European Journal of Science and Mathematics Education, 2022, **10**(4), 455-470
ISSN: 2301-251X (Online)

OPEN ACCESS **Research Article**

Development of an Item Bank for Measuring Students' Conceptual Understanding of Real Functions

Anela Hrnjičić [1,2*]
0000-0003-3846-2502

Adis Alihodžić [1]
0000-0003-0761-1667

Fikret Čunjalo [1]
0000-0001-5996-3109

Dina Kamber Hamzić [1]
0000-0003-4846-2733

[1] Department of Mathematics, Faculty of Science, University of Sarajevo, Sarajevo, BOSNIA & HERZEGOVINA
[2] Department of Mathematics and Informatics, Faculty of Philosophy, University of Zenica, Zenica, BOSNIA & HERZEGOVINA
[*] Corresponding author: anela.hrnjicic@gmail.com

Citation: Hrnjičić, A., Alihodžić, A., Čunjalo, F., & Kamber Hamzić, D. (2022). Development of an Item Bank for Measuring Students' Conceptual Understanding of Real Functions. *European Journal of Science and Mathematics Education, 10*(4), 455-470. https://doi.org/10.30935/scimath/12222

ARTICLE INFO

Received: 16 Feb 2022

Accepted: 27 Jun 2022

ABSTRACT

It is known that students have many misconceptions about concepts related to function. By discovering misconceptions using an appropriate measurement instrument, we can determine what changes we need to make in the real functions curriculum to improve learning outcomes. Therefore, we designed an item bank for measuring conceptual understandings of real functions with items that require ability to move from one representation of the same concept to another. By surveying university professors, we conducted an expert judgment and content validity of the test. Altogether 36 multiple-choice items based on concepts related to real function with a single correct answer and three distractors have been field-tested by means of a paper and pencil survey, which included 80 freshman students from the Faculty of Science at the University of Sarajevo. By surveying students, we checked technical characteristics of items and their cognitive validity. Results from surveying university professors and students show that the test meets the requirements of content and cognitive validity. Results from the item analysis (item difficulty index, item discrimination index, and point-biserial coefficient) and test analysis (test reliability and Ferguson's delta) show that 32 out of 36 items have good psychometric characteristics, and they are reliable for measuring students' understanding and skills in introductory mathematics courses at universities. We noticed that students have a poor understanding of certain concepts, regardless of the representation, and that there is no coordination between representations of the same concept.

Keywords: real functions, multiple representations, item bank, conceptual understandings, item analysis, test analysis

INTRODUCTION

Major challenge for mathematics education researchers (e.g., Bisson et al., 2016; Code et al., 2014; Crooks & Alibali, 2014) is measuring conceptual understanding with acceptable validity and reliability. To achieve increased conceptual understanding in classrooms there need to be valid and reliable measures of conceptual understanding (Bisson et al., 2016). Test validation is the procedure by which evidence is gathered to determine if the test items satisfactorily represent a concept domain and whether the test measures the properties that it proposes to measure (Day & Bonn, 2011). Reliability indicates to whether a test is consistent within itself and across time and can be measured by using statistical calculations that focus on individual items and the test as a whole (Day & Bonn, 2011). By using a valid and reliable measuring instrument, teachers can see the misconceptions that students still hold about the observed concept and help them overcome these misconceptions. In mathematics education research, students' conceptions are typically explored through oral interviews or written surveys. For assessing the conceptions of a larger number of students (e.g., in a university course), written surveys are more efficient (Mešić et al., 2019). Some authors suggest that it is better to develop comprehensive item banks that can be used for tailoring assessment instruments in order to meet different assessment goals then developing a simple assessment instrument (Bjorner et al., 2007).

The theory of function is of vital importance for learning mathematics (Elia & Spyrou, 2006). Since many phenomena and processes from different areas of the economy and everyday environment are described by functions, it is necessary to examine whether students understand the concepts of functions. A significant indicator of conceptual understanding is ability to represent mathematical situations in different ways and knowing how useful can be different representations for different purposes (National Research Council, 2001). Understanding the concept of function or any other concept entails the ability to recognize at least two different representations of the concept and transition coherently from one to another (Hitt, 1998). Because of the importance of different representations in conceptual understanding, our research focused on three fundamental ways of representing functions: graphical, algebraic, and verbal representation. Although there was some highly valuable research in the field of students' misconceptions in the area of real functions and its properties (Bardele & Ferrari, 2011; Bardini et al., 2014; Bezuidenhout, 2001; Carlson et al., 2010; Dreyfus & Eisenberg, 1982; Elia & Spyrou, 2006; Even, 1998; Gagatsis & Shiakalli, 2004; Habre & Abboud, 2006; Hitt, 1998; Lloyd et al., 2010; Nitsch et al., 2015; Sierpinska, 1992), to the authors' knowledge at this moment there is no real function item bank that would allow for assessing university and high school students' understanding of these concepts related to the real function: function recognition, function's zeros, sign of a function, function evenness/oddness, the limit and asymptotes of a function, the extreme and flow of a function, function convexity/concavity and inflection point.

