UTKUN AYDIN and BEHIYE UBUZ

THE THINKING-ABOUT-DERIVATIVE TEST FOR UNDERGRADUATE STUDENTS: DEVELOPMENT AND VALIDATION

Received: 26 September 2012; Accepted: 6 April 2014

ABSTRACT. Two studies were conducted for the development and validation of a multidimensional test to assess undergraduate students' mathematical thinking about derivative. The first study involved two phases: question generation and refinement of the Thinking-about-Derivative Test (TDT). The second study included four phases as follows: test administration, generalizability analysis, confirmatory factor analysis, and subgroup validity analysis. Findings suggested that the 30-item multiple-choice TDT, which comprises 6 mathematical thinking aspects, enactive, iconic, algorithmic, algebraic, formal, and axiomatic thinking, demonstrates acceptable levels of reliability and validity. Followed by additional cross-validation studies, the TDT may be a useful tool for mathematics education researchers and mathematicians. Directions for future research and implications for educational practice are discussed.

KEY WORDS: derivative, mathematical thinking, multiple-choice test, reliability, validity

INTRODUCTION

Although derivative is one of the most important concepts that reside at the core of calculus, many students, even high achievers, encounter difficulties in solving derivative tasks that stimulate different thinking aspects (Artigue, 1991). Tasks used in previous studies on derivative such as defining the derivative (Ubuz, 1996; Zandieh, 1997) or defining prerequisite concepts (i.e. slope) for derivative (Nagle, Moore-Russo, Viglietti & Martin, 2013), proving differentiation theorems (Tsamir & Ovodenko, 2013), solving routine differentiation problems (Martin, 2000; Ubuz, 2001), modeling real-life applications of derivative (Villegas, Castro & Gutiérrez, 2009), interpreting or sketching derivative graphs (Ubuz, 2007), and verifying whether a given function satisfies the hypotheses of a differentiation theorem (Raman, 2002) were set up to encourage a different set of opportunities for students' thinking. The task that requires defining the derivative, for example, focuses on property-based conceptions or thinking. Although individuals may approach mathematical tasks in different ways, with results depending on their thinking styles (Moutsios-Rentzos & Simpson, 2011) or their preferred

International Journal of Science and Mathematics Education (2015) 13: 1279–1303

ways of using the abilities they have (Sternberg, 1999), the nature of tasks can potentially influence and structure the way students think (Henningsen & Stein, 1997). A task that energizes a particular mathematical thinking aspect also puts together a range of different aspects dependent on the context in which the concept occurs. In proof tasks that require making logical deductions to prove theorems, for example, property-based thinking also comes into play in terms of making deductions.

The commonly pursued line of research on the assessment of students' mathematical thinking about derivative addressed how secondary (e.g. Pichat & Ricco, 2001) and/or undergraduate (e.g. García, Llinares & Sánchez-Matamoros, 2011) students think about the differentiation situation by documenting and/or representing their mathematical ideas, particularly using open-ended tasks in the instruments. While producing rich data, the aim of these studies was not to examine or interpret the psychometric properties (i.e. reliability and validity) of the mathematical thinking scores on these instruments. This approach presents limitations for examining the range of mathematical thinking aspects in which a *specific* derivative task activates a *particular* aspect. More adaptive and content-specific instruments are needed for this purpose. Consequently, we developed Thinking-about-Derivative Test to assess various aspects relevant to mathematical thinking. Specifically, the purpose of the present study was to (a) develop a multiple-choice test on mathematical thinking about the derivative, (b) test the validity and reliability of the test, and (c) provide further validity evidence.

MATHEMATICAL THINKING

Thinking is typically defined as the means used by individuals to improve their understanding of, and exert some control over, their environment (Burton, 1984, p. 36). To do this, mathematical thinking lies on particular means such as different registers or representations that can be recognized as arising from or pertaining to the study of mathematics. To address the different aspects of mathematical thinking, we exploited three corresponding theories in which different registers or representations can be interpreted as (a) modes of representation (Bruner, 1966), (b) modes of operation (Hughes-Hallett, 1991), and (c) worlds of mathematics (Tall, 2004). These three theories are not meant to be exhaustive but represent a useful set of core theories that cut across all mathematical domains. Here, it is necessary to mention that mathematical thinking aspects become more sophisticated when the individual becomes more experienced (Tall, 2003).

Bruner's (1966) framework provides an overview of human development in general and mathematical thinking in particular. It presents a long-term development from physical perception and action through the development of symbolism and on to thinking. Bruner classified three modes of human representation in which thinking is formulated in three ways as follows: enactive, iconic, and symbolic. *Enactive* representation involves perceptions of and reflections on real-world objects through which individuals translate daily life situations into a mathematical context (enactive thinking). *Iconic* representation includes mental images by which individuals alter objects and properties of these objects for making mathematical visualizations (iconic thinking). *Symbolic* representation covers symbol systems (e.g. numbers, algebraic expressions, logic, and language) that allow individuals to define mathematical concepts (formal thinking), implement procedural techniques (algorithmic thinking), and make generalizations (algebraic thinking). Clearly, the single category of symbolic mode including a variety of symbol systems needs subdivision.

Bruner's modes of representation can be used to specify operations that can be performed as a sequence of steps because, for instance, throughout the development of symbolism, individuals learn to carry out an operation, to practice it, and then use it as a tool for thinking. In this accordance, Hughes-Hallett (1991) categorized the modes of representation into four ways of operation as follows: numeric, analytic, verbal, and graphic. She suggested that translations among operations, as well as transformations within each, are important processes that lead students to develop robust mathematical thinking. In this regard, effective mathematical thinking is defined as students' being able to work within and among (a) numbers and mathematical notations (numeric) to apply procedures (algorithmic thinking), (b) symbols and algebraic expressions (analytic) to display relationships and generalizations (algebraic thinking), (c) definitions and principles (verbal) to elucidate static and factual information (formal thinking), and (d) graphs, diagrams, and tables (graphic) to make visualizations (iconic thinking). The important contribution here is that the symbolic mode of representation forwarded by Bruner (1966) is distinguished into numeric, analytic, and verbal modes of operation that prompt algorithmic, algebraic, and formal thinking, respectively. Interestingly, however, this categorization lacks in two points. First, the enactive mode of operation that fosters enactive thinking is completely omitted. Second, a theoretical mode of operation that provokes axiomatic thinking is absent. Enactive mode helps to give fundamental mathematical meaning by using embodied ideas, whereas the theoretical mode helps to formulate fundamental mathematical properties as axioms.

Following Bruner (1966) and Hughes-Hallett (1991) theoretical frameworks that build from perception and action and develop through reflection, Tall (2004) built the three-world framework on the tripartite structure of perception, operation, and reason. All three of these aspects arise throughout conceptual-embodied, proceptual-symbolic, and formal-axiomatic world of mathematics. He formulates the transition in thinking about perceptions, operations, and the methods of reasoning within a more global framework that takes into account students' (a) object-based conceptions (conceptual-embodied) to synthesize the properties of the physical environment (enactive thinking) or visualize mental imagery concepts (iconic thinking), (b) action-based conceptions (proceptual-symbolic) to make calculations (algorithmic thinking) and encode correlational information (algebraic thinking), and (c) property-based conceptions (formal-axiomatic) to recall the definitions of concepts (formal thinking) and use these definitions as axioms with which to make logical deductions to prove theorems (axiomatic thinking).

