	0.001	0.997	0.79	30
ID320	1 1 1 1 1 1	0.999	1	0.992	0.985	1	0.981	99
ID321	0 0 0 0 1 0	0.081	0.001	0.007	0.082	0.623	0.431	20
ID322	0 1 0 0 1 0	0.035	0.763	0.036	0.044	0.996	0.037	31
ID323	0 1 1 1 0 0	0	0.654	0.524	0.524	0.112	0.004	30
ID324	0 0 0 0 1 0	0.432	0.488	0.239	0.162	0.569	0.001	31
ID325	0 0 0 0 1 0	0.057	0.005	0.002	0.065	0.936	0.028	18
ID326	0 0 0 0 1 0	0.047	0.438	0.382	0.11	0.969	0.002	32
ID327	0 1 0 0 1 0	0.037	0.767	0.035	0.001	0.998	0.001	30
ID328	0 0 0 0 1 0	0.043	0.495	0.028	0.001	0.999	0.012	26
ID329	1 1 1 0 0 0	0.816	0.772	0.738	0.033	0.176	0	42
ID330	0 1 1 1 1 0	0.485	0.989	0.965	0.940	0.999	0.009	73
ID331	0 0 0 0 1 0	0.013	0.027	0.265	0.100	0.954	0.018	22
ID332	0 0 0 0 1 0	0.006	0.061	0.009	0.008	0.997	0.023	18
ID333	0 0 0 0 1 0	0.005	0.082	0.040	0.04	0.989	0	19
ID334	1 1 1 1 1 0	0.854	0.999	0.964	0.989	1	0.041	80
ID335	0 0 0 0 1 0	0.005	0.023	0.063	0.002	0.944	0	17
ID336	0 0 0 0 0 0	0.007	0	0.057	0.003	0.001	0	1
ID337	0 0 0 0 1 0	0.063	0.1	0.067	0.003	0.88	0.012	18
ID338	0 0 0 0 1 0	0.296	0	0.046	0.299	0.528	0.019	19
ID339	1 0 0 0 0 0	0.844	0	0.006	0.05	0.059	0.006	16
ID340	0 0 0 0 0 0	0.005	0.066	0.069	0.163	0.015	0.137	7
ID341	0 0 0 1 0 1	0.057	0	0.034	0.953	0.047	0.869	32
ID342	0 0 1 1 0 0	0	0.001	0.988	0.988	0.012	0	33
ID343	0 0 0 0 1 0	0.005	0.325	0.089	0.100	0.913	0.012	24
ID344	1 0 0 0 0 0	0.770	0.380	0.209	0.195	0.193	0	28
ID345	0 0 0 0 1 1	0.005	0.170	0.017	0.005	0.983	0.590	29
ID346	1 1 0 1 0 0	0.714	0.733	0.017	0.745	0.256	0.019	41
ID347	1 1 0 1 1 1	0.707	0.987	0.171	0.945	1	0.824	77

ID348	1 1 0 0 1 0	0.570	0.986	0.117	0	1	0	44
ID349	1 1 1 0 0 0	0.758	0.827	0.766	0.001	0.253	0.016	43
ID350	1 1 0 1 0 0	0.838	0.823	0	0.837	0.163	0.07	45
ID351	0 0 0 0 1 0	0.000	0.049	0.049	0.049	0.771	0	15
ID352	0 1 0 0 1 0	0.039	0.883	0.001	0.005	0.994	0	32
ID353	0 1 0 0 1 0	0.313	0.586	0.239	0.045	0.704	0.064	32
ID354	0 1 0 0 1 0	0.024	0.965	0.006	0.001	0.999	0	33
ID355	0 0 0 0 1 0	0.037	0.223	0.003	0.001	0.975	0.002	20
ID356	0 1 0 0 1 0	0.001	0.982	0.164	0.164	0.907	0	36