In this paper we therefore describe the development of conceptual understandings of real functions (CURF) item bank that includes all the above concepts related to the real function and demonstrate its potential for measuring university students' conceptual understanding. We authored and revised multiple-choice items that require manipulation of different representation of the same concept and ability to make a transition from one representation to another. We finally administered these items in a field study and used statistical tests focusing on both item analysis and on the entire test to report on the evidence collected in support of its validity and reliability. Created assessment tool with verified validity and reliability, provides a possibility to measure understanding of concepts across education institutions and over time in a calibrated manner.

Our research questions are:

1. Is the CURF item bank valid and reliable for measuring CURF in introductory mathematics courses at universities?
2. What concepts do students understand and what misconceptions do they have in the area of real function?
3. In which of the transitions between the two representations of the same concept related to the real function are they better?

THEORETICAL BACKGROUND

Understanding the Concepts of a Real Function

In many curricula, the concept of function connects algebra, trigonometry and geometry (Dreyfus & Eisenberg, 1982). The concept of function is complex due to several factors (Dreyfus & Eisenberg, 1982), such as:

- It is not a separate concept, but it is connected with a significant number of sub-concepts (e.g., domain, pre-image, variable, extremes, and growth. We can call them 'functional concepts').

- The concept of function can be used to connect unrelated subjects, such as geometry and algebra. This activity is part of the process of abstraction achieved by using functions.
- The same function may be represented in several settings (e.g., as a table, arrow diagram, graph, formula, or verbal description).

Even/odd functions are essential in many areas of mathematical analysis. They are related to the concepts of symmetry, which most often appears in the visual context (Zazkis, 2014). Bezuidenhout's (2001) research showed that first-year students poorly understand the relationship between the concept of limit, continuity, and differentiation. Research by Tall and Vinner (1981) has shown that students have problems adopting the concept of function limits. Students are not usually allowed to see how the concepts of asymptotes and limit are interrelated (Hornsby & Cole, 1986). Also, previous research has shown that students have problems adopting the concept of function derivation (Orton, 1983). Research by Habre and Abboud (2006) on the conceptual understanding of the function and its derivation showed that students tend to use the first formula and then graphics. Bardelle and Ferrari (2011) examined students' understanding of function monotony. The results showed that freshman students have a poor understanding of the standard definition of increasing function and difficulties in application. The authors are not familiar with previous research on understanding other concepts of the real function of one real variable.

Conceptual Understanding and Multiple Representations of a Real Function

Researchers in mathematics education distinguish between conceptual and procedural understanding in mathematics (e.g., Hiebert & Lefevre, 1986; Schneider and Stern, 2010; Skemp, 1976). Conceptual knowledge is a knowledge of concepts (Rittle-Johnson & Schneider, 2015). In the past, conceptual knowledge has been defined as knowledge rich in connections (e.g., Hiebert, 2013), but recently the richness of connections is viewed as a feature of conceptual knowledge that increases with expertise (Rittle-Johnson & Schneider, 2015). Knowledge of procedures–series of steps and actions done to accomplish a goal is called procedural knowledge (Rittle-Johnson & Schneider, 2015). Knowledge of mathematics includes both conceptual and procedural knowledge. An individual is not competent in mathematics if it is lacking in either kind of knowledge (Hiebert, 2013).

Formal mathematics instruction seems to do a better job of teaching procedures than concepts or relationships between them (Hiebert, 2013). This makes sense, because as Sfard (1991) observes, historically, the operational aspect (procedural knowledge) preceded the structural aspect (conceptual knowledge) and Sfard (1991) argues that the same should be the case for the individual learning process, because the structural approach is more abstract than the operational. However, in schools students often learn only the procedures, without truly understanding the concepts behind them (Even, 1990). This happens due to the lack of meaningful learning, which is connected with conceptual knowledge, while procedures may or may not be learned meaningfully (Hiebert, 2013). Often procedures are learned by heart (rote learning), which means they are tied closely to the context in which they were learned, and this knowledge can only be applied in contexts that look a lot like the original (Hiebert, 2013).

Due to the prevalence of procedural knowledge and sadly rote learning in our schools, we wanted to develop the test items bank that would assess conceptual knowledge of real functions (or more precisely, one important aspect of conceptual knowledge–recognition and transition between different representations of concepts related to functions).

