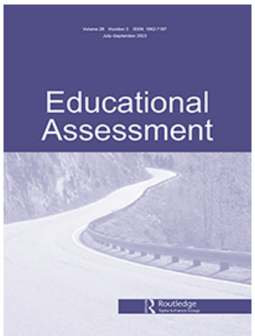


Routledge
Taylor & Francis Group

Educational Assessment

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/heda20

Assessing Mathematical Higher-Order Thinking Skills: An Analysis of Turkish University Entrance Examinations

Utkun Aydin & Bengi Birgili

To cite this article: Utkun Aydin & Bengi Birgili (2023) Assessing Mathematical Higher-Order Thinking Skills: An Analysis of Turkish University Entrance Examinations, Educational Assessment, 28:3, 190-209, DOI: 10.1080/10627197.2023.2202311

To link to this article: https://doi.org/10.1080/10627197.2023.2202311

Published online: 17 Apr 2023.

EDUCATIONAL ASSESSMENT
2023, VOL. 28, NO. 3, 190–209
https://doi.org/10.1080/10627197.2023.2202311

Routledge
Taylor & Francis Group

Assessing Mathematical Higher-Order Thinking Skills: An Analysis of Turkish University Entrance Examinations

Utkun Aydin [a] and Bengi Birgili [b]

[a]American University of the Middle East Kuwait; [b]MEF University Turkey

ABSTRACT
Internationally, mathematics education reform has been directed toward characterizing educational goals that go beyond topic/content/skill descriptions and develop students' problem solving. The Revised Bloom's Taxonomy and MATH (Mathematical Assessment Task Hierarchy) Taxonomy characterize such goals. University entrance examinations have been seen as one way of accomplishing these goals and influence learning, teaching, and assessment in mathematics. The present study analyzed mathematics items ($N =$ 1077) in Turkish university entrance examinations in 1998-2013 and objectives ($N = 621$) in mathematics curricula in 2005, 2011, and 2013 to determine the extent to which they represent the dimensions/categories of these taxonomies and the degree to which items are aligned with objectives in terms of reflecting the dimensions/categories of these taxonomies. The findings reveal that the items demand, to a large extent, automated computational skills; this is also evident in the relevant mathematics curricula. Implications for practice are discussed and could play a role in reforming assessment.

Introduction

The purpose of this study is to examine higher-order thinking skills that the mathematics items in Turkish university entrance examinations and the educational objectives in mathematics curricula reflect. The study is founded on research that points to limitations in students' higher-order thinking (Henningsen & Stein, 1997) and the reforms in curriculum, evaluation, and teaching practices (Ministry of National Education [MoNE], 2018; NCTM, 2000) intended to enhance students' learning of mathematics and become better – higher-order – thinkers. The extent to which higher-order thinking skills are taught and assessed continues to be an area of debate, with many researchers indicating that students display more low-level thinking skills than high level thinking skills that require procedural and conceptual knowledge (e.g., Rittle-Johnson, Schneider, & Star, 2015). Research also indicates that a focus on this type of thinking might weaken students' use of their knowledge in increasingly more complex ways (Anderson & Krathwohl, 2001) and performing in exam tasks (Bergqvist, 2007). These observations and arguments show a connection between a narrow focus on the nature of tasks and their potential to influence and structure student thinking. Most of the international and national results concern the tasks in large-scale assessments such as Trends in International Mathematics and Science Study (TIMSS) and Program for International Student

CONTACT Utkun Aydin Utkun.Aydin@aum.edu.kw Department of Mathematics College of Engineering and Technology American University of the Middle East Kuwait

Utkun Aydin was affiliated with the College of Engineering and Technology, Department of Mathematics at American University of the Middle East at the time of the revision process and is currently affiliated with the College of Social Sciences, School of Education at University of Glasgow, Utkun.Aydin@glasgow.ac.uk

Achievement (PISA), often in situations concerning performance, achievement, and/or competencies (e.g., Incikabi, 2012; Jakwerth, 1999; Rindermann & Baumeister, 2015). The situation regarding university entrance examinations, especially in Turkey, is not thoroughly studied and there are still many questions to be answered.

One related question is *to what extent* are university entrance examinations and mathematics curricula are aligned with each other? It is widely acknowledged that the success of education systems depends upon strong curriculum objectives and assessments that measure expectations of those curriculum objectives (Resnick, Rothman, Slattery, & Vranek, 2004). It has been recognized that when objectives do not suitably match test items, this variability in alignment may influence test scores (Davis-Becker & Buckendahl, 2017). Following on from this, if a specific construct measured by a test item does not align with the curriculum content experienced by particular students, then those students cannot be expected to do well on the test (American Educational Research Association [AERA], American Psychological Association [APA], & American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). University entrance examinations play a pivotal role in mobilizing educational reform movements (Davey, De Lian, & Higgins, 2007). Indeed, the majority of educational/pedagogical practices in secondary education are built upon the results of university entrance examinations rather than *vice versa* (Kuramoto & Koizumi, 2018). It is therefore important to study the extent to which higher-order thinking skills are contained in the university entrance examination mathematics items and educational objectives.

There are several reasons why it is important to study the exams as a part of mathematics education. The exams provide occasions when students, usually very effectively, engage in solving mathematical tasks (i.e., items/questions), and several studies show that students may approach mathematical tasks in different ways, with results depending on their thinking styles (Moutsios-Rentzos & Simpson, 2011). All in all, this indicates that exams, in general, influence the way students study mathematics (Ryan, Ryan, Arbuthnot, & Samuels, 2007). The important role of the university entrance examinations in Turkey is described in Section 3. The knowledge and cognitive processes that an accepted student will be equipped with after the exam are directly related to what is measured by the exam. This feature of the exams shows that the items included in them has practical consequences for students, teachers, and policy makers. Also, on a more abstract level, this threshold function of the exams shows that the design of the mathematics items reflects what the teachers and policymakers see as relevant content and higher-order thinking skills.

Mathematical knowledge and processes: A research framework

In response to the importance of students' higher-order thinking, combined with the crucial role of the university entrance examinations bring to mind the wide range of important learning outcomes and thinking skills students should attain. According to Brookhart (2010), taxonomies (i) guide in understanding how students transfer and transform what they learn; (ii) are useful for categorizing learning objectives and assessments according to level of complexity; and (iii) are effective tools for clarifying whether the instruction and assessment match the intended learning objective in both content (i.e., what the students learn about concepts and procedures) and cognitive complexity (i.e., what the students are able to think with these concepts and procedures – the learning).

There are several theoretical comprehensive frameworks that describe higher-order thinking skills (Biggs & Collis, 2014; Marzano & Kendall, 2007; Niss & Højgaard, 2019), which clearly have the ordering of categories from simple to complex and from concrete to abstract in common. *The Original Bloom's Taxonomy* (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) is the most widely used categorization in many curriculum and teaching materials, and subsequent frameworks tend to be closely linked to Bloom's work. The Original Bloom's Taxonomy classifies cognitive performances into six hierarchical processes, which represent a cumulative hierarchy (i.e., mastery of each simpler category is a prerequisite to mastery of the next more complex one): (1) Knowledge, (2) Comprehension, (3) Application, (4) Analysis, (5) Synthesis, and (6) Evaluation. Anderson and Krathwohl (2001) modified the Original Bloom's Taxonomy, producing the *Revised Bloom's Taxonomy*. Two major differences between the revised and the original are that the new version (1) changed the six categories from nouns to verbs and (2) rearranged them in two dimensions: Knowledge and Cognitive Process. The knowledge dimension includes four types of knowledge: (1) *Factual*, (2) *Conceptual*, (3) *Procedural*, and (4) *Metacognitive* whereas the cognitive process dimension involves six categories: (1) *Remember*, (2) *Understand*, (3) *Apply*, (4) *Analyze*, (5) *Evaluate*, and (6) *Create*. The main aim of this framework was twofold: (1) to classify statements of what we expect or intend students to learn as a result of instruction (Krathwohl, 2002, p. 212) and (2) to perform objective-based evaluation on students' achievement (Lee, Kim, Jin, Yoon, & Matsubara, 2017, p. 11). The operational definitions of the dimensions are summarized with reference to university entrance examination mathematics items in Table 1. For more details about the structure of dimensions, the reader is referred to Anderson and Krathwohl (2001), Krathwohl (2002), and Anderson (2005).

The MATH Taxonomy, originally developed by Smith et al. (1996), was an adaptation of Bloom's Original Taxonomy. Researchers drew on the assumption that changing teaching methods without due attention to assessment methods is not sufficient. Indeed, research has shown that students at all levels of education are more motivated to learn material that is of direct relevance to passing or getting a high grade (e.g., Torrance, 2007). The MATH Taxonomy includes eight categories of mathematical knowledge and skills, which are arranged in three groups: *Group A*, *Group B*, and *Group C*: (1) *Factual Knowledge*, (2) *Comprehension*, (3) *Routine Procedures*, (4) *Information Transfer*, (5) *Application to New Situations*, (6) *Justifying and Interpreting*, (7) *Implications, Conjectures and Comparisons*, and (8) *Evaluation*. The categories of the MATH Taxonomy are described with reference to university entrance examination mathematics items in Table 2. For more details about the framework, the reader is referred to Smith et al. (1996) and Ball et al. (1998).