ID357	0 0 0 0 1 0	0.035	0.184	0.005	0.003	0.992	0.001	20
ID358	0 0 0 0 1 0	0.091	0.016	0.003	0.093	0.908	0	18
ID359	1 1 0 1 0 0	0.930	0.951	0.001	0.93	0.050	0	47
ID360	0 1 0 0 1 0	0.156	0.800	0.051	0	0.966	0.024	33
ID361	1 1 1 1 1 1	1	1	1	1	1	1	99
ID362	0 0 0 1 0 0	0.000	0.383	0.47	0.517	0.239	0.238	30
ID363	0 1 0 0 1 0	0.035	0.752	0.025	0.001	0.999	0.001	30
ID364	0 0 0 0 0 0	0	0.386	0.067	0.066	0.038	0.029	9
ID365	0 1 0 0 1 0	0.048	0.843	0.420	0.261	0.996	0.099	44
ID366	1 1 1 0 0 0	0.855	0.887	0.848	0.001	0.146	0	45
ID367	0 0 0 0 1 0	0.082	0.493	0.005	0.057	0.945	0.013	26
ID368	1 1 0 0 1 0	0.526	0.744	0.002	0.012	0.998	0.010	38
ID369	0 1 0 0 1 0	0.074	0.777	0.079	0.075	0.997	0.022	33
ID370	0 0 0 0 1 0	0.114	0.203	0.062	0.001	0.997	0.018	23
ID371	0 0 0 0 1 0	0.001	0.277	0.236	0.22	0.783	0	25
ID372	0 0 0 0 0 0	0.268	0.170	0.145	0.058	0.200	0.029	14
ID373	0 0 1 0 0 0	0	0	0.839	0.037	0.186	0	17
ID374	1 1 0 1 0 0	0.590	0.778	0.002	0.588	0.433	0.024	40
ID375	0 1 0 0 1 0	0.176	0.862	0.117	0.017	0.870	0.017	34
ID376	0 1 0 0 1 0	0.180	0.881	0.022	0.289	0.803	0.105	37
ID377	1 1 0 1 0 0	0.985	0.989	0.003	0.987	0.013	0	49
ID378	1 1 0 0 0 0	0.830	0.956	0.482	0.345	0.169	0	46
ID379	0 1 1 1 1 0	0.037	0.978	0.838	0.810	1	0.001	61

ID380	0 0 0 0 1 1	0.025	0.012	0.001	0.022	0.978	0.833	31
ID381	0 0 0 0 1 1	0.052	0.178	0.237	0.211	0.998	0.521	36
ID382	0 1 0 1 1 1	0.020	0.989	0.105	0.970	1	0.891	66
ID383	0 0 0 0 1 1	0.041	0.045	0.025	0.014	0.960	0.9	33
ID384	1 1 1 1 1 1	1	1	1	1	1	1	99
ID385	0 0 0 0 1 0	0.000	0.002	0.133	0.134	0.867	0	18
ID386	0 0 0 0 1 0	0.257	0.270	0.228	0.010	0.742	0.111	26
ID387	0 1 0 0 1 0	0.037	0.767	0.035	0.001	0.998	0.001	30
ID388	0 0 0 0 0 0	0.003	0	0.028	0.02	0.001	0.026	1
ID389	0 1 0 0 1 0	0.151	0.551	0	0.001	0.999	0	28
ID390	0 0 0 0 1 0	0.003	0.045	0.007	0.008	0.995	0.074	18
ID391	1 1 1 0 1 0	0.796	0.964	0.738	0.002	1	0	58
ID392	0 0 0 0 0 0	0.002	0	0.057	0.004	0.001	0	1
ID393	1 1 0 1 0 0	0.876	0.954	0.070	0.95	0.051	0.001	48
ID394	0 0 0 0 1 0	0.101	0.026	0.008	0.018	0.883	0	17
ID395	1 1 0 1 1 1	0.759	0.999	0.006	0.952	1	0.948	77