All these theoretical perspectives commonly point that mathematical thinking evolves through six aspects as follows: enactive, iconic, algorithmic, algebraic, formal, and axiomatic thinking, each with its own way of describing different processes. The term thinking in these theories were mainly used to describe development. This focusing on thinking can result in neglecting the types of mathematical tasks and ways of participating in these tasks (Rasmussen, Zandieh, King & Teppo, 2005). Although students' way of participating in, for example defining tasks, may be different based on their thinking style, they are required to express their formal thinking. We therefore deal with the particular aspect of mathematical thinking that a task mandates by its very nature. A detailed explanation for each of these mathematical thinking aspects along with an example of derivative tasks (see Table 1) that incorporates the activation of a particular mathematical thinking aspect are provided below.

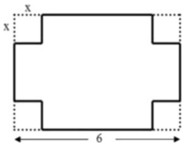
Enactive, Iconic, Algorithmic, Algebraic, Formal, and Axiomatic Thinking

Enactive thinking (*ENACTHK*) can be viewed as roughly equivalent to the process of model-building that students develop and use during their efforts to solve a real-world problem (Lesh & Doerr, 2003). Broadly

TABLE 1

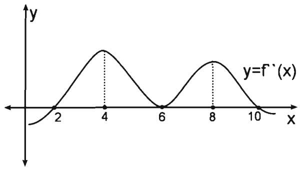
Specimen items of the TDT

Enactive thinking
Item 30:



An open-top rectangular prism box is to be made by cutting congruent squares of side length x from the corners of a 6 cm square plate as illustrated in the figure above. What is the largest volume of the box?
A) 8 B) 12 C) 16 D) 18 E) 20

Iconic thinking
Item 17:



The figure shown gives the graph of the second derivative of a function f. Find the inflection points of f?
A) 2, 6, and 10 B) 2, 6, and 8
C) 2, 4, and 6 D) 4 and 6 E) 4 and 8

Algorithmic thinking
Item 25:
Let $f(x) = 2x^3 + ax^2 + (b + 1)x - 3$ be a function that has a local extremum at $x = -1$ and an inflection point at $x = \frac{-1}{12}$.
Find a. b?
A) −3 B) −2 C) 4 D) 6 E) 12

Algebraic thinking
Item 12:
Let f: [6, 15] $\rightarrow R$ be a function with $f(6) = -2$ and f′. If f satisfies the hypotheses of the mean value theorem on [6, 15], then which one of the following is true?
I. The maximum value of $f(15)$ is 88.
II. The average value of the function f on [6, 15] is 45.
III. The function f takes on the value of 45 at least once on [6, 15].
A) I only B) II only C) I and III
D) II and III E) I, II, and III

Formal thinking
Item 2:
Which one of the statements is true with respect to the definition of the *inflection point*?
I. It is the point where the first derivative of function $f\colon[a,b]\rightarrow R$ equals to zero.
II. It is the point where the second derivative of function $f\colon[a,b]\rightarrow R$ equals to zero.
III. It is the point where function $f\colon[a,b]\rightarrow R$ changes from increasing to decreasing or from decreasing to increasing.
IV. It is the point where function $f\colon[a,b]\rightarrow R$ changes from convex to concave or from concave to convex.
A) I and III B) II and IV
C) III and IV D) II only E) IV only

Axiomatic thinking
Item 7:
"Rolle's theorem: Let f: [a, b]$\rightarrow R$ be a continuous function on the closed interval [a, b] and a differentiable function on the open interval (a, b). If $f(a) = f(b)$, then there is a number c in (a, b) such that $f'(c) = 0$."
Which inference is true with respect to the theorem given above?
A) The function f has more than one root on (a, b).
B) The function f has at least one critical point on (a, b).
C) The tangent drawn to the function f at the point (c, $f(c)$) is perpendicular to the x-axis.
D) The first derivative of the function f is always positive or always negative on (a, b).
E) The slopes of the secant line and the tangent line drawn to the function f on (a, b) are different from each other.

speaking, it includes movement among the real-world situations and the mathematical solutions to make sense of the ways to manipulate the physical environment (Zbiek & Conner, 2006). ENACTHK, in this sense of the term, is activated when (a) examining various attributes of a particular mathematical, physical, or social context (i.e. identify the real-world phenomenon); (b) embodying key aspects of these attributes into a mathematical model (i.e. simplify the phenomenon); (c) relating a subset of those key aspects through operations, equations, or functions (i.e. express the simplified phenomenon mathematically); and (d) using resulting internal and external representations to solve problems (i.e. verify, interpret, and solve the model) (Gainsburg, 2006). An example of ENACTHK about derivative is constructing and interpreting a real model (e.g. diagram and graph) to understand the given conditions about the optimization situation, structuring a mathematical model (e.g. equation and function) to explore the quantities that are to be maximized or minimized, and investigating various models to find the maximum/minimum value of a function (see item 30 in Table 1).

Iconic thinking (*ICONTHK*) refers to the act of visualizing through which individuals reflect and interpret upon images, diagrams, or graphs with the purpose of depicting and communicating information (Arcavi, 2003). It covers four main processes as follows: graph reading, graph interpretation, graph construction, and graph evaluation. More specifically, graph reading involves extracting data from a graphical display and generating information by operating on data shown in the graphical display (Meletiou-Mavrotheris & Lee, 2010). Graph interpretation involves gaining meaning and/or making inferences from a graph (Sharma, 2006). Graph construction includes displaying and organizing data sets using graphs or understanding graph conventions (Friel, Curcio & Bright, 2001). Graph evaluation entails analyzing whether a graph is correctly constructed or whether a graph is effective to represent the interrelationships between the given context and data (Curcio, 1987). An example of ICONTHK about derivative is reading and interpreting the graph of a second derivative function to determine the inflection points of the original function (see item 17 in Table 1).

Algorithmic thinking (*ALGOTHK*) is procedural in nature—it is characterized by the automatized processes such as computation, calculation, and execution (Martin, 2000). It centers on selecting and applying the appropriate procedures to solve a problem (Martin, 2000). Furthermore, ALGOTHK involves (a) verifying or justifying the correctness of procedures and (b) explaining the successive steps involved in various standard operations (Fischbein, 1983). An example of

ALGOTHK about derivative is selecting and applying the appropriate inflection point and local extrema algorithms to find the product of unknown variables in a given function (see item 25 in Table 1).

Algebraic thinking (*ALGETHK*) can be defined as the use of various representations that handle quantitative situations in a relational way (Kieran, 1996, pp. 4 – 5). It is considered to be important in both *seeing a generality through the particular* and *seeing the particular in the general* (Mason, 1996, p. 65). Three central themes are at the core of ALGETHK as follows: (a) viewing variables, symbols, expressions, and equations as structures of general representation (Stacey & MacGregor, 2000); (b) understanding of how and when algebraic expressions should be used to display relationships and generalizations (Arcavi, 1994); and (c) applying a given argument to a broader context by making empirical and/or theoretical generalizations (Harel & Tall, 1991). An example of ALGETHK about derivative is using a conditional argument given for a function at a certain interval to display the relationships among the hypotheses of the mean value theorem (see item 12 in Table 1).