The major distinction between the Revised Bloom's Taxonomy and MATH Taxonomy is that Anderson and Krathwohl's revisit is the only framework that has explicitly highlighted the importance, use, and assessment of metacognition. It is therefore a comprehensive framework for investigating teaching, learning, and assessment in mathematics (Radmehr & Drake, 2018, 2019). It is a powerful tool for fitting teachers' instructional goals and provides a concise visual representation of the alignment between these goals and assessment (see the Special Issue of the Theory into Practice (2002) devoted to "Revising Bloom's Taxonomy"). It has been widely used in mathematics (Radmehr & Drake, 2019), science (Lee, Kim, Jin, Yoon, & Matsubara, 2017), and curriculum (Porter, 2006) studies. In addition, there is theoretical and empirical support for matching the dimensions of the Revised Bloom's Taxonomy and the categories of the MATH Taxonomy (for details, see Smith et al., 1996). Drawing on previous research (e.g., Anderson & Krathwohl, 2001; Smith et al., 1996), Figure 1 presents the associations between the dimensions and categories of the two taxonomies. Aware of the strengths and weaknesses of the taxonomies (see Ari, 2011), we chose to build our investigation on two taxonomies for three reasons: (1) The Revised Bloom's Taxonomy extends beyond more complex aspects of learning (Anderson, 2005), which reflects the aim of university entrance examinations; (2) The Revised Bloom's Taxonomy provides an in-depth understanding of assessment formats (Airasian & Miranda, 2002), which can be integrated into the development of university entrance examinations; (3) the MATH Taxonomy is identified as useful for analyzing assessments in undergraduate mathematics courses (Bennie, 2005), which can be well-suited to the analysis of university entrance examinations.

Research context: university entrance examinations and mathematics curricula in Turkey

Turkey is a country of tests where students complete several large-scale national assessments. For university placement, central examinations dominate, managed by the Measurement, Selection and Placement Center (MSPC) attached to the Higher Education Council (HEC). The goal of MSPC is to

Table 1. Descriptions of Revised Bloom's Taxonomy Dimensions/Subdimensions along with Specimen University Examination Mathematics Items.

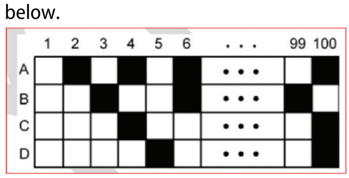
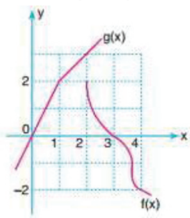
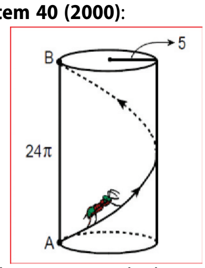
Dimension/Subdimension	Item (Year)
Knowledge: a range from concrete to abstract	
Factual Knowledge: The basic elements that students must know to be acquainted with mathematics. "What," "which," and "when" type of questions.	**Item 26 (1998)**: What is the range of $f(x) = \frac{2x+1}{x-1}$, defined in $R/\{1\}$? A) R B) $R/\{3\}$ C) $R/\{2\}$ D) $R/\{1\}$ E) $R/\{0\}$
Conceptual Knowledge: The interrelationships among the basic elements within a larger structure that enable the elements to function together. "Why" type of questions.	**Item 7 (2002)**: Let $a, b,$ and c be integers. If $a.b = 2c - 1$, then which one of the below is true? A) a and b are odd. B) a and b are even. C) a is even, b is odd. D) $a - b$ is odd. E) $a + b$ is odd.
Procedural Knowledge: How to use mathematical methods of inquiry, and criteria for using appropriate mathematical skills, algorithms, techniques, and methods. "How" type of questions.	**Item 10 (2007)**: Simplify $\frac{3^{2x}-2.3^{x+y}+3^{2y}}{3^{2x}-3^{x+y}}$. A) $3^x - 3^y$ B) $3^x + 3^y$ C) $1 + 3^{y-x}$ D) $1 - 3^{x+y}$ E) $1 - 3^{y-x}$
Metacognitive Knowledge: Knowledge of cognition in general, and awareness of one's own cognition, beliefs, and thinking, in particular	Not Applicable
Cognitive Process: a continuum of increasing cognitive complexity	
Remember: Retrieve relevant knowledge from long-term memory. Action verbs include list, label, define, describe, show, locate, cite, underline, tabulate, identify, and name.	**Item 12 (2000)**: Let $\sqrt[3]{2\sqrt[5]{x}} = \sqrt[3]{2}\sqrt[5]{3}$. What is x? A) 3^3 B)3^4 C) 3^6 D) 2^7 E) 2^8
Understand: Determine the meaning of instructional messages, including verbal, symbolic, algebraic, and graphic representations. Action verbs involve clarify, represent, translate, illustrate, categorize, abstract, generalize, predict, contrast, map, match, and model.	**Item 25 (1999)**: How many integers are there which satisfy the inequality $\lvert x+2 \rvert \leq 4$? A) 13 B) 9 C) 8 D) 7 E) 6
Apply: Carry out a procedure or use a technique in a given situation; use information in concrete situations or in another familiar situation. Action verbs include calculate, compute, solve, illustrate, demonstrate, and draw.	**Item 20 (2011)**: Given $f(x) = 3x - 6$ and $g(x) = (x-2)^2$. Find $(gof^{-1})(x)$. A) $\frac{3x^2}{2} - 1$ B) $(3x+4)^2$ C) $x^2 - 4x + 2$ D) $\frac{x^2}{9}$ E) $(3x-8)^2$
Analyze: Break material/information into its constituent parts and determine the relationship among the parts; explore understandings about how the parts relate to an overall structure or purpose. Action verbs consist of discriminate, distinguish, focus, select, find coherence, integrate, outline, parse, structure, deconstruct, interrogate, compare, contrast, and organize.	**Item 28 (2012)**: A design is generated by shading some of the squares on a 4x100 graph paper as is illustrated in the figure below.  In this design, the squares which correspond to columns with the multiples of 2 in row A, the multiples of 3 in row B, the multiples of 4 in row C, and the multiples of 5 in row D are shaded. Accordingly, in how many of the columns are the squares at the rows A and D shaded, and the others unshaded? A) 3 B) 4 C) 5 D) 6 E) 7
Evaluate: Make judgments about the value of a process or product based on criteria; detect the appropriateness of a procedure for a given problem. Action verbs include test, judge, experiment, defend, decide, support, and justify.	Not Applicable
Create: Put elements and ideas together to form a coherent/ functional whole; come up with hypotheses based on criteria; invent a product. Action verbs involve hypothesize, design, construct, invent, plan, produce, revise.	Not Applicable

Table 2. Descriptions of MATH Taxonomy Categories along with Specimen University Examination Mathematics Items.

Categories	Item (Year)
Factual Knowledge: Recalling prior knowledge/information; remembering a specific formula, definition, and proof.	**Item 42 (1998)**: What is the symmetry of the line $y = 2x - 1$ about the point $A(\frac{1}{2}, 3)$? A) $y = -\frac{1}{2}x + 3$ B) $y = \frac{1}{2}x + 1$ C) $y = -2x + 3$ D) $y = 2x + 1$ E) $y = 2x + 5$
Comprehension: Recognize examples and counterexamples; decide whether or not conditions of a simple definition are satisfied; understand the significance of the symbols in a formula; make substitutions in a formula.	**Item 2 (2013)**: An isosceles triangle has an apex angle measure α and one of the base angles measure $\sqrt{\beta}$, and $\sqrt{\sin\beta < \sin\alpha}$. Which one of the following is true? A) $\sqrt{0^\circ < \alpha < 30^\circ}$ B) $\sqrt{30^\circ < \alpha < 45^\circ}$ C) $\sqrt{45^\circ < \alpha < 60^\circ}$ D) $\sqrt{0^\circ < \alpha < 60^\circ}$ E) $\sqrt{60^\circ < \beta < 90^\circ}$
Routine Use of Procedures: Carry out algorithm steps; drill and practice.	**Item 39 (2013)**: Evaluate $\lim_{x\to\infty} \frac{e^{-3x}+e^{2x}}{lnx+3e^{2x}}$. A) $\frac{1}{2}$ B) $\frac{3}{2}$ C) $\frac{1}{3}$ D) 0 E) 1
Information Transfer: Transform information from one form to another; decide whether or not conditions of a conceptual definition are satisfied; recognize the applicability of a formula/method in different and unusual contexts; recognize the inapplicability of a formula/method in a context; summarize mathematics in "non-technical terms" for a different audience or paraphrase; generate a mathematical argument from a verbal outline of the method; explain mathematical processes; explain relationships between parts of the material; reorganize the parts of the mathematical argument in a logical order.	**Item 27 (1998)**:  Given above the graph of $f(x)$ and $g(x)$. Use the information in the graph to find $\frac{g(1)+(fog)(2)}{f(4)}$. A) $-\frac{1}{2}$ B) -1 C)0 D) 1 E) $\frac{1}{2}$
Application in New Situations: Model real life situations; prove a theorem using nonroutine procedures; extrapolate known/familiar procedures to new/unfamiliar situations; choose and apply appropriate techniques; choose and apply appropriate algorithms.	**Item 40 (2000)**:  The point A on the lower base of a right cylindrical box with a radius of 5 cm, height 24, and point B on its upper base is on the same vertical line. What is the path taken by an ant going from A to B in the shortest way to B, by moving from A as shown in the figure and making a single circulation only on the lateral surface of the box? A) 26π B) 25π C) $24\sqrt{2}\pi$ D) $25\sqrt{3}$ E) $25\sqrt{2}$

(*Continued*)

Table 2. (Continued).