The ability to recognize and represent the same concept differently, the flexibility to move from one representation to another, provides insight into relationships, develops deeper and enhanced conceptual understanding, and strengthens problem-solving skills (Even, 1998). A central goal of mathematics teaching is thus taken to be enabling students to move from one representation to another without falling into contradictions (Hitt, 1998). Principles and standards for school mathematics (NCTM, 2000) include standards that relate exclusively to representations and emphasize using multiple representations in learning mathematics. *The common core mathematics curriculum based on learning outcomes* (APOSO, 2015) within algebra contains the component *algebraic expressions, functions, proportions,* and *applications*. The learning outcome includes analyzing and displaying mathematical situations and structures using algebraic symbols and different notations, graphs and diagrams, and making generalizations. Primary mathematical forms of representation and transition play a central role in subject-related didactics (Nitsch et al., 2015). Different mathematical representations of functional dependence and transitions between them have proven critical factors for successful individual learning (Nitsch et al., 2015).

Gagatsis and Shiakalli (2004) examined the transitional abilities of university students when it comes to the concept of function–the verbal, graphic and algebraic representation of that concept. Students understand two representations of the same concept (verbal and graphic) as different concepts rather than as different ways of representing the same concept (Gagatsis & Shiakalli, 2004). This fact indicates that they do not have a complete and coherent cognitive structure for the concept of function (Gagatsis & Shiakalli, 2004). Habre and Abboud (2006) concluded that for most *calculus 1* course students, algebraic representation dominates when they think about function. Kalchman and Koedinger (2005) noticed that some students lack coordinated conceptual understanding of functions and how they appear in different representations (graphical, tabular, and symbolic representation), and they designed a change in mathematics curriculum, which would teach students to reason about multiple representations of mathematical functions.

In this research, we have concentrated on the verbal, algebraic and graphical representation of content. By verbal representation we mean that the concept is described in written words in the way that is most accessible to students and as described in words in the textbooks. Algebraic representation involves the use of mathematical symbols and representation through mathematical expressions or formulas. The graphical representation of the function represents the geometric representation of the function in the coordinate system. Sierpinska (1992) states that students have difficulties connecting different representations of the concept of function (formulas, graphs, diagrams, and verbal descriptions) when interpreting function graphs and manipulating symbols related to the function. Some authors associate these difficulties with the way concepts are represented in schools. The teacher who understands the concept of function as an action emphasizes function as a chain of operations (Sánchez & Llinares, 2003). Such teachers put more emphasis on algebraic representations and arithmetic activities in teaching. A teacher's goals are determined by different aspects of the featured concept, its specific connections, and how representations are used (Sánchez & Llinares, 2003).

RESEARCH METHODOLOGY

Sample and Sample Size

The student sample was obtained by convenience sampling and it included altogether 80 freshman students of the Faculty of Science of the University of Sarajevo: 39 students from the Department of Information Technology (IT), 25 students from the Department of Theoretical Computer Science (TCS), and 16 students from the Department of Chemistry. 33 students were female and 47 were male. All students attended lectures on *real functions of a real variable–examining and drawing graphs of functions* in the final grade of high school ten months earlier. We conducted our research on a sample of first-year university students because previous research has shown that the performance in elementary mathematics of first-year university students is comparable to the performance of high school students.

Development of an Item Bank

The development of an item bank involved several stages, which mainly follow the procedure proposed by Crocker and Algina (1986, as cited in Liu, 2010):

(i) defining the construct,

(ii) delineating the construct into items,

(iii) item survey,

(iv) final field testing, and

(v) item and test analysis.

The item bank CURF covers the area of *real functions of one real variable–examination and graphing of functions*, which is studied in the final grade of high school (grammar and technical school) and introductory mathematics courses at universities. Based on the learning outcomes, our experience teaching high school

Table 1. Items and concepts related to function

Concepts	Items			
	Card A	Card B	Card C	Card D
The type of a function	A1	B1	C1	D1
The zero of a function	A2	B2	C2	D2
Evenness of a function	A3	B3	C3	D3
Oddness of a function	A4	B4	C4	D4
The sign of a function	A5	B5	C5	D5
Asymptotes	A6	B6	C6	D6
Flow (rise/fall) of the function	A7	B7	C7	D7
Extreme and an inflection point	A8	B8	C8	D8
Convexity/concavity	A9	B9	C9	D9

If you look at the graph of the function, what can you conclude about the sign of the function?

a) The function is negative only on the interval $(-1,0)$, but it is positive only on the interval $(-2,-1)$.

b) The function is negative only on the interval $(-\infty,-1)$, but it is positive only on the interval $(-1,+\infty)$.

c) The function is positive only on the intervals $(-\infty,-2)$ and $(-1,0)$, but it is negative only on the intervals $(-2,-1)$ and $(0,+\infty)$.

d) The function is negative only on the intervals $(-\infty,-2)$ and $(-1,0)$, but it is positive only on the intervals $(-2,-1)$ and $(0,+\infty)$.