ID396	1 1 0 0 1 0	0.511	0.871	0.417	0.023	0.56	0.208	43
ID397	0 1 1 1 1 0	0.253	0.869	0.808	0.681	1	0.018	60
ID398	1 1 0 0 1 0	0.543	0.748	0.001	0	1	0.002	38
ID399	0 0 0 0 1 0	0.008	0.016	0.006	0.001	0.991	0	17
ID400	0 0 0 0 0 0	0.039	0.002	0.339	0.041	0.150	0.030	10
ID401	1 1 1 1 1 1	1	1	1	1	1	1	99
ID402	1 1 1 1 1 1	0.904	0.999	0.963	0.882	1	0.885	93
ID403	1 1 1 1 1 1	0.976	1	0.988	0.983	1	0.946	98
ID404	1 1 1 1 1 1	0.979	1	0.997	0.988	1	0.976	98
ID405	1 1 0 0 1 0	0.685	0.988	0.001	0.001	1	0	44
ID406	1 1 1 1 1 1	0.988	1	0.999	0.999	1	0.832	96
ID407	1 1 1 0 0 0	0.956	0.925	0.889	0.036	0.042	0.045	48
ID408	0 0 0 0 1 0	0.003	0.063	0.073	0.056	0.920	0.008	18
ID409	0 1 1 1 1 0	0.000	0.997	0.965	0.933	0.987	0	64
ID410	0 0 0 0 1 0	0.000	0.009	0.011	0.004	0.988	0.024	17
ID411	1 1 1 0 1 0	0.621	0.751	0.617	0.001	0.998	0	49

ID412	0 0 0 0 1 0	0.020	0.464	0.004	0.004	0.999	0	24
ID413	0 0 0 1 0 1	0.000	0	0.008	0.999	0.001	0.991	33
ID414	0 0 1 1 0 0	0	0	0.984	0.984	0.016	0.011	33
ID415	0 0 0 0 0 0	0.003	0	0.004	0.035	0.001	0.033	1
ID416	0 0 0 0 1 0	0.008	0.018	0.002	0.002	0.998	0	17
ID417	1 1 1 1 1 1	0.999	1	0.993	0.899	1	0.996	98
ID418	0 0 0 1 0 1	0	0	0.009	0.983	0.017	0.981	33
ID419	1 1 0 0 1 0	0.730	0.934	0.001	0	1	0.019	44
ID420	1 1 0 0 1 0	0.559	0.816	0	0.462	0.536	0.002	39
ID421	1 1 1 1 1 1	0.975	0.999	0.988	0.983	1	0.946	98
ID422	1 1 1 1 1 0	0.813	0.998	0.987	0.953	1	0	79
ID423	0 0 0 0 1 0	0.001	0.217	0.085	0.003	0.926	0	20
ID424	1 1 1 1 1 1	0.995	0.999	0.992	0.992	1	0.966	99
ID425	0 0 0 0 0 0	0	0.296	0.468	0.301	0.181	0.068	21
ID426	0 1 0 0 1 0	0.259	0.904	0.228	0.009	0.766	0.007	36
ID427	1 1 1 1 1 1	0.999	1	1	1	1	1	99
ID428	0 0 1 0 0 0	0	0	0.577	0.041	0.027	0.022	11
ID429	0 0 0 0 1 0	0.022	0.056	0.004	0.023	0.915	0.009	17
ID430	1 1 1 0 0 0	0.677	0.692	0.663	0.001	0.322	0.007	39
ID431	0 0 0 0 1 0	0.113	0.153	0.112	0.009	0.842	0.021	20
ID432	0 0 0 0 1 0	0.008	0.013	0.015	0.001	0.975	0.024	17
ID433	0 1 0 0 1 0	0.050	0.768	0.011	0.010	0.997	0.006	30
ID434	0 1 1 1 0 0	0.013	0.843	0.807	0.981	0.019	0.143	46