Categories	Item (Year)
Justifying and Interpreting: Prove a theorem for justifying a result, method, or model; find errors in reasoning; recognize the limitations and the appropriateness of a model; recognize the computational limitations and sources of error; interpret a result, method, or model; discuss the significance of examples and counterexamples; recognize the unstated assumptions.	**Item 17 (2012)**: A student made a mistake proving the following claim which he/she thought was right: *Claim*: Let A, B, and C any sets, $A \setminus (B \cap C) \subseteq (A \setminus B) \cap (A \setminus C)$. *Proof*: If I show that each element of the set $A \setminus (B \cap C)$ is also an element of the set $(A \setminus B) \cap (A \setminus C)$, then the proof is valid. Now take $x \in A \setminus (B \cap C)$. (I) Then $x \in A$ and $x \notin (B \cap C)$. (II) So $x \in A$ and $(x \notin B \text{and} x \notin C)$. (III) Now $(x \in A$ and $x \notin B)$ and $(x \in A$ and $x \notin C)$. (IV) So $x \in (A \setminus B)$ and $x \in (A \setminus C)$. (V) Then $x \in [(A \setminus B) \cap (A \setminus C)]$. In which of the numbered steps did the student make an error? A) I B) II C) III D) IV E) V
Implications, Conjectures, and Comparisons: Make conjectures based on inductive or heuristic arguments; prove conjectures by rigorous methods; compare within and among algorithms; deduce the implications of a given result; construct examples and counterexamples.	**Item 18 (2012)**: A geometric drawing is made following the steps below: (i) Draw two parallel lines d_1 and d_2, which are 2 units away from each other. (ii) Take a point A on d_1 and draw a circle with center A and radius 3*cm*. Let B and C be the points where this circle cuts the line d_2. (iii) Draw a circle with center C and radius $\|BC\|$. Let D and E be points where this circle cuts the line d_1. According to this drawing, what is the distance between the points D and E? A) 5 B) 6 C) 7 D) 8 E) 9
Evaluation: make judgments; argue the merits of an algorithm coherently; use creative thinking to restructure the information into a new whole and develop implications.	Not Applicable

administer entrance examinations that select and place students in higher education institutions of their choice by taking into account (i) students' scores in single- or two-stage entrance examinations, (ii) students' high school grade point averages, (iii) their personal preferences of higher education programs, (iv) universities' capacities/quotas, and (v) the prerequisites of university programs (Measurement, Selection and Placement Center [MSPC], 2019). The number of high school students taking exams increases each year, helping to meet the growing needs of the country's economy.

Similar to entrance examinations in Japan, England, France, and Germany, the Turkish university entrance examination is based on a curriculum established by the Ministry of National Education [MoNE]. It is designed to measure students' academic achievement in secondary school. Under the system of MSPC, students take tests in nine subjects in five academic fields (Turkish Literature, mathematics, science, social sciences, and foreign language) to enter university. While the examinations in Japan, England, France, and Germany typically include open-ended items that require application of knowledge and oral items that require students to express themselves verbally (Stevenson, Harold, & Lee, 1997), the examinations in Turkey rely solely on multiple-choice items. In contrast to the U.S., Japan, England, and France, in Turkey and Germany no attention is paid to the particular abilities (e.g., essay, student's demonstrated interest, counselor and teacher recommendations, class rank, and extracurricular activities) of students when selecting them for admission to university. The Turkish university entrance examination parallels college-entrance testing in the U.S. (e.g., the SAT), where approved calculators are allowed for some items and the top factors in the entrance are overall high school GPA and admission test scores. Here, it is important to note that, similar to Japan, the Turkish higher education system has been highly centralized. Like their counterparts in other countries, Turkish students make their preferences known upon receiving their examination scores.

Starting in 2005, three major reform movements were carried out in Turkish mathematics teaching giving attention to higher-order thinking skills. This, in turn, led to changes to university

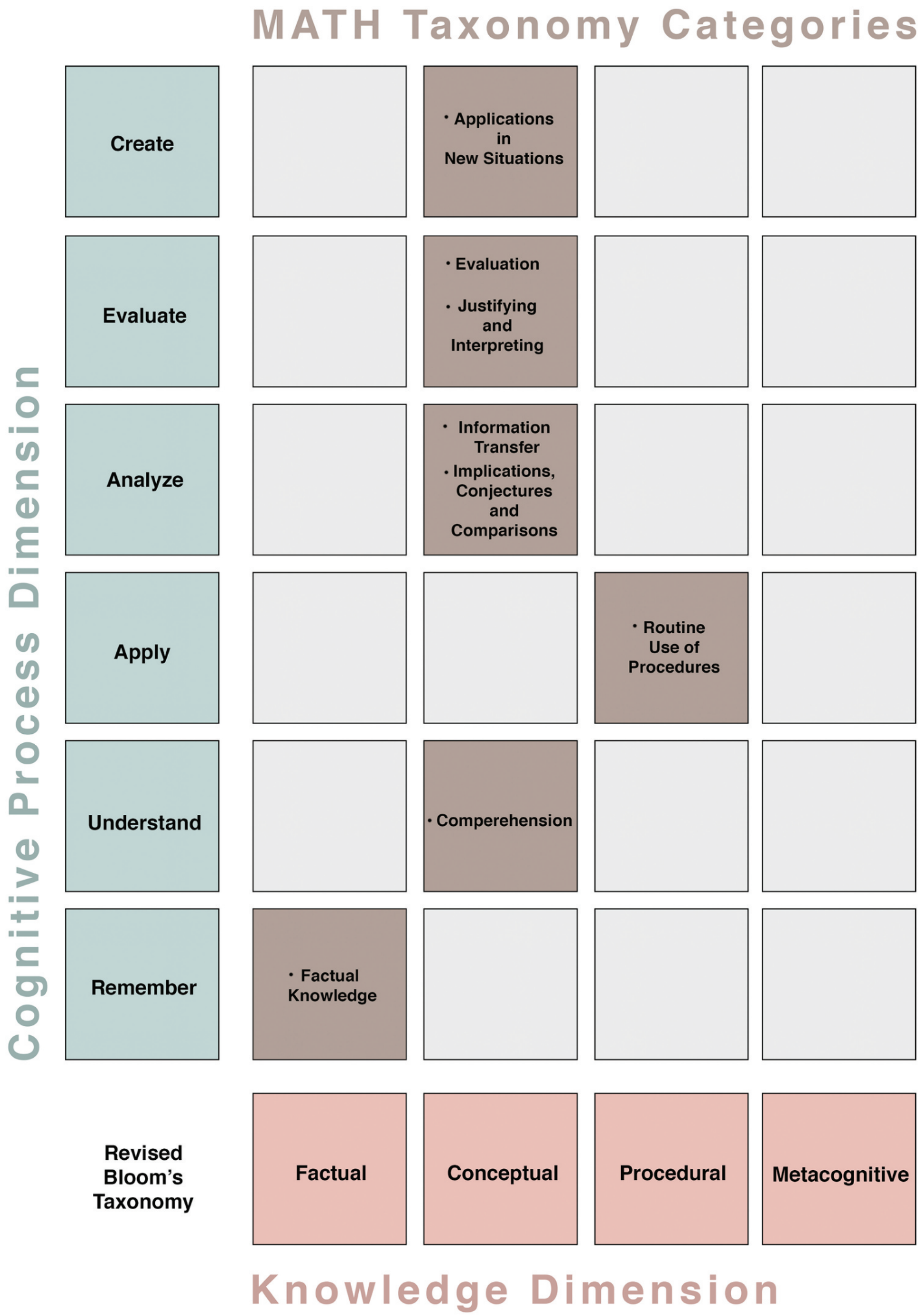


Figure 1. The Matrix of Relationships among Dimensions of the Revised Bloom's Taxonomy and Categories of the MATH Taxonomy.

entrance examinations. The reform movement shares commonalities with the reform movement in the U.S. (Şahin, Isiksal, & Ertepınar, 2010). In the mid-2000s and 2010s, reform curricula were developed with the goal of increasing the cognitive demand of educational objectives (MoNE, 2018). In contrast to large-scale international assessments built on a theoretical framework, the

Turkish university entrance examinations are based on a content analysis of the intended curriculum of secondary education; that is, an analysis of what students are expected to have learned as a result of exposure to a prescribed curriculum. This is typically done in a table of specifications with learning objectives and content area.