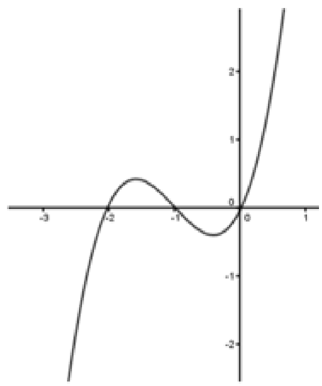

Figure 1. Item A5

and university mathematics, a review of school textbooks, resources and materials from practical and theory classes at the faculty, we defined the construct as a cross-section of the content being studied at university and high school. We focused on the concepts related to functions presented in **Table 1**.

Delineating Constructs into Items

We created items that require the ability to transition from source to target representations of the given concept. We divided 36 items into four cards (each card containing nine items). The cards were organized in the following way:

- Question card A–items require a transition from graphic to verbal representation (the source representation is graphic, and the target representation of the same concept is verbal).
- Question card B–items require a transition from graphic (source) to algebraic representation (target).
- Question card C–items require a transition from algebraic (source) to graphic representation (target).
- Question card D–items require a transition from algebraic (source) to verbal representation (target).

The English version of the question cards is provided at the following address: https://sites.google.com/view/supplementalmaterialaphd.

Most of the items are developed by the authors except items B1 (Bardini et al., 2014) and B6 (Szydlik, 2000). Each item has four offered answers (alternatives), with one correct answer and three distractors (incorrect alternatives). In order to get an insight into cognitive processes, students were expected to explain each chosen answer. During the creation of distractors, we chose situations for which we assumed would trigger student misconception. As an example of representing one concept through cards, we present the concept of the sign of a function trough **Figure 1**, **Figure 2**, **Figure 3**, and **Figure 4**.

Test Validation

The content validation procedure was conducted by questioning five professors from the Faculty of Science at the University of Sarajevo. The professors had to answer three questions for each item:

If you look at the graph of the function $f(x)$, what can you conclude about the sign of the function?

a) $f(x) < 0$, only for $x \in (-1,0)$, $f(x) > 0$ only for $x \in (-2,-1)$.

b) $f(x) < 0$, only for $x \in (-\infty,-1)$, $f(x) > 0$ only for $x \in (-1,+\infty)$.

c) $f(x) > 0$ for $x \in (-\infty,-2) \cup (-1,0)$, $f(x) < 0$ for $x \in (-2,-1) \cup (0,+\infty)$.

d) $f(x) < 0$ for $x \in (-\infty,-2) \cup (-1,0)$, $f(x) > 0$ for $x \in (-2,-1) \cup (0,+\infty)$.

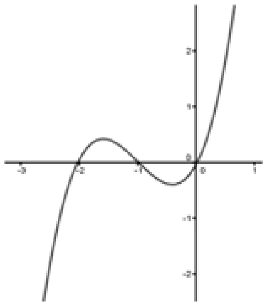

Figure 2. Item B5

Which function $f(x)$ represented in the graph corresponds to $f(x) < 0$ for $x \in (-\infty,-1) \cup (0,1)$, $f(x) > 0$ for $x \in (-1,0) \cup (1,+\infty)$?

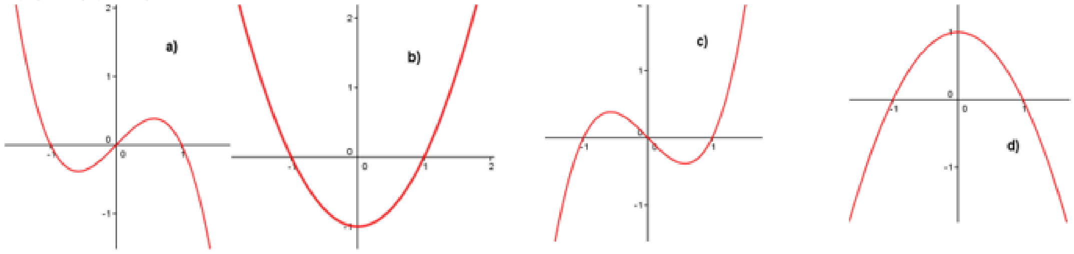

Figure 3. Item C5

What can you conclude about the sign of the function $y = x^3 - x$?

a) The function is always positive no matter what value of x we choose.

b) The function is always negative no matter what value of x we choose.

c) The function is positive on the x intervals $(-\infty,-1)$ and $(0,1)$, but it is negative on the x intervals $(-1,0)$ and $(1,+\infty)$.

d) The function is negative on the x intervals $(-\infty,-1)$ and $(0,1)$, but it is positive on the x intervals $(-1,0)$ and $(1,+\infty)$.