Formal thinking (*FORMTHK*) focuses on the factual information that undergirds basic mathematical terminology (Fischbein, 1983). Accordingly, it is represented by the construction of meaning from definitions, principles, facts, and symbols (Tall, 2004). In a more figurative sense, it involves (a) using, connecting, and interpreting various conceptual representations; and (b) recalling, distinguishing, and integrating definitions, principles, facts, and symbols in a mathematical setting (Martin, 2000). An example of FORMTHK about derivative is integrating the definitions of various differentiation concepts (i.e. first derivative, second derivative, increasing/decreasing functions, and convex/concave) to recall the definition of inflection point (see item 2 in Table 1).

Axiomatic thinking (*AXIOTHK*) closely aligns with the notion of proof and the process of proving, and thus follows a path between attempts to generate valid arguments and criticisms of these attempts (Stylianides, 2007). In this context, it is crucial for (a) making a connected sequence of assertions for or against a mathematical claim (Stylianides, 2007); (b) understanding how and why a statement works (Tall, 2004); and (c) demonstrating whether and why propositions are true or false (Ko & Knuth, 2009). The act of AXIOTHK that culminates in proving involves exploring mathematical relationships to identify significant facts and arrange specified assumptions into meaningful patterns, using the patterns to formulate conjectures, and testing and revising these conjectures (Stylianides, 2009). An example of AXIOTHK about derivative is identifying the logic behind the hypotheses that are used either explicitly or implicitly in the statement of Rolle's theorem, analyzing the way in which these different hypotheses are connected in the proof of Rolle's theorem, and formulating cases where Rolle's theorem is plausible and subsequently chaining all these conditions as a whole to deduce the truth of statements (see item 7 in Table 1).

Viewed together, the six mathematical thinking aspects are distinct but interrelated. Ideally, they should cooperate in any mathematical task and their partial vitality depends upon students' building appropriate links among them. To this end, mathematical thinking is not characterized by the replacement of one aspect of thinking by another that supposedly is "higher" or "more abstract"; rather it is characterized by the development and interlinking of different aspects of thinking that can develop alongside and in combination with one another. Empirical findings also provided a sizable body of evidence that there are unique and joint effects of mathematical thinking aspects on one another (Hähkiöniemi, 2006; Zandieh, 1997). Collectively, research results indicated that (a) FORMTHK provides a ground on which ENACTHK, ICONTHK, ALGOTHK, ALGETHK, and AXIOTHK depend; (b) AXIOTHK and ALGETHK coevolve in a dialectic process of hypothesizing, verifying, and generalizing; and (c) ICONTHK, by virtue of its concreteness, can accompany ENACTHK in the very act of building, comparing, and investigating models.

METHOD

Along with the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999), we used a multistep process to provide evidence that scores on the Thinking-about-Derivative Test (TDT) are reliable and valid. Two studies were conducted for the construction, refinement, and validation of the TDT. The whole test is available from the authors upon request.

Study 1: Question Generation and the Refinement of Thinking-About-Derivative Test

Phase 1: The Generation of Multiple-Choice Test Items. The purpose of this phase was to construct a multiple-choice item pool on different aspects of mathematical thinking related to derivative and its applications. In assessing students' mathematical thinking, researchers generally had a tendency to use open-ended item format (e.g. Zandieh, 1997). Although such open-ended items may convey useful information for researchers to identify the aspects of mathematical thinking that students activate within a specific content (Moreno & Mayer, 1999), they lack the potential for extensive domain coverage, thus yield lower reliability and ambiguous individual-level scores (Parke, Lane & Stone, 2006). Multiple-choice items, on the other hand, enable researchers to ask a large number of questions on a wider range of topics and use plausible distracters that reflect typical student errors (Bridgeman & Rock, 1993). Although using multiple-choice items presents a risk of guessing, in view of providing students enough test items and including sufficient number of distracters for these items, we do, however, consider the chance of a student guessing the right answer is decreased (Haladyna, 2004). Besides, depending on the particular aspect of mathematical thinking that a task mandates by its very nature, multiple-choice items can elicit complex cognitions (Martinez, 1999), including making comparisons/contrasts, building cause/effect relationships, or drawing generalizations (Haladyna, 2004). That is, students capitalize on information embedded in the response options of a multiple-choice item that mandates a particular aspect of mathematical thinking. They further use more complex strategies such as response elimination (Martinez, 1999) to narrow down the presented response options. While eliminating options that are implausible, students would be likely to activate different mathematical thinking aspects. They may further integrate their thinking more carefully into the evaluation of the remaining set of responses to yield the best selection given the effectiveness of a particular mathematical thinking aspect they activate. It is appropriate to conclude that a set of well-constructed multiple-choice items, in which the goal of testing for a particular mathematical thinking aspect is made explicit, can provide very reliable assessment of *what* processes students demonstrate rather than *how* students make progress in activating these processes (Osterlind, 1998, p. 164).

From the investigation of calculus textbooks, course materials, journal articles, dissertations, and matriculation exam questions, an initial item pool containing 100 open-ended and 83 multiple-choice items based on the definition of each mathematical thinking aspect was constructed. Forty-nine of these 100 open-ended items had been used effectively in previous studies (e.g. Aspinwall et al., 1997) to reveal unique insights about the abilities and understandings of students, misconceptions they held, or mathematical thinking they activated when learning or using the derivative concept to solve problems. The remaining 51 open-ended items were from calculus textbooks and course materials. On the other hand, 83 multiple-choice questions were from calculus textbooks (e.g. Stewart, 2003), course materials (i.e. Calculus, General Mathematics), and national matriculation exam questions. With regard to the derivative content covered in undergraduate courses, we decided to delimit the content of (a) ENACTHK items on real-life applications of derivative; (b) ICONTHK items on graph interpretations/constructions of the derivative; (c) ALGOTHK items on routine differentiation problems; (d) ALGETHK items on the hypotheses and/or assumptions of differentiation theorems; (e) FORMTHK items on the definitions, principles, facts, and symbols relevant to the derivative; and (f) AXIOTHK items on differentiation theorems and their proofs. At the end of this delimiting process, 60 open-ended items measuring all 6 mathematical thinking aspects and 43 multiple-choice items measuring ENACTHK, ICONTHK, and ALGOTHK were retained. There were no multiple-choice items in the literature that assess FORMTHK, ALGETHK, and AXIOTHK.

The further investigation of the selected 60 open-ended items revealed that 20 of them require self-explanatory responses (e.g. What does the derivative mean in practice? For what the derivative can be used? How can the derivative be determined?). So, they were eliminated. This process led us to have 40 open-ended items to be adapted to researcher-developed multiple-choice items. Converting an open-ended item to a multiple-choice item involved (a) using the overall mathematical and methodological essence provided in the open-ended items (i.e. "In your own words, define the derivative of a function."; "Prove that differentiability implies continuity.") for constructing the stems of multiple-choice items; (b) analyzing possible responses given to these open-ended items in order to phrase the distracters of multiple-choice items. As one example, item 2 (see Table 1) was developed by the authors to assess students' FORMTHK about inflection point. The work on that item involved (a) identifying the most common student responses, and (b) developing statements for each response. In defining the inflection point, the most common responses provided by students were "second derivative positive, concave up" and "second derivative equal to zero, inflection point" (Carlson, 1998), "change in concavity of a graph" (Berry & Nyman, 2003), or "the rate changing from increasing to decreasing" (Carlson et al., 2002). Following that, the explanations involved in these responses were used to form statements for answer choices (see Table 1).