This study focuses on the analysis and comparison of the distribution of dimensions/categories in the university entrance examinations and mathematics curricula. The research questions are as follows: (1) Which dimensions/categories of the Revised Bloom and MATH taxonomies are reflected and tested by the university entrance examination mathematics items across 1998–2013? (2) What are the trends in dimensions/categories of the taxonomies placed in university entrance examination mathematics items across pre-reform (1998–2004) and first reform (2005–2013) periods? and (3) To what extent do university entrance examination mathematics items align with mathematics curricula objectives in terms of reflecting the dimensions/categories of the taxonomies across time periods (2005, 2011, and 2013)?

The significance of the present study is three-fold. First, it provides a general portrayal of Turkish university entrance examinations. Second, it discerns the diversity of university entrance examinations in different reform periods. Third, it gives significant information about the alignment of university entrance examinations and mathematics curricula in certain years after the reform.

Method

Data source and design

The mid-2000s to the early 2010s was a period of major changes in the Turkish education system. Indeed, considered a driving force of change, the dimensions/categories of taxonomies were introduced on a large scale in all elementary and secondary schools throughout the country in the national curricula. By applying a descriptive approach to quantitative content analysis (Neuendorf, 2002), we initially investigated the dimensions/categories of taxonomies reflected in Turkey's university entrance examinations and secondary mathematics curricula.

Content analysis is an extremely useful technique as a means of analyzing in an unobtrusive way (Fraenkel, Wallen, & Hyun, 2014, p. 489) and producing counts of key categories (Fink, 2009). The exam booklets were downloaded from the online archive (http://www.osym.gov.tr/TR,15045/osys-cikmis-sorular.html) with permission of the MSPC for the years 1998 to 2013. Since the mathematics curricula were changed several times by the MoNE between 1998 and 2013, the only available curricula were for the years 2005, 2011, and 2013, which captures the first reform period (2005–2013). The mathematics curricula were downloaded from the online archive (https://ttkb.meb.gov.tr) of the MoNE.

Coding procedure

The unit of analysis for coding each university entrance examination and mathematics curriculum was an item and for the mathematics curriculum as an educational objective. In order to investigate the trend across reform periods, university entrance examination items from 1998 to 2013 were grouped into before (1998–2004) and after (2005–2013) reform periods. Moving to the educational objectives, to answer the third research question, coding schemes for curriculum year, dimensions of the Bloom's Revised Taxonomy, and categories of the MATH Taxonomy were determined. The three mathematics curricula were from 2005, 2011, and 2013 with educational objectives ($N = 621$) from 2005 ($N = 234$), 2011 ($N = 203$), and 2013 ($N = 184$). The university entrance examinations were from eight booklets 2005 ($N = 2$), 2011 ($N = 3$), and 2013 ($N = 3$) containing mathematics items ($N = 285$) from 2005 ($N = 45$), 2011 ($N = 120$), and 2013 ($N = 120$).

To determine each coding scheme, two authors (i.e., the coders) reviewed the literature, analyzed each of the mathematics items and educational objectives completely, and discussed these until consensus about dimensions/categories of items and objectives was reached. After the coding schemes were determined, we blind-coded independently the mathematics items and educational objectives. Where there was disagreement among the codings, we discussed them until consensus was reached. All of these coding categories were disjointed each other. What follows are the coding schemes used in the study and the inter-coder agreement.

Exam year

The exam year of each mathematics item ($N = 1077$) in 28 electronic booklets was coded into the years from 1998 to 2013: 1998 (coded as 1), 1999 (coded as 2), . . . , 2013 (coded as 16).

Reform period

The reform period included two categories of exam years (1) before and (2) after the reform movement took place. Exam years before the reform (1998–2004) included pre-reform (coded as 1) and after the reform (2005–2013) included first reform (coded as 2).

Curriculum year

The curriculum year of each educational objective ($N = 621$) in three available curricula included 2005 ($N = 234$; coded as 1), 2011 ($N = 203$; coded as 2), and 2013 ($N = 184$; coded as 3).

Revised bloom's taxonomy

Adapted from Anderson and Krathwohl (2001), the dimension of each mathematics item and educational objective included knowledge (coded as 1) and cognitive process (coded as 2). The sub-dimension of each mathematics item and educational objective categorized in the knowledge dimension included factual, conceptual, procedural, and metacognitive, which were coded as 1, 2, 3, and 4, respectively. The sub-dimension of each mathematics item categorized in the cognitive process dimension included remember, understand, apply, analyze, evaluate, and create, which were coded from 5 to 10. We considered each mathematics item and educational objective, attending specifically to their potential relationships with dimensions/sub-dimensions regarding three processes based on a comprehensive literature review: (1) identifying the operational definitions of; (2) determining action verbs related to; and (3) analyzing sample items relevant to the dimensions/sub-dimensions. We take the two dimensions and ten sub-dimensions as mutually exclusive rather than overlapping. That is, each mathematics item and educational objective received a single code with respect to the relevant dimension/sub-dimension, which yields four codes in total, two codes (e.g., Item1: knowledge dimension and procedural sub-dimension; cognitive process dimension and apply sub-dimension). When an item or objective called for more than one type of mathematical activity, for example knowing a particular definition (*factual*) and knowing how to use this definition in computations (*procedural*), these processes were not approached as separate units but as one single segment by taking into account the item stem with regards to the abovementioned three processes. We agreed to select the code that best described the majority of the mathematics item and educational objective (i.e., the predominant dimension/sub-dimension). In this way, it was possible to differentiate between mathematics items as well as educational objectives, which only required low-level thinking from those which prompted thinking more extensively.

MATH taxonomy

Adapted from Smith et al. (1996), the category of each mathematics item included factual knowledge; comprehension; routine use of procedures; information transfer; application in new situations; justifying and interpreting; implications, conjectures, and comparisons; and evaluation. For each of the eight categories, a code was determined from 1 to 8. We consider each mathematics item and educational objective, attending specifically to their potential relationships with categories in the same way that we did for categorizing the dimensions/sub-dimensions of the Revised Bloom's Taxonomy with regards to the three processes mentioned above. In a similar vein, all the eight were disjoint from each other. Whenever it was noticed that the categories were not totally exclusive and that a single item or objective may be classified into more than one category, decisions about interpretation were made collaboratively through discussion. We then confirmed whether the item/objective fitted exactly into the categorization. If inconsistencies arose, we sought consensus. For example, an item "Given the graph of $f(x)$ above, what is the sum of $\lim_{x \to a^+} f(x) + \lim_{x \to b^-} f(x) + \lim_{x \to c^+} f(x)$?" indicates graph reading and therefore fits the category of "information transfer." However, it also expresses visualizing one-sided limits on the graph and then implementing addition procedures, thus fitting the category of "routine use of procedures." Consequently, this item was coded as "information transfer" considering the *meta-level* mathematical higher-order thinking skill that it triggers.

Inter-coder agreement

The two authors separately coded all items using the coding system. Inter-coder agreement, which was indexed by κ (kappa), was high for both the Revised Bloom's Taxonomy (κ= .88) dimensions and the MATH Taxonomy (κ= .91) categories. This indicated satisfactory inter-coder agreement according to recommendations by Landis and Koch (1977). To gather additional inter-coder reliability evidence, two external coders (an expert in mathematics education and an expert in test development) separately coded 30% (N = 377) of the randomly selected items after being trained on the coding system and completing 15 practice items. Two different analytical techniques were used. Firstly, we looked at the percentage of items in which the internal and external coders had matching codings (92% for the Revised Bloom's Taxonomy and 95% for the MATH Taxonomy). Secondly, we computed κ (kappa) to assess the inter-coder agreement between the coding of the two sets of coders. Agreement remained high throughout the coding process for both the Revised Bloom's Taxonomy (κ= .86) dimensions and the MATH Taxonomy (κ = .91) categories. In the same vein, the two authors coded each objective (N = 621) independently. The initial inter-coder agreement was .84 and .79 for the Revised Bloom's Taxonomy dimensions and the MATH Taxonomy categories, respectively. An additional test of inter-coder agreement on 10% of the objectives (N = 62) was performed with the external coder, who was an expert in mathematics education. The inter-coder agreement was .88 and .92 for the Revised Bloom's Taxonomy dimensions and the MATH Taxonomy categories, respectively. All these values indicate satisfactory agreement (Landis & Koch, 1977) among the three coders. Throughout this process, the coders resolved the few discrepancies and disagreements through discussion and established consensus.