Figure 4. Item D5

1. *Does the item measure student's understanding well?*
2. *Can the item measure the ability of transition from source to target representation of the content?*
3. *How difficult is the item (1 to 3)* (Mešić et al., 2019)?

The questionnaire results showed that in 86% of all cases, the professors agreed that the test item measures the aspect of understanding, and in 87% that it measures the ability to transition from one representation to another. The average difficulty rate of the items was 2.01 on a scale from 1 to 3. In their opinion, the test is of average difficulty level. For card A, on average, the item difficulty was estimated to be 1.88. For card B, it was 2.07; for card C, it was 2.02; and for card D, it was 2.07. Based on these results, we can conclude that card A is slightly more accessible than the other cards. Professors gave suggestions for improving some items (**Table 2**), which we accepted and applied on the items.

Final Field-Testing

The test was conducted in the classroom under the strict supervision of researchers. Through a group interview, we first checked whether the students had learned the concepts related to the function in the final grade of high school. Those students who did not learn about it were excluded at the beginning of the testing. Students voluntarily participated in the research. In the beginning, we emphasized that each question has four answers offered, out of which only one is correct. Students were required to give a written explanation for each of their chosen answer.

Table 2. Professors' suggestions for improving items

Items	Comment
A4, C4	Insert a graph of a function that is neither even nor odd.
A5	In alternatives add "only on the interval…"
A6	In alternatives *a* and *c* use the plural.
A8	Replace points F and A in alternative *d*.
B4	Change alternative *d* to the $f(-x) \neq -f(x)$
B5	In alternatives *a* and *b* add "only".
B7	Use the symbol "∪" instead of "and".
B9	Rephrase the question, improve the visibility of the graph.
C7	Include only the property of the growing function, and omit the first derivative.
D6	In the alternative b add "vertical, oblique and horizontal"
D8	Change the alternative b into: "The maximum of the function is at the point (0, 0)."

Table 3. Questionnaire for each item

Underline every word (including the answers offered) you don't understand or you are not familiar with.
Is any part of the question or the answers offered confusing? If your answer is yes, can you explain what is confusing:
How sure are you of your answer? 1- not sure, just guessing; 2- not sure, but I would exclude the following answer/s:__________; 3- completely sure
How would you rate the level of difficulty of the question: 1 (not at all difficult); 2; 3; 4; 5 (very difficult)?
How much effort did you need to answer the question: 1 (very little); 2; 3; 4; 5 (very much)?
Do you consider the question appropriate for Grammar School students in their final year? YES NO

At this stage we wanted to check wording and clarity of items. For each of the 36 items, students were asked several questions with the purpose to gain additional insight into technical characteristics of items and their cognitive validity (**Table 3**). Some questions in **Table 3** are original, while others are similar to questions in previous research (Mešić et al., 2019). Since this test will also serve as a mock test which will be the base for the test for measuring the understanding of high school graduate students, **Table 3** contains the following question: *Do you consider the question appropriate for grammar school students in their final year?*

Item Analysis

In order to determine which items in the pool of potential test items will be the best to construct the most efficient, reliable, and valid test we conducted an item analysis. After final field-testing of the students, we calculated descriptive statistics: item difficulty index (p), item discrimination index (D), and point-biserial correlation coefficient (r_{pbis}). Item difficulty index (p) presents the percentage of students who got the correct answer compared to the total number of students. A value of p closer to 1 indicates a more accessible item, while a value closer to 0 indicates a more difficult one. The value of the item difficulty index between 0.3 and 0.9 is considered acceptable (Doran, 1980).

In order to determine how well each item in a test distinguishes between higher-achieving and lower-achieving students, according to their test scores, we calculated item discrimination index (D). Ebel and Frisbie (1991) classify items according to the following rules: D>0.4 (excellent discrimination), D between 0.30 and 0.39 (good discrimination), D between 0.20 and 0.29 (acceptable discrimination), while items lower than 0.2 should be excluded or revised. When we want to discover if an item is good, we determine the correlations between the success of students on the item and their success on the whole test (item-total correlations) (Husremović, 2016). For this purpose, we used a point-biserial correlation coefficient (r_{pbis}). The value of the point-biserial correlation coefficient ranges between -1 and 1. We can consider an item to be good if it was correctly solved by students who did well in the entire test or vice versa. The point-biserial correlation coefficient is considered acceptable if $r_{pbis} \geq 0.2$ (Kline, 1986).