Phase 2: The Refinement of Multiple-Choice Test Items. The purpose of this phase was twofold: to provide (a) evidence based on test content via expert evaluations, and (b) evidence based on response processes via student evaluations. The TDT including 83 multiple-choice items

(40 researcher-developed and 43 from the resources) was submitted to the experts with a table presenting the number, objective, and thinking aspect that corresponds to the context of the item along with the solution and answer keys. To rank how well the item context fit with the characteristics of relevant mathematical thinking aspect, the definitions of each mathematical thinking aspect were also submitted. The experts were requested to analyze the items in terms of their (a) contribution to the relevant mathematical thinking aspect, and (b) clarity, comprehensiveness, and accuracy. A mathematics educator, who is Behiye Ubuz, and a mathematician were the experts. The main research fields of Behiye Ubuz are the development of mathematical thinking and the teaching and learning of calculus, whereas the mathematician is specialized in real analysis and calculus. The nature of the study and difficulty in finding an expert in mathematics education to evaluate the test items allowed us to have Behiye Ubuz as an expert, since mathematics educator as an expert needed to have knowledge on derivative, mathematical thinking, measurement and evaluation, and so forth. Being unfamiliar with a specific area (i.e. mathematical thinking about derivative) can limit him/her to use the necessary knowledge to make accurate, reliable, and unbiased judgements (Bolger & Wright, 1992).

Based on Behiye Ubuz suggestions, (1) the content of ENACTHK items which initially covered rate of change, population growth, and optimization was restricted to optimization; (2) the content of ALGETHK items involving elementary applications of differentiation theorems was revised to reveal the more complicated linkages between the theorem statement and its hypotheses by focusing on the algebraic manipulations as well as mathematical inferences that may derive from these linkages; (3) the structure, syntax, and distracters of FORMTHK items were revised by taking into consideration the static and factual information inherent in the basic differentiation terminology and common student misconceptions as well as errors; (4) the hints that has been included in the distracters of the AXIOTHK items were removed; and (5) the familiar function graphs in the distracters of ICONTHK items were replaced with more unfamiliar ones, and the number of items on graph construction and graph interpretation were made equal. Based on the suggestions of the mathematician, (1) the statements of differentiation theorems were integrated into the stem of AXIOTHK items; (2) some of the ALGETHK items including functions on which the hypotheses of differentiation theorems would be applied were revised; and (3) the overlaps in some of the ENACTHK, ALGOTHK, and FORMTHK items were reconsidered, and items that covered the same content and/or context were eliminated. In all, 53 items of these 83 items were deleted from the TDT.

Taken together the results of the adaptation and elimination processes, the TDT involved 30 multiple-choice items, 17 of which were researcher-developed (items 1 – 15, 18, 19) and 13 of which were taken from calculus textbooks, course materials, or national matriculation exams (items 16, 17, 20 – 30). More specifically, 30 TDT multiple-choice items were grouped in 6 FORMTHK, 5 AXIOTHK, 4 ALGETHK, 5 ICONTHK, 5 ALGOTHK, and 5 ENACTHK items.

For providing evidence based on response processes, three first year students majoring in Elementary Mathematics Education, two PhD students majoring in Secondary Science and Mathematics Education, and two second year students majoring in Mathematics were asked to evaluate the remaining 30 items in terms of their clarity and intelligibility. All the participants were enrolled in a calculus course in their first or second years at the university. Six of the seven students were requested to complete the TDT in a classroom environment within one block-class period (90 min) in order to determine the duration, while one student was requested to solve each item using a think-aloud procedure in the presence of the first researcher. Based on students' suggestions, minor revisions were made as follows: (1) ordering of ENACTHK and ICONTHK items were changed, and (2) spelling/editing errors in FORMTHK and AXIOTHK items were corrected. The revised final version of the TDT containing 30 items was reviewed in its entirety, and no further changes were made. Each item was scored either 0 (incorrect) or 1 (correct). The total testing time was decided to take one block-class period long (90 min).

Study 2: Instrument Testing

Phase 1: Test Administration. During 2010/2011 academic year, the TDT including 30 items was administered to 766 undergraduates from 9 public universities in Turkey. Cross-sectional data were collected from year 1 (N = 293; freshmen), year 2 (N = 131; sophomores), year 3 (N = 254; juniors), and year 4 (N = 88; seniors) students who were attending to the Faculty/School of Education (N = 253), Faculty/School of Arts and Sciences (N = 284), and Faculty/School of Engineering (N = 229).

In many countries such as Turkey, the derivative concept is introduced to the students at the 12th grade in high school. The emphasis is placed on the essential components of the derivative and its applications such as defining derivative verbally, graphically, symbolically, and physically (FORMTHK); explaining the relationship between continuity and differentiability (ALGETHK); solving derivative problems by using differentiation rules (ALGOTHK); sketching and interpreting derivative graphs (ICONTHK); building models to find the maximum/minimum of a given function (ENACTHK); and proving differentiation theorems (AXIOTHK). When entering university, students are expected to understand the basics of derivative and to be able to employ different aspects of mathematical thinking about it. All science, education, and engineering students take introductory mathematics courses (i.e. Calculus I & II, Fundamentals of Mathematics) as a must course in the first year of the university. The derivative content taught in the introductory mathematics courses at the university is similar to the content taught at high school. Drawing on a broad base of fundamentals built in these courses, students activate and extend different aspects of mathematical thinking about derivative in subsequent mathematics courses as well as other mathematics-related departmental courses during the 4 years of the university studies. In common, all science, engineering, and education students take Introduction to Differential Equations and Basic Linear Algebra in year 2 and Differential Equations and Linear Algebra in year 3. The content of these courses covers basics of differentiation rules and applications necessary for introducing further concepts of mathematics (e.g. first-order differential equations). Along with the introductory mathematics courses in year 1, the courses in years 2 and 3 serve as a prerequisite or a corequisite for the mathematics-related scientific courses that at least to some extent depending on the conceptualization and the application of derivative. In the Faculty/School of Arts and Sciences, some mathematics-related departmental courses are Mathematics for Chemists, Mathematical Methods in Physics, whereas laboratory courses are Experimental Physics and Analytical Chemistry Laboratory, and so forth. Similarly, in the Faculty/School of Engineering, students continue to build on their earlier experiences with the derivative concept and use that understanding to explore new situations or systems in both theoretical (i.e. Numerical Methods for Engineers) and technical (i.e. Mathematical Modeling and Applications) courses as they proceed through the years. In the Faculty/School of Education, students who have acquired a firm grounding in the theories and applications of the derivative, further strengthen their mathematical thinking base in various must courses (i.e. Special Teaching Methods) as well as elective courses (i.e. Problem Solving in Mathematics) as they move through the years. Viewed together, all the participants were at least to some degree familiar with the fundamentals and the applications of the derivative concept.