Data analyses

Statistical analyses were performed in 4 interrelated stages: (1) classify mathematics items with regard to the Revised Bloom's Taxonomy knowledge and cognitive process; (2) classify mathematics items with regard to the categories of knowledge and skills in the MATH Taxonomy; (3) identify the distribution of the dimensions/categories of the taxonomies reflected and tested by the university entrance examination across reform periods; and (4) identify the distribution of the dimensions/categories of the taxonomies reflected in the objectives of the mathematics curricula. Along with appropriate identification codes (e.g., year, session, reform period), codes for the categories described earlier under "Coding procedure" were entered into the IBM Statistical Package for Social Sciences 21.0 (SPSS, 2012) database.

Initial analysis reported the descriptive statistics – frequencies and percentages – for each possible code across all 1,077 mathematics items by 16 years (*Stages 1, 2, and 3*) and across all 621 objectives by 3 years (*Stage 4*). Percentages and their statistical significance were checked using the bootstrapping method with a chi-square test of independence (*Stages 3 and 4*), in which 5,000 random samples are generated with 95% bias-corrected confidence intervals. These intervals should not include zero for a significant descriptive statistics (LaFlair, Egbert, & Plonsky, 2015).

Results

Distribution of mathematics items across the revised bloom's taxonomy dimensions and the MATH taxonomy categories

Table 3 presents the distribution of mathematics items across the Revised Bloom's Taxonomy dimensions and the MATH Taxonomy categories. Specifically, in terms of the knowledge dimension, the results indicate that 69.8% of the items reflect procedural knowledge ($N = 752$) whereas 28.7% reflect conceptual knowledge ($N = 309$), followed by 1.5% factual knowledge ($N = 16$). In a related vein, with respect to the cognitive process dimension, the results show that the majority of items reflect apply (62.7%, $N = 675$), while 23.4% reflect analyze ($N = 252$), 8.6% remember ($N = 93$), and 5.3% understand ($N = 57$). None of the items show metacognitive knowledge, evaluate, or create.

In the distribution of mathematics items across the MATH Taxonomy categories, the results suggest that the majority of the items reflect routine use of procedures, followed by information transfer and, to a lesser extent, comprehension. More specifically, 66.6% of the items mirror routine use of procedures ($N = 717$), 15.4% information transfer ($N = 166$), and 14% comprehension ($N = 151$). On the other hand, 2.3% of the items represent application in new situations ($N = 25$), 1.1% factual knowledge ($N = 12$), 0.5% implications, conjectures, and comparisons ($N = 5$), and 0.1% justifying and interpreting ($N = 1$). None of the items reflect evaluation.

Distribution of university examination mathematics items classified across the dimensions/categories of the taxonomies by reform periods

The bootstrapped results of the percentages and the bias-corrected (BC) confidence intervals with 95% confidence are illustrated in Table 4. Findings revealed that the BC confidence intervals of the Revised

Table 3. Distribution of Dimensions/Categories of Taxonomies in University Entrance Examination Mathematics Items.

	Dimension/Category	Frequency	Percent
Revised Bloom's Taxonomy	Knowledge		
	Factual	16	1.5
	Conceptual	309	28.7
	Procedural	752	69.8
	Metacognitive	0	0
	Cognitive Process		
	Remember	93	8.6
	Understand	57	5.3
	Apply	675	62.7
	Analyze	252	23.4
	Evaluate	0	0
	Create	0	0
	Total	1077	100
MATH Taxonomy	Factual knowledge	12	1.1
	Comprehension	151	14
	Routine use of procedures	717	66.6
	Information Transfer	166	15.4
	Application in new situations	25	2.3
	Justifying and interpreting	1	.1
	Implications, conjectures, and comparisons	5	.5
	Evaluation	0	0
	Total	1077	100

Table 4. Percentages and Bootstrapped Chi-Square Results for Dimensions/Categories of Taxonomies in University Entrance Examination Mathematics Items ($N = 1077$) across Reform Periods.

Taxonomies	1998–2004 pre-reform %	2005–2013 first reform %	χ^2 *Value*	*df*	*p*	*95% CI*
Revised Bloom's Taxonomy						
Factual Knowledge	1.9	1.3				
Conceptual Knowledge	26.9	29.4				
Procedural Knowledge	71.2	69.3	1.56	2	.56	[.002, .146]
Metacognitive Knowledge	0	0				
Remember	8.3	8.8				
Understand	7.4	4.4				
Apply	69.2	60				
Analyze	15.1	26.8	19.69	3	.00*	[.073, .214]
Evaluate	0	0				
Create	0	0				
Total	100	100				
MATH Taxonomy						
Factual Knowledge	2.2	.7				
Comprehension	17.6	12.5				
Routine Use of Procedures	69.9	65.2				
Information Transfer	7.1	18.8	30.57	6	.00*	[.104, .257]
Application in New Situations	2.6	2.2				
Justifying and Interpreting	0	.1				
Implications, Conjectures, and Comparisons	.6	.4				
Evaluation	0	0				
Total	100	100				

*$p < .05$.
Note. CI = confidence interval.

Bloom's Taxonomy knowledge and cognitive dimensions, and the MATH Taxonomy categories were significant (95% BC CIs [.002, .146], [.073, .214] and [.104, .257], respectively). This implied that the distribution of mathematics items classified by the dimensions/categories of the taxonomies in the pre-reform period was significantly different from the distribution displayed in the first reform period. Additionally, there appears to be an association between the distribution of the dimensions/categories of the taxonomies and reform periods ($\chi^2(2, N = 1077) = 1.56$, $p = .56$; $\chi^2(3, N = 1077) = 19.69$, $p < .05$ and $\chi^2(6, N = 1077) = 30.57$, $p < .05$).

A closer look at Table 4 in terms of the Revised Bloom's Taxonomy dimensions reveals that the percentage of the *factual knowledge* and *procedural knowledge* items in the knowledge dimension slightly decreased from pre-reform (1.9% and 71.2%, respectively) to first reform (1.3% and 69.3%, respectively), whereas the percentage of *conceptual knowledge* items slightly increased from pre-reform (26.9%) to first reform (29.4%). With regards to the cognitive process dimension, the percentage of *understand* items slightly decreased from pre-reform (7.4%) to first reform (4.4%), and the percentage of *apply* items decreased from pre-reform (69.2%) to first reform (60%). Although there was a negligible increase in the *remember* items from pre-reform (8.3%) to first reform (8.8%), there was a substantial increase in the percentage of analyze items from pre-reform (15.1%) to first reform (26.8%).

Similarly, regarding the MATH Taxonomy, there was a decline in the percentages of *factual knowledge* (2.2% – 0.7%), *comprehension* (17.6% – 12.5%), and *routine use of procedures* (69.9% – 65.2%) items over the reform periods. While the percentage of *application* in new situations (2.6% – 2.2%) and *implications, conjectures and comparisons* (0.6% – 0.4%) slightly decreased, there was a sharp increase in the percentage of *information transfer* items from pre-reform (7.1%) to first reform (18.8%). As shown in Figure 2, although no items were related to *metacognitive knowledge, evaluate*, or *create* in the Revised Bloom's Taxonomy or *evaluate* in the MATH Taxonomy, there appears to be a trend toward increasing higher-order thinking across the two reform periods.