ID435	0 1 0 0 1 0	0.066	0.978	0.494	0.001	0.999	0	42
ID436	0 0 0 0 1 0	0.013	0.266	0.091	0	0.999	0.464	30
ID437	0 0 0 0 0 0	0.000	0.040	0.45	0.053	0.174	0.291	16
ID438	0 1 0 0 1 0	0.046	0.800	0.184	0.059	0.999	0.005	34
ID439	0 1 1 0 1 0	0.088	0.770	0.502	0.151	0.998	0.013	42
ID440	0 0 1 0 0 0	0	0	0.874	0.046	0.005	0.012	15
ID441	0 0 0 0 0 0	0.002	0	0.417	0.048	0.002	0.032	8
ID442	0 0 1 0 0 0	0	0.078	0.593	0.099	0.016	0.023	13
ID443	0 0 0 0 0 0	0.002	0	0.055	0.050	0.001	0.048	2

ID444	0 0 0 0 1 0	0.000	0.009	0.269	0.110	0.901	0.019	21
ID445	0 0 0 0 1 0	0.011	0.008	0.019	0.011	0.986	0.009	17
ID446	0 0 1 0 0 0	0	0.040	0.973	0.083	0.020	0.008	18
ID447	0 0 1 0 0 0	0	0.040	0.973	0.083	0.020	0.008	18
ID448	0 0 1 0 0 0	0.036	0.133	0.790	0.127	0.035	0	18
ID449	0 0 0 0 1 0	0.038	0.090	0.051	0	0.965	0.008	19
ID450	0 0 0 0 1 0	0.005	0.027	0.119	0.072	0.983	0.314	25
ID451	0 0 0 0 1 0	0.000	0.006	0.005	0.004	0.994	0.354	22
ID452	0 1 1 1 1 0	0.000	0.932	0.942	0.933	0.877	0.016	61
ID453	0 1 1 1 1 0	0.423	0.917	0.776	0.776	0.990	0.179	67
ID454	0 0 0 0 1 0	0.031	0.076	0.002	0.002	0.971	0.022	18
ID455	0 1 1 1 1 0	0.132	0.837	0.512	0.729	1	0.307	58
ID456	0 0 0 0 1 0	0.067	0.147	0.006	0.003	0.992	0.020	20
ID457	1 0 0 1 0 0	0.775	0	0.004	0.779	0.221	0.038	30
ID458	1 1 1 1 1 0	0.807	0.989	0.98	0.971	1	0.013	79
ID459	0 0 0 0 1 1	0.002	0.027	0.011	0.009	0.987	0.63	27
ID460	0 0 0 0 1 0	0.014	0.062	0.102	0.013	0.996	0.036	20
ID461	0 0 0 0 1 0	0.120	0.246	0.023	0.003	0.995	0.001	23
ID462	0 0 1 0 1 0	0.001	0.001	0.508	0.408	0.598	0.017	25
ID463	0 0 0 0 1 0	0.012	0.007	0.052	0.019	0.983	0.024	18
ID464	1 1 0 0 1 0	0.544	0.742	0.072	0	1	0.024	39
ID465	0 0 0 0 1 0	0.048	0.057	0.010	0.012	0.771	0.026	15
ID466	1 1 1 1 1 1	0.980	1	1	1	1	0.999	99
ID467	0 1 0 0 1 0	0.082	0.732	0.425	0	0.999	0.002	37
ID468	0 1 0 0 1 0	0.282	0.984	0.249	0.010	0.745	0	37
ID469	1 1 1 0 1 0	0.870	0.999	0.820	0.301	1	0.331	72
ID470	0 0 0 0 0 0	0.161	0	0.041	0.150	0.006	0.141	8
ID471	0 0 0 0 0 0	0.001	0	0.057	0.003	0.001	0	1
ID472	1 1 1 1 1 0	0.876	0.991	0.985	0.951	1	0.001	80