Test Analysis

In order to determine test reliability, we calculated Cronbach's alpha, which is the most widely used objective measure of reliability in the case of only one test. George and Mallery (2003) provide the following rules: α>0.9 (excellent reliability), α>0.8 (good), α>0.7 (acceptable), α>0.6 (questionable), α>0.5 (poor), and α<0.5 (unacceptable). In order to measure the discriminatory power of the entire test, we used Ferguson's delta (δ). The discriminatory power refers to the potential of the test to differentiate among the examinees

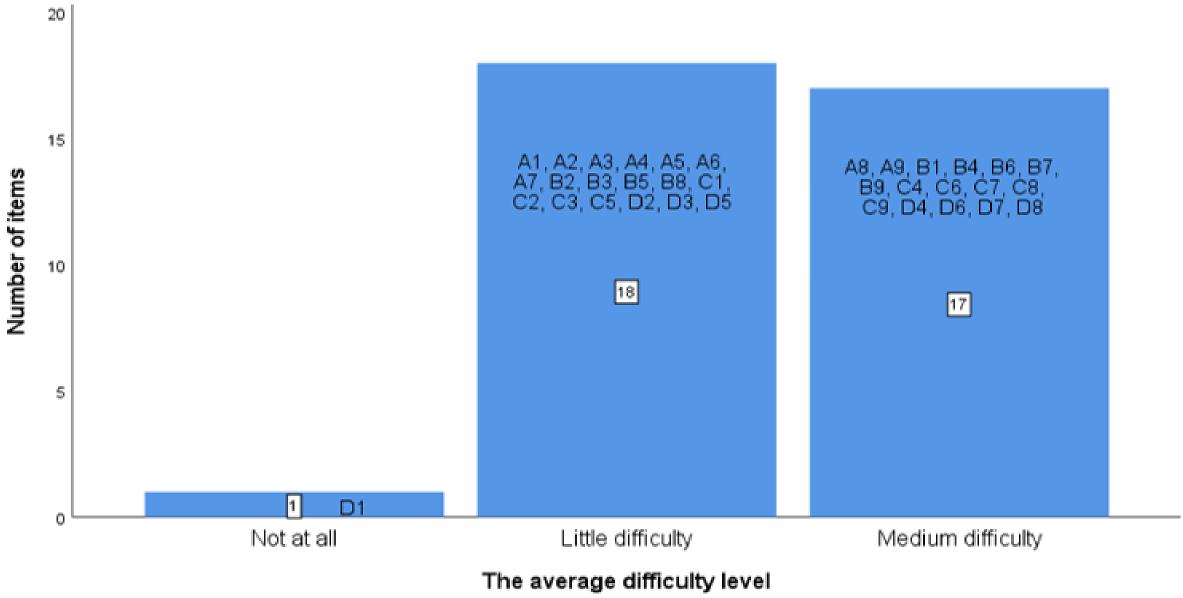

Figure 5. The average difficulty level of the item according to students' opinions

according to a specific characteristic (Husremović, 2016). If the test has a Ferguson's delta greater than 0.90, it is considered to provide a good discrimination, and it is a good test (Kline, 1986). Calculation of Ferguson's delta relies on the relationship between the overall test scores of any two students (Day & Bonn, 2011).

RESULTS

Questionnaire Results

Results from the questionnaire conducted with university professors showed that most of the items measure conceptual understandings and the ability of transition from one representation to another. Professors consider the test to be of moderate severity. Based on written student surveys we found that students understand the questions. For most students, the test is well-formulated and clear. Students rated most of the items as not complex or moderately difficult. Also, the questionnaire results showed that this test does not represent a cognitive load for students. 96% of students claimed no words they did not understand, while the others stated that they were not familiar with words such as 'even', 'odd' and 'inflection point'. For item A2, two students stated that the graph of the function was confusing. Two students reported that they were not familiar with the concept of asymptotes and that the coordinate system was confusing for them. Two students were not familiar with the concepts of convexity and concavity. For item B4, one student stated that answers 'c' and 'd' were confusing. In item B6, the graph was confusing for three students. One student stated that the answers offered in item B9 were confusing. For most items (80.6%), students reported that they needed minimal effort to solve them. 19.4% of items were estimated to have an effort level of 3. Students' opinions about the level of difficulty of the items are shown in **Figure 5**.

Results of Item and Test Analysis

In order to address the first research question, we analyzed items and the test to determine which items in the pool of potential test items will be the best to construct the most efficient, reliable, and valid test. For item and test statistics we used data from final field-testing of the students. We used SPSS 25 software for data processing and coded the students' answers with numbers 0 (incorrect answer), 1 (correct answer) and 9 (missing answer).

The difficulty indices p for each item are shown in **Table 4**. The average value of the item difficulty index is 0.49, which brings it very close to the optimal of 0.5. Items with a difficulty index value outside the recommended range (0.3-0.9) are items A9, B4, B6, B9, C9, and D8 (difficult items) and D1 (accessible item). We can conclude that 29 (90.6%) items have an acceptable difficulty index.