The 90-min TDT was administered to classes during regular course sessions by Utkun Aydın.

Phase 2: Generalizability Analysis. Reliability of test scores was examined within the context of generalizability theory (Brennan, 2001). This theory was used to characterize specific sources of error variance (i.e. items and students) that contaminate the measurement of mathematical thinking such that future measurements can be made more accurately. Thus, we initially carried out a generalizability study (G-study) to investigate the primary sources of variation in the measurement of mathematical thinking by means of the TDT. By further identifying the simultaneous influence of multiple sources of measurement error variance, the G-study allowed us to estimate the accuracy of generalizing from a student's observed score on the TDT to the average score that student would have received under the possible conditions such as all possible items or all possible mathematical thinking aspects. Estimates of variance attributable to each source in the G-study were then used to conduct a decision study (D-study) to determine the effects of the number of items. More precisely, we used a mixed two-facet partially nested *s* x (*i*:*a*), where *s*, *i*, and *a* represent students, items, and aspects (subscales), respectively, while ":" denotes "nested within." Since the six aspects contain different number of items, the design is described as unbalanced with five sources of variance: ${\sigma_s}^2$, ${\sigma_a}^2$, ${\sigma_{i:a}}^2$, ${\sigma_{sa}}^2$, and ${\sigma_{si:a,e}}^2$. Students were considered as the object of measurement, whereas items were treated as a random facet, and aspects were treated as a fixed facet. The full data tables for the generalizability analysis are available upon request from the authors.

Averaging Over Aspects. We used a three-step procedure for averaging over the aspects by using analysis of variance (ANOVA) procedures to partition variance attributable to each source. All the main calculations were done by using PASW Statistics 18 (SPSS Inc, 2010). In the first step of the G-study, we treated all sources of variance as random (${\sigma_s}^2$, ${\sigma_a}^2$, ${\sigma_{i:a}}^2$, ${\sigma_{sa}}^2$, and ${\sigma_{si:a,e}}^2$). For the fully random analysis with all six aspects, we selected four items at random from each aspect to create a balanced design (Brennan, 2001). When randomly deleting data, however, we compared the results of several random deletions to make sure that any particular selection of items is not unusual. For the TDT data set, we estimated variance components for several randomly selected sets of four items per aspect. We observed that the estimated variance components were very similar for all sets of randomly selected items. Results revealed that the main effect of aspects is nonnegligible (${\sigma_a}^2 = 0.28$). This shows that some aspects yielded somewhat higher average mathematical thinking scores than others. The more substantial effect for the interaction between students and aspects ($\sigma_{sa}{}^2$ = 0.73) shows that the relative standing of students differed from one aspect of mathematical thinking to another. In the second step, we identified the random portion of the mixed design (students crossed with items, *s* x *i*) and the associated variance components to be calculated ($\sigma^2_{s^*}$, $\sigma^2_{i^*}$, and σ^2_{si,e^*}). In the third step, we estimated the variance components for the random portion of the mixed design. Findings demonstrated that the differences between items were substantial ($\sigma^2_{i^*}$ = 0.16). This indicates that some items were more difficult than others, averaging over all the students. Similarly, the substantial variance component for the interaction between students and items (σ^2_{si,e^*} = 0.10) indicates that the relative standing of students differed somewhat across items, unmeasured variation, or both. Indeed, this suggests that there are important sources of variance not accounted for by differences between students, differences in item difficulty, or both.

Analyzing Each Aspect Separately. We carried out a separate generalizability analysis with *s* x *i* design for each aspect. The six separate analyses allowed us to use all of the items in the original aspects (6, 5, 4, 5, 5, 5), not the four items randomly selected to create a balanced *s* x (*i*:*a*) design. For comparison, however, the analyses using the four randomly selected items were also carried out and similar results were obtained. Although relative decisions (i.e. norm-referenced) will ordinarily be more important with the use of the TDT, absolute decisions (i.e. criterion-referenced) may sometimes be of interest in educational settings. Hence, generalizability coefficients (i.e. ρ^2, reliability estimates for relative decisions) as well as indexes of dependability (i.e. $\varnothing$, reliability estimates for absolute decisions) were calculated. The separate analyses of the six mathematical thinking aspects yielded similar results relevant to the estimated variance components. Findings showed that there were substantial differences among students in their scores on each of the six mathematical thinking aspects. More specifically, the estimated variance component for students accounted for most of the variation in each scale. Results revealed that differences in the relative standing of students are the largest for AXIOTHK (1.28) and ICONTHK (1.28), whereas they are smaller for ALGETHK (0.91) and ALGOTHK (0.99). This suggests that averaging over all the aspects, students differed more in activating theoretical or visual processes, and to a lesser extent, in mobilizing hypothetical or technical processes. The variance components for items accounted for a smaller variation. The ALGOTHK aspect showed the greatest variation (0.47), while the FORMTHK aspect showed the least variation (0.15) across items. Taken together, this implies that averaging over all the students, items including the implementation of procedures and techniques were more difficult than the items involving the recognition of definitions, symbols, and facts. The residual component also accounted for substantial variation in each aspect. For instance, the residual component of ICONTHK is 0.27, indicating that there are important sources of variance not accounted for the relative standing of students in solving visual/spatial items, differences in these items (i.e. graph interpretation/construction), or both.

Using these estimated variance components from the G-study, we compute the variance components for alternative D-studies to estimate how the consistency of students' mathematical thinking scores would change if different conditions such as using different number of items per aspect were taken into consideration. This allowed us to examine how the generalizability coefficients changed under different circumstances, and consequently determine the ideal conditions under which our measurement of mathematical thinking would be the most reliable. If the users of the TDT intend to use an average score across aspects, using 24 items (4 per aspect) yields acceptable levels of generalizability. If they intend to use separate scores for each aspect, using 6, 5, 4, 5, 5, and 5 items for the FORMTHK, AXIOTHK, ALGETHK, ICONTHK, ALGOTHK, and ENACTHK, respectively, would yield coefficients for absolute decisions of at least 0.86. In addition, generalizability values increase as the number of items within the fixed 6 aspects increases to 12, 10, 8, 10, 10, and 10 for FORMTHK, AXIOTHK, ALGETHK, ICONTHK, ALGOTHK, and ENACTHK, respectively. The generalizability coefficient values range from 0.95 to 0.98, whereas the indexes of dependability range from 0.91 to 0.97, indicating sufficiently high reliability for all six mathematical thinking aspects. Collectively, the relationship between the number of items and the generalizability value has a similar pattern to that of the dependability value. In sum, the coefficients show the diminishing returns for increasing the number of items per aspect. Furthermore, we, as researchers, must balance cost considerations and test length in choosing the optimal number of items. Henceforth, we retained 6, 5, 4, 5, 5, and 5 items for the FORMTHK, AXIOTHK, ALGETHK, ICONTHK, ALGOTHK, and ENACTHK, respectively.