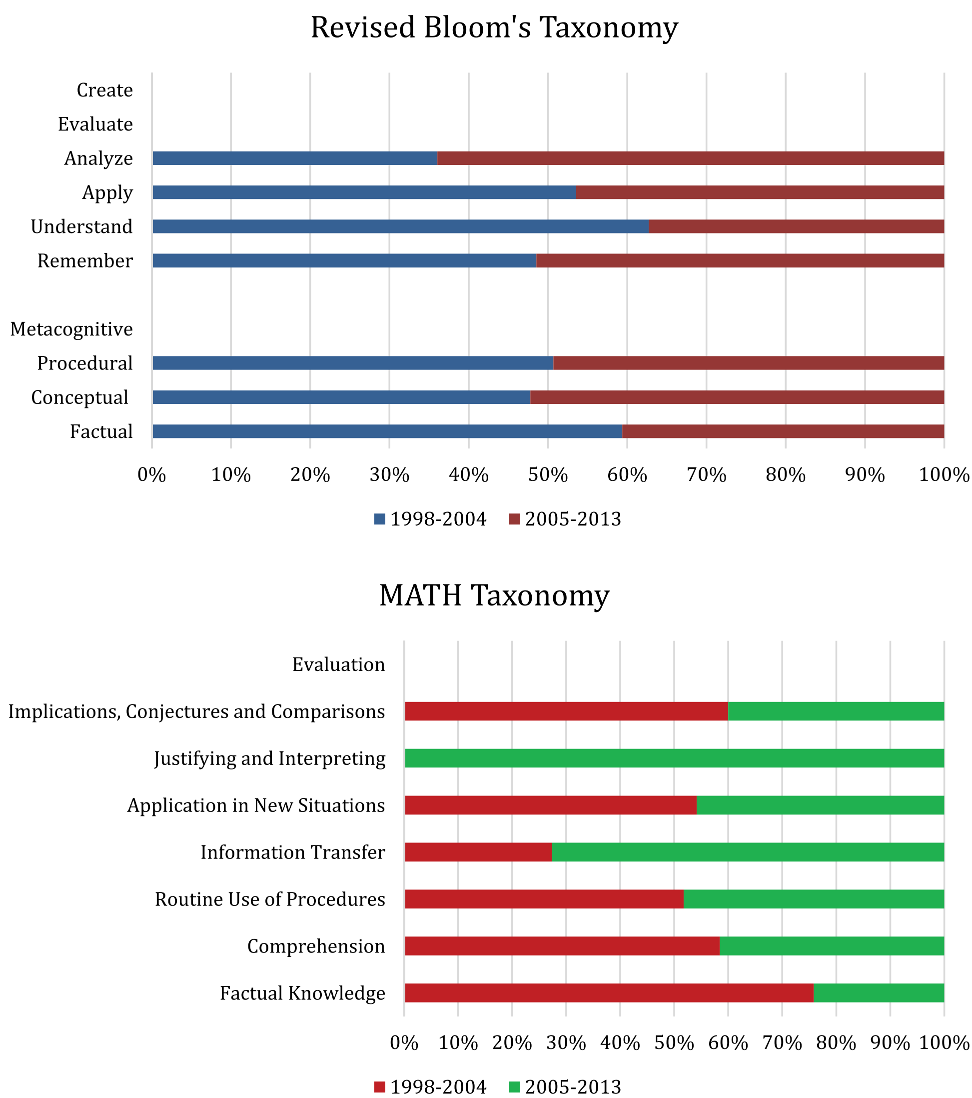


Figure 2. Trends in University Entrance Examination Mathematics Items Across Reform Periods Stratified by Dimensions/Categories of Taxonomies.

Distribution of mathematics curriculum objectives and university examination mathematics items classified across the dimensions/categories of the taxonomies by time period

The bootstrapped results of the percentages and the bias-corrected confidence intervals with 95% confidence are illustrated in Table 5. According to the bootstrapped results, the percentage distributions of the Revised Bloom's Taxonomy knowledge and cognitive dimensions and the MATH Taxonomy categories were significant (95% BC CIs [.058, .273], [.141, .373] and [.259, .488], respectively), providing evidence that the distribution of educational objectives classified by the dimensions/categories of the taxonomies differed significantly across the years 2005, 2011, and 2013. There appears to be an association between distribution of the dimensions/categories of the taxonomies and time periods ($\chi^2(4, N = 621) = 12.85, p < .05$;$\chi^2(10, N = 621) = 34.38, p < .05$; and $\chi^2(14, N = 621) = 75.44, p < .05$, respectively).

Table 5. Percentages and Bootstrapped Chi-Square Results of Taxonomies Reflected in Mathematics Curricula Objectives ($N = 621$) and University Examination Items ($N = 285$) by Time Period.

	Mathematics curricula							University examinations						
	Years			χ^2				Years			χ^2			
	2005	2011	2013					2005	2011	2013				
Taxonomies	%	%	%	Value	*df*	*p*	95%CI	%	%	%	Value	*df*	*p*	95%CI
Revised Bloom's Taxonomy														
Factual	20.1	20.7	15.8					0	2.5	0				
Conceptual	31	43.3	32.1					31.1	26.7	41.7				
Procedural	48.3	36	52.2	12.85	4	.00*	[.058, .273]	68.9	70.8	58.3	19.56	4	.00*	[.115, .271]
Metacognitive	0	0	0					0	0	0				
Total	100	100	100					100	100	100				
Remember	12.8	7.9	4.3					11.1	12.5	4.2				
Understand	24.4	30.5	31					6.7	5	4.2				
Apply	60.3	57.6	52.2					53.3	63.3	55.8				
Analyze	1.7	3	7.6	34.38	10	.00*	[.141, .373]	28.9	19.2	35.8	24.74	6	.00*	[.115, .349]
Evaluate	0.9	1	2.7					0	0	0				
Create	0	0	2.2					0	0	0				
Total	100	100	100					100	100	100				
MATH Taxonomy														
Factual Knowledge	12.8	7.9	3.8					0	0	0				
Comprehension	23.9	29.1	30.4					20	11.7	16.7				
Routine Use of Procedures	59	48.8	41.8					57.8	70.8	51.7				
Information Transfer	2.1	5.4	3.3	75.44	14	.00*	[.259, .488]	22.2	16.7	25	15.87	8	.00*	[.117, .425]
Application in New Situations	0.4	1.5	10.9					0.4	0.8	5				
Justifying and Interpreting	0.4	2.5	2.2					0	0	0				
Implications, Conjectures and Comparisons	1.3	4.9	4.9					0	0	1.7				
Evaluation	0	0	2.7					0	0	0				
Total	100	100	100					100	100	100				

*$p < .05$.

Note. CI = confidence interval. Percentages of the objectives are shown for 2005 ($n = 234$), 2011 ($n = 203$), and 2013 ($n = 184$). Percentages of the items are shown for 2005 ($n = 45$), 2011 ($n = 120$), and 2013 ($n = 120$).

A closer look at Table 5 in terms of the Revised Bloom's Taxonomy dimensions reveals a declining trend in *factual knowledge, remember*, and *apply objectives* over 2005 (20.1%, 12.8%, and 60.3%, respectively), 2011 (20.7%, 7.9%, and 57.6%, respectively), and 2013 (15.8%, 4.3%, and 52.2%, respectively). While there was a negligible increase in *procedural knowledge* (48.3%, 36%, and 52.2%, respectively) and *understand* (24.4%, 30.5%, and 31%, respectively) objectives, there was an increase in *analyze*, *evaluate*, and *create objectives* over 2005 (1.7%, .9%, and 0%, respectively), 2011 (3%, 1%, and 0%, respectively), and 2013 (7.6%, 2.7%, and 2.2%, respectively). Regarding the MATH Taxonomy categories, a similar declining trend was observed in *factual knowledge* (12.8%, 7.9%, and 3.8%, respectively), *routine use of procedures* (59%, 48.8%, and 41.8%, respectively) objectives over the years. Although the percentage of *comprehension objectives* slightly increased (23.9%, 29.1%, and 30.4%, respectively), the increase was noteworthy for *information transfer* (2.1%, 5.4%, and 3.3%, respectively), application in new situations (0.4%, 1.5%, and 10.9%, respectively), *justifying and interpreting* (0.4%, 2.5%, and 2.2%, respectively), *implications, conjectures, and comparisons* (1.3%, 4.9%, and 4.9%, respectively), and *evaluation* (0%, 0%, and 2.7%, respectively) objectives across the years.

Moving to the university entrance examination items, the bootstrapped results showed that the percentage distributions of the Revised Bloom's Taxonomy knowledge and cognitive dimensions and the MATH Taxonomy categories were significant (95% BC CIs [.115, .271], [.115, .349] and [.117, .425], respectively). This implies that the distribution of objectives classified by the dimensions/categories of the taxonomies differed significantly across 2005, 2011, and 2013. There appears to be an association between distribution of the dimensions/categories of the taxonomies and time periods ($\chi^2(4,\ N = 285) = 19.56,\ p < .05; \chi^2(6,\ N = 285) = 24.74,\ p < .05$ and $\chi^2(8,\ N = 285) = 15.87,\ p < .05$, respectively).

The results revealed that there was a variation in the proportion of examination items classified across the Revised Bloom's Taxonomy in each dimension/category. The general trend was a decline in proportion of *factual knowledge* (0%, 2.5%, and 0%, respectively), *procedural knowledge* (68.9%, 70.8%, and 58.3%, respectively), *remember* (11.1%, 12.5%, and 4.2%, respectively), and *apply* (53.3%, 63.3%, and 55.8%, respectively) items and, to a lesser extent, for *understand* (6.7%, 5%, and 4.2%, respectively) items over the three years.Conversely, the percentage of *conceptual knowledge* and *analyze* items saw an overall increase from 2005 to 2013: 2005 (31.1% and 28.9%, respectively), 2011 (26.7% and 19.2%, respectively), and 2013 (41.7% and 35.8%, respectively). Regarding the MATH Taxonomy categories, while there was an overall decrease in *routine use of procedures* (57.8%, 70.8%, and 51.7%) items, there were increases in *application in new situations* (0%, 0.8%, and 5%), *implications, conjectures, and comparisons* (0%, 0%, and 1.7%), and *information transfer* (22.2%, 16.7%, and 25%, respectively).

Viewed together, the results for the distribution of taxonomies reflected in mathematics curricula objectives and university mathematics examination items can be visualized in one single graph. Figure 3 illustrates our results using a radar graph. The graph consists of 10 equiangular sub-dimensions of the Revised Bloom's Taxonomy on the right-hand side and 8 categories of the MATH Taxonomy on the left-hand side. A blue line was drawn connecting the data values (i.e., percentages) for educational objectives from the mathematics curricula of the years 2005, 2011, and 2013, while a red line was drawn for mathematics items from the university entrance examinations of the pre-reform (1998–2004) and first reform (2005–2013) periods.