Table 4. Item descriptive statistics for the CURF

Item	N	Item difficulty index (p)	Item discrimination index (D)	Point-biserial correlation coefficient (r_{pbis})	Corrected item-total correlation	Cronbach's alpha if item deleted
A1	79	.54	0.75	0.55	.497	.774
A2	80	.55	0.75	0.51	.363	.780
A3	72	.60	0.3	0.29	.131	.789
A4	73	.38	0.65	0.49	.542	.771
A5	78	.86	0.45	0.39	.291	.784
A6	73	.34	0.3	0.27	.152	.788
A7	78	.83	0.25	0.23	.117	.788
A8	76	.43	0.2	0.24	.159	.788
A9	74	.16	0	-0.01	-.155	.796
B1	78	.60	0.65	0.47	.359	.780
B2	77	.78	0.55	0.4	.363	.780
B3	74	.55	0.6	0.44	.199	.787
B4	72	.18	0.35	0.41	.386	.779
B5	77	.65	0.85	0.6	.663	.766
B6	75	.25	0.15	0.21	.058	.791
B7	75	.45	0.6	0.46	.500	.773
B8	75	.60	0.75	0.55	.372	.779
B9	74	.18	0.25	0.29	.211	.786
C1	79	.76	0.45	0.37	.280	.783
C2	75	.73	0.35	0.22	.188	.787
C3	76	.45	0.55	0.44	.348	.780
C4	76	.43	0.5	0.43	.475	.774
C5	77	.77	0.45	0.37	.219	.785
C6	75	.36	0.35	0.31	.236	.785
C7	68	.32	0.4	0.44	.420	.777
C8	74	.61	0.75	0.53	.431	.777
C9	68	.21	0.15	0.16	.029	.792
D1	76	.91	0.25	0.22	.201	.786
D2	74	.57	0.7	0.52	.363	.780
D3	74	.49	0.6	0.47	.487	.774
D4	72	.36	0.65	0.49	.272	.783
D5	74	.51	0.65	0.49	.285	.783
D6	70	.33	0.35	0.22	-.114	.798
D7	69	.33	0.2	0.2	.158	.788
D8	67	.18	-0.2	-0.2	-.003	.791
D9	65	.45	0.25	0.13	.030	.794

Based on the results in **Table 4**, 26 items have good or excellent discrimination index ($D \geq 0.3$). Lower discrimination index values indicate that there is no difference between higher-achieving and lower-achieving students. It implies that the item is not well formulated or is too difficult or too easy (Ding & Beichner, 2009). Item B5 (**Figure 2**) has the highest discrimination index value (0.85), so we can state that this item is the best formulated and defined item because it distinguishes well between higher-achieving and lower-achieving students. Item D8 has the worst discrimination index value (D=-0.2). If we consider its difficulty index value, we can conclude that almost all students solve it incorrectly as it belongs to the group of complex items. The average discrimination index value is D=0.44, which represents excellent discriminatory power.

Based on the r_{pbis} coefficient from **Table 4**, we can conclude that most items in terms of quality belong to the category of very good items. Only four items (A9, C9, D8, and D9) have a point-biserial coefficient $r_{pbis}<0.2$, so they belong to the category of bad items and have an unacceptable correlation coefficient. Such items should be excluded or revised. The average point-biserial coefficient is 0.35 (acceptable).

In order to produce the trustworthiness of the test, we used Cronbach's alpha coefficient to determine the internal consistency of the item scale. The initial Cronbach's alpha coefficient for the whole scale is 0.788, which shows that the test has good reliability. The discriminatory power of the test is determined using Ferguson's delta. Ferguson's delta is δ=0.9716, which shows that our test is good and has good discriminatory power.

Table 5. A list of the most frequently chosen distractors

Item	A4	A8	A9	B6	C7	C9
Distractor	c (36.3%)	c (51.3%)	b (35%)	b (36.3%)	c (37.5%)	b (37.5%)

Results of Distractor Analysis

In order to get an insight into the quality of the distractor, we conducted a distractor analysis. Through distractor analysis we addressed the second research question, since the choice of a distractor as the correct answer is an indicator of students' difficulties and misconceptions. Distractor analysis is fundamental in the analysis of multiple-choice items. A good item offers equally attractive alternatives to students who do not know the answer (Husremović, 2016). Distractors should be clearly defined so that they are not confusing to higher-achieving students. The total number of distractors is 108. A distractor that was chosen by less than 5% of students was considered a non-functional distractor (NF-D) (Tarrant et al., 2009). According to this criterion, we detected 13 NF-Ds. We will only consider the distractors chosen by at least 35% of the students as indicators of students' difficulties and misconceptions (Mešić et al., 2019). We have detected six items that met this criterion (**Table 5**).