Phase 3: Confirmatory Factor Analysis. A *confirmatory factor analysis* (CFA) was conducted to provide supportive evidence for the six-factor structure of the TDT. The analyses employed the LISREL 8.7 (Jöreskog & Sörbom, 1993) in calculating weighted least squares (WLS) estimates. The results from the theory-driven CFA are shown in Table 2. The

TABLE 2

Standardized estimates and reliability coefficients of the items in the TDT

Items	*Enactive thinking*	*Iconic thinking*	*Algorithmic thinking*	*Algebraic thinking*	*Formal thinking*	*Axiomatic thinking*	R^2
24	1.00						0.70
29	0.69						0.83
28	0.76						0.79
30	0.60						0.89
27	0.65						0.85
20		1.00					0.74
19		0.87					0.87
16		0.85					0.74
18		0.94					0.81
17		0.67					0.77
23			1.00				0.77
25			0.92				0.78
22			0.79				0.70
21			0.81				0.73
26			0.65				0.83
15				1.00			0.90
14				0.88			0.88
12				0.77			0.80
13				0.87			0.86
4					1.00		0.86
5					0.85		0.84
3					0.73		0.84
2					0.74		0.83
6					0.93		0.88
1					0.90		0.88
9						1.00	0.80
10						0.77	0.87
8						0.67	0.77
11						0.90	0.84
7						0.89	0.80

obtained fit indexes for the six-factor model were χ^2 (789.77, n = 766) = 2.12, RMSEA = 0.03, GFI = 0.96, AGFI = 0.95, and CFI = 0.98. These indices suggest that the model fits the data according to the multiple criteria including a chi-square ratio of three or less, a GFI above 0.90, an AGFI above 0.90, a RMSEA from 0.06 to 0.08, and a CFI above 0.95 (Kline, 2005). Taken together, the indices confirm that the underlying structure of the TDT is formed by six factors that measure FORMTHK (items 1, 2, 3, 4, 5, and 6), AXIOTHK (items 7, 8, 9, 10, and

11), ALGETHK (items 12, 13, 14, and 15), ICONTHK (items 16, 17, 18, 19, and 20), ALGOTHK (items 21, 22, 23, 25, and 26), and ENACTHK (items 24, 27, 28, 29, and 30).

Standardized estimations were higher than 0.45 and appeared between 0.60 and 0.93, showing that all items are relevant in defining the corresponding mathematical thinking aspect (Jöreskog & Sörbom, 1993). Squared multiple correlation (R^2) of individual items were higher than 0.50 ranging from 0.70 to 0.90. This indicates that the reliability of the items was substantial in size, and that these items can be explained by the corresponding mathematical thinking aspects (Tabachnick & Fidell, 2007).

In an attempt to validate the theoretical model, we tested three alternative models as follows: a common factor model, a three-factor model, and a null model (see Table 3). The common factor model was specified such that all items loaded on a general single factor as mathematical thinking. This model proposed that ENACTHK, ICONTHK, ALGOTHK, ALGETHK, FORMTHK, and AXIOTHK are not conceptually or statistically distinct. The three-factor model was specified such that items loaded on three factors. Based on Tall's (2004) theory, this model proposed that mathematical thinking can be distinguished in conceptual-embodied, proceptual-symbolic, and formal-axiomatic thinking. It is impossible to untie the merging of ENACTHK, ICONTHK, ALGOTHK, ALGETHK, FORMTHK, and AXIOTHK in that every mathematical task calls out some paths of embodiment, symbolism, and formalism. Drawing on this perspective, in the three-factor model, ENACTHK and ICONTHK items were combined to constitute the conceptual-embodied thinking dimension, ALGOTHK and ALGETHK items were gathered under proceptual-symbolic thinking dimension, and FORMTHK and AXIOTHK items were put together to compose the formal-axiomatic thinking dimension. The null model implied that all the items are uncorrelated, and that each item of the TDT constitutes a single factor.

TABLE 3

Comparison of models: goodness-of-fit indices

Model	χ^2	*df*	χ^2/df	*RMSEA*	*GFI*	*AGFI*	*CFI*	$\Delta\chi^2$	Δdf	ΔCFI
Target	789.77	372	2.1	0.03	0.98	0.95	0.98			
Common factor	1284.83	350	3.6	0.05	0.92	0.91	0.95	495.06	1.5	0.03
Three-factor	989.79	341	2.8	0.05	0.94	0.93	0.96	199.02	0.7	0.02
Null	1,580.14	390	4.5	0.06	0.91	0.89	0.95	790.37	2.4	0.03

In comparison across goodness of fit indices and the chi-square difference test, the model fit was poorer for the common factor model ($\Delta\chi^2 = 495.06$, $\Delta df = -22$, $p < 0.001$); the three-factor model ($\Delta\chi^2 = 199.02$, $\Delta df = -31$, $p < 0.001$), and the null model ($\Delta\chi^2 = 790.37$, $\Delta df = 18$, $p < 0.001$). In sum, CFAs corroborated the proposed six-factor structure of the TDT and indicated that undergraduate students distinguish among ENACTHK, ICONTHK, ALGOTHK, ALGETHK, FORMTHK, and AXIOTHK.

Phase 4: Subgroup Validity To investigate the discriminant validity of the TDT, we examined whether it functions differently for science, engineering, and education students. To analyze the relevance of these differences—particularly in light of the faculty/school affiliation referring to a student's department within a university/school/college concerned with a major division of knowledge—we sought to replicate the faculty/school affiliation-related differences in students' mathematical thinking that have consistently been found in previous studies (e.g. Bingolbali & Monaghan, 2008; Ubuz & Kırkpınar, 2000; Ubuz, 2011). Bingolbali & Monaghan (2008) remarked that engineering students were more successful than science students in activating ENACTHK and ALGOTHK, whereas science students were more competent than engineering students in activating ICONTHK and AXIOTHK. Ubuz & Kırkpınar (2000) reported that mathematics students were more successful than mathematics education students in activating ALGOTHK. Accordingly, we hypothesized that there would be significant differences among science, engineering, and education faculty students in the accumulation and exchange of various mathematical thinking aspects. Our results of the multivariate analysis revealed a significant main effect for faculty affiliation-related difference (Wilk's lambda = 0.74, F (12, 1516) = 19.99, $p < 0.001$, $\eta^2 = 0.13$) suggesting that science, education, and engineering students differed on a linear combination of the six dimensions of the TDT. The partial eta squared of 0.137 would be interpreted as a large effect (Cohen, 1988). The follow-up univariate analyses indicated that there were significant differences in ENACTHK (F (2, 763) = 66.11, $p = 0.001$, $\eta^2 = 0.14$), ICONTHK (F (2, 763) = 55.41, $p = 0.001$, $\eta^2 = 0.14$), ALGOTHK (F (2, 763) = 83.59, $p = 0.001$, $\eta^2 = 0.21$), ALGETHK (F (2, 763) = 57.37, $p = 0.001$, $\eta^2 = 0.15$), FORMTHK (F (2, 763) = 53.78, $p = 0.001$, $\eta^2 = 0.12$), and AXIOTHK (F (2, 763) = 58.11, $p = 0.001$, $\eta^2 = 0.13$) favoring science faculty students. Science students were more competent than education and engineering students in energizing mathematical thinking aspects.