The radar chart helped us to easily examine the relative values for each dimension of the taxonomies in addition to locating similar or dissimilar results so that we can make relevant and useful visual comparisons. Turning first to the university entrance examinations, the figure clearly illustrates that there was a high percentage of mathematics items which reflect *procedural knowledge* and *apply* sub-dimensions in the knowledge and cognitive process dimensions of the Revised Bloom's Taxonomy. Similarly, there was a high percentage of mathematics items which display *routine use of procedures* category of the MATH Taxonomy. The figure, on the contrary, illustrates that there was a low percentage of mathematics items, which reflect *factual* knowledge, *remember*, and *understand* sub-dimensions in the knowledge and cognitive process dimensions of the Revised Bloom's Taxonomy.

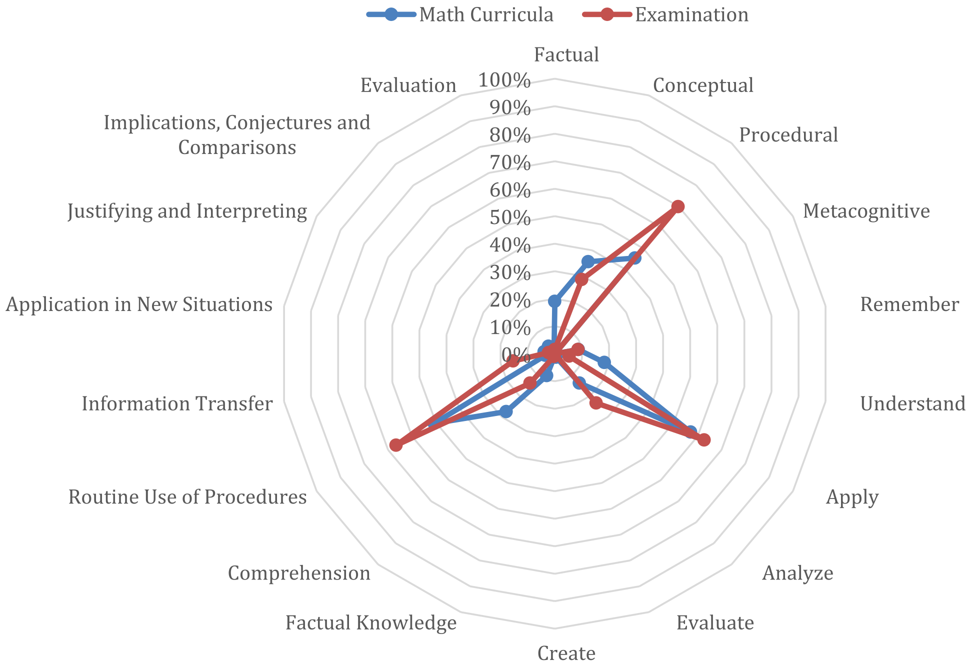


Figure 3. Radar Graph for the Distribution of Taxonomies Reflected in Mathematics Curricula Objectives and University Examination Mathematics Items: Percentages.

There was a low percentage of mathematics items which display *factual knowledge* or *application in new situations* category of the MATH Taxonomy. Unfortunately, no information existed about *metacognitive* knowledge, and *evaluate* and *create* sub-dimensions in the knowledge and cognitive process dimensions of the Revised Bloom's Taxonomy. No information existed about *evaluation* for the MATH Taxonomy.

Moving to the mathematics curricula, the figure showed that there was a high percentage of educational objectives from the mathematics curricula which reflect *procedural* knowledge, and *apply* sub-dimensions in the knowledge and cognitive process dimensions of the Revised Bloom's Taxonomy. Similarly, there was a high percentage of educational objectives which display *routine use of procedures* category of the MATH Taxonomy. The same figure, on the other hand, showed that there was a low percentage of educational objectives which reflect *factual* knowledge, and *evaluate* and *create* sub-dimensions in the knowledge and cognitive process dimensions of the Revised Bloom's Taxonomy. There was a low percentage of educational objectives which display *justifying and interpreting* or *evaluation* category of the MATH Taxonomy.

Discussion

The work in this paper adds to the literature on mathematics education by detailing for the first time how the Revised Bloom's Taxonomy can be used in conjunction with its adapted version the MATH Taxonomy to explore university entrance examinations. The findings show that mathematics items most frequently reflected *procedural knowledge* and *apply* sub-dimensions of the Revised Bloom's Taxonomy while they mirrored *routine use of procedures* category of the MATH Taxonomy. Additionally, none of the items displayed higher-order skills: neither *evaluate* nor *create* from the Revised Bloom's Taxonomy nor *evaluation* from the MATH Taxonomy.

The parallels between curricula and university entrance examinations might suggest that the two taxonomies would be a worthwhile pursuit for future research. For example, the two taxonomies might be used to tease apart aspects of assessment and curricular design and/or align educational objectives in the curricula, teaching activities in textbooks, and cognitive demands of items in assessments.

Given the fact that preparation of university examinations in Turkey is not organized in a theoretical framework, our findings provide a lens on the degree to which the university entrance examinations and mathematics curricula are aligned with one another. Curriculum reform has long been a line of scholarly inquiry (Howson, Keitel, & Kilpatrick, 1981) and is often viewed as a powerful tool for improving teaching, learning, and instructional materials (Cai & Howson, 2013) including assessments (Stanick & Kilpatrick, 1992). Reform-guided mathematics curricula in Turkey have put great emphasis on measuring complex cognitive skills and students' higher-order thinking. Turkish test developers have made some effort to integrate items that require high cognitive demand into national examinations at all levels of education. Although our data shows that the Turkish university entrance examinations do contain higher-order items (e.g., analyze, information transfer), the percentage of such items in each of the examinations is still quite low even following the reform movements. A related finding is also evident in the distribution of the dimensions/categories of taxonomies in university entrance examinations and curricula over time periods. Our findings clearly indicate that although there was a decline in factual knowledge objectives/items, there was no significant change in procedural knowledge, apply, and routine use of procedures.

Our findings therefore reflect the national (Dursun & Çoban, 2006) and international (Brown, 2010; Drijvers, Kodde-Buitenhuis, & Doorman, 2019) work by researchers who show that there is a general tendency in examinations to mainly include items with a low level of cognitive demand (Levy & Murnane, 2004; Pettersen & Braeken, 2019). This is a point of concern, as the presence of mathematical thinking/reasoning is at the heart of the mathematics education reform movements (Palha, Dekker, & Gravemeijer, 2015) and is generally associated with problem posing, problem solving, and analyzing information (Wagner, 2014). Considering 21st century skills through the lens of the Revised Bloom's Taxonomy and the MATH Taxonomy, particular strengths can be observed in terms of developing analyzing, evaluating, and creating; and transferring information, justifying and interpreting. At this high end, these taxonomies are very much in alignment with the modern focus on 21st century skills. Essentially, university entrance examination is at its best when the actions taken by policymakers embody the learning objectives of the curriculum (Barnes, Clarke, & Stephens, 2000), which can be seen as an amalgam of instruction, content, materials, and assessment (Clarke, 1996).

In connection with the purpose of the university entrance examinations (i.e., to determine which students are prepared to enter higher education) and the particular importance of mathematics items, a question about the validity of university entrance examinations arises for future research: What is the degree to which mathematics items on specific dimensions/categories of each taxonomy can predict success in university persistence and/or degree completion? Researchers have theorized that alignment between test content and the intended curriculum influences students' test performance, and such an effect could also be detectable for test items (Traynor, 2017).

There are a number of noteworthy limitations to the present study. Firstly, university entrance examinations from 1998 to 2013 were the sole source of information because of permission restrictions. Secondly, only two taxonomies were used in the analyses, rather than including other frameworks (e.g., the SOLO). Lastly, think-aloud protocols with students, teachers, or test developers would allow a stronger argument to be made regarding, for instance, a comparison between the Turkish university entrance examination system and its equivalent in other countries.

In sum, successful change in university entrance examinations requires a fundamental change in curricula, classroom tests, and teaching methods (Gravemeijer, Stephan, Julie, Lin, & Ohtani, 2017; Wagner, 2014). This is only possible with the support of policymakers and other stakeholders including students, teachers, school principals, and parents.

Acknowledgments

The authors would like to thank the editors and anonymous reviewers who provided helpful feedback on earlier drafts. Their valuable comments led us to put our heart and soul into the challenge of improving the manuscript during the hard times of COVID-19 quarantine.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Utkun Aydin http://orcid.org/0000-0002-1380-5911
Bengi Birgili http://orcid.org/0000-0002-2990-6717

References

Airasian, P. W., & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice*, *41*(4), 249–254. doi:10.1207/s15430421tip4104_8

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.

Anderson, L. W. (2005). Objectives, evaluation, and the improvement of education. *Studies in Educational Evaluation*, *31*(2–3), 102–113. doi:10.1016/j.stueduc.2005.05.004

Anderson, L. W., & Krathwohl, L. R. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Longman Press.