ID473	0 1 1 1 1 1	0.032	1	1	1	1	0.993	83
ID474	1 1 1 1 1 0	0.924	1	0.871	0.969	1	0.348	85
ID475	1 1 1 1 1 1	1	1	1	1	1	1	99

ID476	0 0 0 0 1 0	0.057	0.004	0.004	0.006	0.741	0	13
ID477	0 1 1 1 1 0	0.180	0.994	0.967	0.967	1	0	68
ID478	1 1 0 0 1 0	0.624	0.819	0.439	0	1	0.002	48
ID479	0 1 1 1 1 0	0.172	0.764	0.561	0.561	0.999	0.001	50
ID480	0 0 0 0 1 1	0.020	0.088	0.095	0.073	0.997	0.716	33
ID481	0 1 0 0 1 0	0.004	0.627	0.357	0.235	0.973	0.008	36
ID482	0 1 1 1 1 0	0.359	0.997	0.973	0.941	0.999	0.001	71
ID483	0 1 1 1 1 0	0.351	0.942	0.536	0.687	1	0.456	66
ID484	0 0 0 1 0 1	0.002	0.012	0.042	0.833	0.167	0.792	30
ID485	0 0 0 0 0 0	0.072	0.001	0.05	0.006	0.062	0	3
ID486	0 0 0 0 1 0	0.374	0.424	0	0	1	0	29
ID487	0 0 0 0 1 1	0.001	0.147	0.004	0.003	0.997	0.694	30
ID488	1 1 1 1 1 1	0.709	1	0.999	0.999	1	0.965	94
ID489	1 1 1 0 1 0	0.872	0.987	0.785	0	1	0.104	62
ID490	0 1 0 0 1 0	0.157	0.982	0.023	0.001	0.999	0	36
ID491	1 1 1 1 1 0	0.832	0.999	0.935	0.789	1	0.047	76
ID492	1 1 1 1 1 1	0.980	1	1	1	1	0.998	99
ID493	1 1 1 1 1 0	0.871	0.995	0.978	0.954	1	0.011	80
ID494	1 1 1 1 1 0	0.845	0.999	0.948	0.957	1	0.048	79
ID495	0 0 0 0 1 0	0	0.270	0.148	0.007	0.756	0.042	20
ID496	0 1 0 0 1 0	0.165	0.803	0.117	0.104	0.980	0.020	36
ID497	1 1 1 1 1 0	0.707	0.950	0.905	0.747	1	0.002	71
ID498	0 1 0 0 1 0	0.068	0.623	0.03	0.005	0.993	0.003	28
ID499	0 1 1 1 1 0	0.094	0.990	0.867	0.859	0.999	0.007	63
ID500	0 1 0 1 1 1	0.096	0.997	0.390	0.975	1	0.662	68
ID501	0 0 0 0 1 0	0.148	0.190	0.019	0.218	0.783	0.068	23
ID502	0 1 1 1 1 0	0.384	0.980	0.603	0.692	1	0.381	67
ID503	0 1 0 0 1 0	0.037	0.767	0.061	0.078	0.998	0.030	32
ID504	1 1 1 1 1 1	0.979	1	0.997	0.996	1	0.989	99
ID505	0 0 0 0 1 0	0.064	0.017	0.022	0.063	0.891	0	17
ID506	1 1 1 1 1 1	0.980	1	1	1	1	0.992	99
ID507	0 1 0 0 1 0	0.449	0.773	0	0	0.998	0.002	37

ID508	0 1 0 0 1 0	0.123	0.778	0.011	0.114	0.920	0.025	32
ID509	1 0 0 1 0 0	0.937	0.001	0.001	0.938	0.063	0	32
ID510	1 1 1 1 1 1	0.840	0.999	0.679	0.952	1	0.921	89
ID511	0 1 0 0 1 0	0.191	0.884	0.186	0.402	0.999	0.240	48