DISCUSSION AND CONCLUSION

The results of the present research provide support for the reliability and validity of the TDT assessing undergraduate students' mathematical thinking about derivative. CFAs provided evidence for the underlying structure of the TDT. Fit indices of the six-factor model were excellent and standardized estimates for the items on the six aspects were high, suggesting that the TDT with 30 multiple-choice items measures 6 dimensions of mathematical thinking about derivative: ENACTHK, ICONTHK, ALGOTHK, ALGETHK, FORMTHK, and AXIOTHK. Fit indices associated with the competing common factor, three-factor, and null models indicated poor fit, thus providing additional support for the six-factor model. Among the other alternative models, the fit indices for the three-factor model were the closest ones to the six-factor model. This finding supports previous research (Stewart & Thomas, 2009), emphasizing that whether researchers are speaking of—embodiment, symbolism, and formalism, in general, and conceptual-embodied, proceptual-symbolic, and formal-axiomatic thinking, in partial—they hold to the same premise that mathematical thinking echoes the essence of ENACTHK, ICONTHK, ALGOTHK, ALGETHK, FORMTHK, and AXIOTHK.

Reliabilities obtained from the G-study were good or excellent for all six mathematical thinking aspects. Moreover, results from the D-study for the preferred nested two-facet design showed that similar reliability coefficients can be obtained by doubling the number of items in each aspect. Taken together, the generalizability analysis showed that the TDT has high internal consistency for the six mathematical thinking aspects and our choice of measurement design—including 6, 5, 4, 5, 5, and 5 items for FORMTHK, AXIOTHK, ALGETHK, ICONTHK, ALGOTHK, and ENACTHK, respectively—is appropriate for the intended use of the TDT.

For subgroup validity evidence, we found that science students, compared to education and engineering students, were more effective in activating ENACTHK, ICONTHK, ALGOTHK, ALGETHK, FORMTHK, and AXIOTHK. This result is consistent with previous research (Bingolbali & Monaghan, 2008), indicating that science students were more competent than engineering students in activating ICONTHK and ALGOTHK, and further had a tendency to energize AXIOTHK. Our results also agree with that reported by Ubuz & Kırkpınar (2000), who found that mathematics (science faculty) students outperformed mathematics education (education faculty) students in activating ALGOTHK.

This result is in accordance with the fact that faculty affiliation conveys students' mathematical needs and expectations, departments' features and demands, and the higher education curricula (Reid, Wood & Petocz, 2005). This result would indicate that instructors may approach the same mathematical concept from different perspectives in regard to their faculty's vision or students' indicators (e.g. calculus readiness, attitude, and confidence) for success might be different (Pyzdrowski, Sun, Curtis, Miller, Winn & Hensel, 2013). Indeed, the teaching practice in science departments is theoretical-oriented while it is application- and practice-oriented in the engineering and education departments, respectively. This particular focus on the theory-driven aspects, in itself, may provide science students with the opportunity to make transformations within a broader amalgam of visual/spatial aspects (ENACTHK, ICONTHK) and hypothetical/verbal/logical aspects (ALGETHK, FORMTHK, and AXIOTHK).

It is important to consider the implications of the present research for educational and methodological practice. The TDT is a valuable tool for mathematics educators working in research settings to assess students' mathematical thinking about derivative or the relationships among different mathematical thinking aspects (e.g. Aydın & Ubuz, 2014). Furthermore, it can be used to examine the determinants of mathematical thinking aspects (i.e. gender) external to the TDT (e.g. Aydın & Ubuz, 2014). Not only by mathematics education researchers, but also by mathematicians and mathematics teachers who are eager to improve calculus instruction and enhance mathematical thinking aspects the TDT can be used. The structure of students' mathematical thinking at the university level highlights the necessity of research on mathematical thinking in high school levels. Students who are not performing well at these lower educational levels may adopt detrimental mathematical thinking patterns, which in turn may impair their activation of effective mathematical thinking at the university. To gain knowledge about how to prevent such a vicious cycle, it seems highly important to investigate mathematical thinking with respect to precalculus and calculus courses and implement interventions designed to improve students' mathematical thinking at high school/secondary and university/tertiary levels of education. For this purpose, the TDT may serve important diagnostic functions. On the part of implications for assessment, the present research has started a stream of work to develop a multiple-choice test for mathematical thinking about calculus, in general and derivative, in particular. The TDT provides directions that can be specified to account for other topics such as limits, integral, and so forth.

References

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington DC: Author.

Arcavi, A. (1994). Symbol sense: Informal sense-making in formal mathematics. *For the Learning of Mathematics, 14*(3), 24–35.

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics, 52*(3), 215–241.

Artigue, M. (1991). Analysis. In D. O. Tall (Ed.), *Advanced mathematical thinking* (pp. 167–198). Dordrecht, The Netherlands: Kluwer.

Aspinwall, L., Shaw, K. L., & Presmeg, N. C. (1997). Uncontrollable mental imagery: Graphical connections between a function and its derivative. *Educational Studies in Mathematics, 33*(3), 301–317.

Aydın, U. & Ubuz, B. (2014). Predicting undergraduate students' mathematical thinking about derivative concept: A multilevel analysis of personal and institutional factors. *Learning and Individual Differences, 32*, 80–92.

Aydın, U. & Ubuz, B. (2014). Multilevel modeling of undergraduate students' mathematical thinking about the derivative concept. *Educational Studies in Mathematics*, (under review).

Berry, J. & Nyman, M. (2003). Promoting students' graphical understanding of the calculus. *Journal of Mathematical Behavior, 22*(4), 481–497.

Bingolbali, E. & Monaghan, J. (2008). Cognition and institutional setting. In A. Watson & P. Winbourne (Eds.), *New directions for situated cognition in mathematics education* (pp. 233–260). New York: Springer.

Bolger, F. & Wright, G. (1992). Reliability and validity in expert judgment. In G. Wright & F. Bolger (Eds.), *Expertise and decision support* (pp. 47–71). New York: Plenum Press.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.

Bridgeman, B. & Rock, D. (1993). Relationships among multiple-choice and open-ended analytical questions. *Journal of Educational Measurement, 30*(4), 313–329.

Bruner, J. S. (1966). *Towards a theory of instruction*. Cambridge: Harvard.

Burton, L. (1984). Mathematical thinking: The struggle for meaning. *Journal for Research in Mathematics Education, 15*(1), 35–49.

Carlson, M. (1998). A cross-sectional investigation of the development of the function concept. In E. Dubinsky, A. H. Schoenfeld, & J. J. Kaput (Eds.), *Research in collegiate mathematics education, II. Issues in Mathematics Education, 7*, 115–162.

Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education, 33*(5), 352–367.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382–393.

Fischbein, E. (1983). Intuition and analytical thinking in mathematics. *International Reviews on Mathematical Education, 15*(2), 68–74.

Friel, S. N., Curcio, F. R. & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*(2), 124–158.

Gainsburg, J. (2006). The mathematical modeling of structural engineers. *Mathematical Thinking and Learning, 8*(1), 3–36.

García, M., Llinares, S. & Sánchez-Matamoros, G. (2011). Characterizing thematized derivative schema by the underlying emergent structures. *International Journal of Science and Mathematics Education, 9*(5), 1023–1045.