Ari, A. (2011). Finding acceptance of bloom's revised cognitive taxonomy on the international stage and in Turkey. *Educational Sciences: Theory & Practice*, *11*(2), 767–772. https://files.eric.ed.gov/fulltext/EJ927376.pdf

Ball, G., Stephenson, B., Smith, G., Wood, L., Coupland, M., & Crawford, K. (1998). Creating a diversity of mathematical experiences for tertiary students. *International Journal of Mathematical Education in Science and Technology*, *29*(6), 827–841. doi:10.1080/0020739980290605

Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: The engine of systemic curricular reform? *Journal of Curriculum Studies*, *32*(5), 623–650. doi:10.1080/00220270050116923

Bennie, K. (2005). The MATH taxonomy as a tool for analysing course material in mathematics: A study of its usefulness and its potential as a tool for curriculum development. *African Journal of Research in Mathematics, Science and Technology Education*, *9*(2), 81–95. doi:10.1080/10288457.2005.10740580

Bergqvist, E. (2007). Types of reasoning required in university exams in mathematics. *The Journal of Mathematical Behavior*, *26*(4), 348–370. doi:10.1016/j.jmathb.2007.11.001

Biggs, J. B., & Collis, K. F. (2014). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. The classification of educational goals. Handbook 1: Cognitive domain*. David McKay Press .

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD Press.

Brown, R. G. (2010). Does the introduction of the graphics calculator into system-wide examinations lead to change in the types of mathematical skills tested? *Educational Studies in Mathematics*, *73*(2), 181–203. doi:10.1007/s10649-009-9220-2

Cai, J., & Howson, A. G. (2013). Toward an international mathematics curriculum. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & K. S. F. Leung (Eds.), *Third international handbook of mathematics education research* (pp. 949–978). Springer Press.

Clarke, D. J. (1996). Assessment. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (pp. 327–370). Kluwer Press.

Davey, G., De Lian, C., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education*, *31*(4), 385–396. doi:10.1080/03098770701625761

Davis-Becker, S., & Buckendahl, C. W. (Eds.). (2017). *Testing in the professions: Credentialing policies and practice*. Routledge Press.

Drijvers, P., Kodde-Buitenhuis, H., & Doorman, M. (2019). Assessing mathematical thinking as part of curriculum reform in the Netherlands. *Educational Studies in Mathematics*, *102*(3), 435–456. doi:10.1007/s10649-019-09905-7

Dursun, Ş., & Çoban, A. (2006). Evaluation of geometry course in terms of high school programs and ÖSS questions. *Cumhuriyet University Social Sciences Journal*, *30*(2), 213–221. https://kutuphane.dogus.edu.tr/mvt/pdf.php

Fink, E. L. (2009). The FAQs on data transformation. *Communication Monographs*, *76*(4), 379–397. doi:10.1080/03637750903310352
Fraenkel, J., Wallen, N., & Hyun, H. (2014). *How to design and evaluate research in education* (9th ed.). McGraw-Hill Education Press.
Gravemeijer, K., Stephan, M., Julie, C., Lin, F. L., & Ohtani, M. (2017). What mathematics education may prepare students for the society of the future? *International Journal of Science and Mathematics Education*, *15*(1), 105–123. doi:10.1007/s10763-017-9814-6
Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*(5), 524–549. doi:10.2307/749690
Howson, G., Keitel, C., & Kilpatrick, J. (1981). *Curriculum developments in mathematics*. Cambridge University Press.
Incikabi, L. (2012). After the reform in Turkey: A content analysis of SBS and TIMSS assessment in terms of mathematics content, cognitive domains, and item types. *Education as Change*, *16*(2), 301–312. doi:10.1080/16823206.2012.745758
Jakwerth, P. (1999). TIMSS performance assessment results: United States. *Studies in Educational Evaluation*, *25*(3), 277–281. doi:10.1016/S0191-491X(99)00028-0
Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–264. doi:https://doi.org/10.1207/s15430421tip4104_2
Kuramoto, N., & Koizumi, R. (2018). Current issues in large-scale educational assessment in Japan: Focus on national assessment of academic ability and university entrance examinations. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 415–433. doi:10.1080/0969594X.2016.1225667
LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, t tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (1st ed., pp. 46–77). Routledge Press.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. doi:https://doi.org/10.2307/2529310
Lee, Y. J., Kim, M., Jin, Q., Yoon, H. G., & Matsubara, K. (2017). *East-Asian primary science curricula: An overview using revised Bloom's Taxonomy*. Springer Press.
Levy, F., & Murnane, R. J. (2004). Education and the changing job market. *Educational Leadership*, *62*(2), 80–84 (EJ716780. https://eric.ed.gov/?id=EJ716780
Marzano, R. J., & Kendall, J. S. (2007). *The new taxonomy of educational objectives* (2nd ed.). Sage Press.
Measurement, Selection and Placement Center [MSPC]. (2019). *About MSPC*. Retrieved from: https://www.osym.gov.tr/TR,8832/hakkinda.html
Ministry of National Education [MoNE]. (2018). *Mathematics curriculum (primary and middle school 1,23,4,567 and 8th grades)*. http://mufredat.meb.gov.tr/Dosyalar/201813017165445MATEMATİK%20ÖĞRETİM%20PROGRAMI%202018v.pdf
Moutsios-Rentzos, A., & Simpson, A. (2011). University mathematics students and exam-style proving questions: The AB-Δ strategy classification scheme. *International Journal for Mathematics in Education*, *3*, 45–64.
NCTM. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.
Neuendorf, K. A. (2002). *The content analysis guidebook*. Sage Press.
Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, *102*(1), 9–28. doi:10.1007/s10649-019-09903-9
Palha, S., Dekker, R., & Gravemeijer, K. (2015). The effect of shift-problem lessons in the mathematics classroom. *International Journal of Science and Mathematics Education*, *13*(6), 1589–1623. doi:10.1007/s10763-014-9543-z
Pettersen, A., & Braeken, J. (2019). Mathematical competency demands of assessment items: A search for empirical evidence. *International Journal of Science and Mathematics Education*, *17*(2), 405–425. doi:10.1007/s10763-017-9870-y
Porter, A. C. (2006). Curriculum assessment. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 141–159). Lawrence Erlbaum Press.
Radmehr, F., & Drake, M. (2018). An assessment-based model for exploring the solving of mathematical problems: Utilizing revised bloom's taxonomy and facets of metacognition. *Studies in Educational Evaluation*, *59*, 41–51. doi:10.1016/j.stueduc.2018.02.004
Radmehr, F., & Drake, M. (2019). Revised Bloom's taxonomy and major theories and frameworks that influence the teaching, learning, and assessment of mathematics: A comparison. *International Journal of Mathematical Education in Science and Technology*, *50*(6), 895–920. doi:10.1080/0020739X.2018.1549336
Resnick, L. B., Rothman, R., Slattery, J. B., & Vranek, J. L. (2004). Benchmarking and alignment of standards and testing. *Educational Assessment*, *9*(1&2), 1–27. doi:10.1080/10627197.2004.9652957
Rindermann, H., & Baumeister, A. E. (2015). Validating the interpretations of PISA and TIMSS tasks: A rating study. *International Journal of Testing*, *15*(1), 1–22. doi:10.1080/15305058.2014.966911
Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. doi:10.1007/s10648-015-9302-x

Ryan, K. E., Ryan, A. M., Arbuthnot, K., & Samuels, M. (2007). Students' motivation for standardized math exams. *Educational Researcher*, *36*(1), 5–13. doi:10.3102/0013189X06298001

Şahin, E., Isiksal, M., & Ertepınar, H. (2010). In-service elementary school teachers' beliefs in science teaching practices. *Hacettepe University Journal of Education*, *39*, 296–306. https://dergipark.org.tr/tr/pub/hunefd/issue/7799/102183

Smith, G., Wood, L., Coupland, M., Stephenson, B., Crawford, K., & Ball, G. (1996). Constructing mathematical examinations to assess a range of knowledge and skills. *International Journal of Mathematical Education in Science and Technology*, *27*(1), 65–77. doi:10.1080/0020739960270109

SPSS. (2012). *IBM SPSS statistics for windows, version 21*. International Business Machines Corporation.

Stanick, G. M., & Kilpatrick, J. (1992). Mathematics curriculum reform in the United States: A historical perspective. *International Journal of Educational Research*, *17*(5), 407–417. doi:10.1016/S0883-0355(05)80002-3

Stevenson, H., Harold, W., & Lee, S. (1997). *International comparisons of entrance and exit examinations: Japan, United Kingdom, France, and Germany*. U.S. Department of Education.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, *14*(3), 281–294. doi:10.1080/09695940701591867

Traynor, A. (2017). Does test item performance increase with test-to-standards alignment? *Educational Assessment*, *22* (3), 171–188. doi:10.1080/10627197.2017.1344092

Wagner, T. (2014). *The global achievement gap*. (*Updated edition* ed.). Perseus Books Group Press.