Hähkiöniemi, M. (2006). Associative and reflective connections between the limit of the difference quotient and limiting process. *The Journal of Mathematical Behavior, 25*(2), 170–184.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Erlbaum.

Harel, G. & Tall, D. (1991). The general, the abstract and the generic in advanced mathematics. *For the Learning of Mathematics, 11*(1), 38–42.

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.

Hughes-Hallett, D. (1991). Visualisation and calculus reform. In W. Zimmerman & S. Cunningham (Eds.), *Visualisation in teaching and learning mathematics* (pp. 121–126). Mathematical Association of America Notes No. 19.

Jöreskog, K. & Sörbom, D. (1993). *Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kieran, C. (1996). The changing face of school algebra. In C. Alsina, J. Alvarez, B. Hodgson, C. Laborde, & A. Pérez (Eds.), *8th International Congress on Mathematical Education: Selected lectures* (pp. 271-290). Seville, Spain: S.A.E.M. Thales.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Ko, Y. Y. & Knuth, E. (2009). Undergraduate mathematics majors' writing performance producing proofs and counterexamples about continuous functions. *The Journal of Mathematical Behavior, 28*(1), 68–77.

Lesh, R. & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics teaching, learning, and problem solving*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Martin, T. (2000). Calculus students' ability to solve geometric related-rates problems. *Mathematics Education Research Journal, 12*(2), 74–91.

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist, 34*(4), 207–218.

Mason, J. (1996). Expressing generality and roots of algebra. In N. Bednarz, C. Kieran & L. Lee (Eds.), *Approaches to algebra: Perspectives for research and teaching* (pp. 65–86). Dordrecht: Kluwer Academic Publishers.

Meletiou-Mavrotheris, M., & Lee, C. (2010). Investigating college-level introductory statistics students' prior knowledge of graphing. *Canadian Journal of Science, Mathematics and Technology Education, 10*(4), 339–355.

Moreno, R. & Mayer, R. E. (1999). Gender differences in responding to open-ended problem-solving questions. *Learning and Individual Differences, 11*(4), 355–364.

Moutsios-Rentzos, A. & Simpson, A. (2011). University mathematics students and exam-style proving questions: A-B-Δ strategy classification scheme. *International Journal for Mathematics in Education, 3*, 45–64.

Nagle, C., Moore-Russo, D., Viglietti, J. & Martin, K. (2013). Calculus students' and instructors' conceptualizations of slope: A comparison across academic levels. *International Journal of Science and Mathematics Education, 11*(6), 1491–1515.

Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*. Boston: Kluwer Academic Publishers.

Parke, C. S., Lane, S. & Stone, C. A. (2006). Impact of a state performance assessment program in reading and writing. *Educational Research and Evaluation, 12*(3), 239–269.

Pichat, M. & Ricco, M. (2001). Mathematical problem solving in didactic institutions as a complex system: The case of elementary calculus. *The Journal of Mathematical Behavior, 20*(1), 43–53.

Pyzdrowski, L. J., Sun, Y., Curtis, R., Miller, D., Winn, G. & Hensel, R. A. (2013). Readiness and attitudes as indicators for success in college calculus. *International Journal of Science and Mathematics Education, 11*(3), 529–554.

Raman, M. (2002). Coordinating informal and formal aspects of mathematics: Student behavior and textbook messages. *The Journal of Mathematical Behavior, 21*(2), 135–150.

Rasmussen, C., Zandieh, M., King, K. & Teppo, A. (2005). Advancing mathematical activity: A practice-oriented view of advanced mathematical thinking. *Mathematical Thinking and Learning, 7*(1), 51–73.

Reid, A., Wood, L. N. & Petocz, P. (2005). Intention, approach and outcome: University mathematics students' conceptions of learning mathematics. *International Journal of Science and Mathematics Education, 3*(4), 567–586.

Sharma, S. V. (2006). High school students interpreting tables and graphs: Implications for research. *International Journal of Science and Mathematics Education, 4*(2), 241–268.

Statistical Package for the Social Sciences Inc (2010). *PASW statistics 18*. Chicago: SPSS Inc.

Stacey, K. & MacGregor, M. (2000). Learning the algebraic method of solving problems. *The Journal of Mathematical Behavior, 18*(2), 149–167.

Sternberg, R. J. (1999). *Thinking styles*. New York: Cambridge University Press.

Stewart, J. (2003). *Calculus* (5th ed.). Belmont, CA: Brooks/Cole Thomson Learning.

Stewart, S. & Thomas, M. O. J. (2009). A framework for mathematical thinking: The case of linear algebra. *International Journal of Mathematical Education in Science and Technology, 40*(7), 951–961.

Stylianides, A. J. (2007). Proof and proving in school mathematics. *Journal for Research in Mathematics Education, 38*(3), 289–321.

Stylianides, G. J. (2009). Reasoning-and-proving in school mathematics textbooks. *Mathematical Thinking and Learning, 11*(4), 258–288.

Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson Education.

Tall, D. (2003). Using technology to support an embodied approach to learning concepts in mathematics. In L.M. Carvalho & L.C. Guimarães. *História e Tecnologia no Ensino da Matemática* [History and Technology in the Teaching of Mathematics], vol. 1 (pp. 1–28). Rio de Janeiro.

Tall, D. (2004). Introducing three worlds of mathematics. *For the Learning of Mathematics, 23*(3), 29–33.

Tsamir, P. & Ovodenko, R. (2013). University students' grasp of inflection points. *Educational Studies in Mathematics, 83*(3), 409–427.

Ubuz, B. (1996). Evaluating the impact of computers on the learning and teaching of calculus. *Unpublished PhD Thesis*, University of Nottingham, UK.

Ubuz, B. (2001). First year engineering students' learning of point tangency, numerical calculation of gradients, and the approximate value of a function at a point through computers. *Journal of Computers in Mathematics and Science Teaching, 20*, 113–137.

Ubuz, B. (2007). Interpreting a graph and constructing its derivative graph: Stability and change in students' conceptions. *International Journal of Mathematical Education in Science and Technology, 38*(5), 609–637.

Ubuz, B. (2011). Factors associated with success in a calculus course: An examination of personal variables. *International Journal of Mathematical Education in Science and Technology, 42*(1), 1–12.

Ubuz, B. & Kırkpınar, B. (2000). Factors contributing to learning of calculus. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education*, vol. 4 (pp. 241–248). Hiroshima: Department of Mathematics Education, Hiroshima University.

Villegas, J. L., Castro, E. & Gutiérrez, J. (2009). Representations in problem solving: A case study with optimization problems. *Electronic Journal of Research in Educational Psychology, 7*(1), 279–308.

Zandieh, M. (1997). *The evolution of student understanding of the concept of derivative*. Unpublished doctoral dissertation, Oregon State University.

Zbiek, R. & Conner, A. (2006). Beyond motivation: Exploring mathematical modeling as a context for deepening students' understandings of curricular mathematics. *Educational Studies in Mathematics, 63*(1), 89–112.

Utkun Aydın and Behiye Ubuz

Secondary Science and Mathematics Education
Middle East Technical University
Ankara, Turkey
E-mail: ubuz@metu.edu.tr

Present Address:
Utkun Aydın

Department of Elementary Mathematics Education
MEF University
İstanbul, Turkey