Distractor 'c' in item A4 was chosen by as many as 36.3% of students, while the correct alternative has been chosen by 35% of students. Based on the results, we can conclude that most students cannot recognize the graph of an odd function or its symmetry. Based on the answers in item A8, we can conclude that most students know how to determine the point of the local minimum and local maximum from the graphs, but they do not know how to determine the inflection point. By choosing the distractor 'c', most students think the inflection point is on the x-axis. In item A9, most students chose the distractor 'b', while only 15% of students chose the correct answer. As item C9 is also related to the concepts of convexity and concavity, most students do not understand the connection between the sign of the second derivative of the function with the convexity/concavity of the function. Although they were required to choose a concave graph in one interval and convex in another based on the second derivative sign, most students chose a graph of the increasing function. Based on the results in item B6, we see that most students do not know how to determine the values of the function limit based on graphs. They did not recognize that the function has a horizontal asymptote y=0, so the limit of the function would be zero. Most students do not understand the algebraic notation of an increasing function (item C7), so they did not even choose a graph of an increasing function. Consequently, they have not connected the algebraic and graphic representation of the increasing function.

Students' Success on the Test and Understanding of Concepts

By analyzing students' explanations provided for the selected answer and their success on the items, we gained insight into their understanding of the functional concepts. The test was successfully solved by 32.5% of students (their total test score is higher than 18). The first question in each card referred to recognizing the type of function (square, exponential or cubic). Generally, students were successful in all four items, but they achieved the best results in items describing quadratic function and were less successful in items with exponential and cubic functions.

When it comes to the understanding of the concept of zero of the function, most students understand graphical, algebraic, and verbal representation of the zeros of the functions. Students who chose the correct answer mainly provided the correct explanations. Students who chose the wrong alternative mainly provided the following wrong explanations: 'the graph has three zeros because it intersects the x-axis at two points and the y-axis at one point', 'the function must pass through the origin'. In item D2, where an algebraic expression for a rational function was given, 16 students based their explanation of the incorrect alternative on the denominator, claiming that the zeros of functions are examined using denominators and equating the denominators to zero.

In items referred to the concept of an even function, the best result was achieved in item A3 (60% of correct answers), in which they were required to select the graph of the even function. 45% of students selected the correct graph for the offered algebraic formula of the even function [$f(-x)=f(x)$]. 49% of students induced correctly the parity of the function based on the algebraic notation of the function. Some students provided the following reason of their incorrect alternative: 'it passes through the origin', 'it intersects the x-axis in the origin', 'it is symmetrical with the *x*-axis', 'it must not be negative, and the graph does not go below *x*-axis', 'it does not touch the *x*-axis'. We can conclude that some of the students associated the concept of parity with the zero and sign of the function. Most students did not answer correctly in items that referred to the concept of an odd function. Only 18% of students realized that the graph of the odd function had the following property $f(-x)=-f(x)$. We can also witness that the graph of the odd function is not observed from the aspect of symmetry.

In items that referred to the sign of a function the best results were achieved in item A5 (86% of correct answers), where they were expected to verbally conclude from the graph at which intervals the function was positive/negative. The worst results were obtained in item D5 (51% of correct answers), where they were expected to deduce the sign of the function by calculation based on the algebraic formula of the function. The reason for such a result may be seen in the fact that students were expected to conclude a calculation procedure.

When it comes to the understanding of the limits and asymptotes of the function, only 34% of students successfully found the asymptotes in a graph, while only 33% determined the asymptotes based on the algebraic expression of the rational function. Only 25% of the students were able to determine the values of the limit from the function's graph. Most students could not connect the limits of the function from the provided graph with marked asymptotes. Furthermore, most students (64%) do not know how to determine which graph corresponds to the calculated limit values, although asymptotes are marked on the graphs (item C6).

Most students (83%) answered correctly in the item requiring conclusion from the presented graph in which intervals of the function increases and decreases, but they still do not associate the flow of the function with the sign of the first derivative. We see this in item B7, where the same function graph is given as in item A7. Now, they were expected to conclude from the graph the sign of the first derivative. Also, most of the students (68%) do not understand the algebraic notation of function growth because, in item C7, they were expected to choose a graph of an increasing function since an algebraic notation of an increasing function with an additional representation of the first derivative was given. Only 33% of students chose the correct alternative in item D7, where the function was represented through algebraic notation, and they were assumed to conclude through a calculation process. However, most students chose the correct alternative by guessing, while others entered the point values from the offered intervals. Only two students came to a conclusion based on the first derivative value.

Most students were successful in items C8 and B8 relating to the extremes of the function. Although in item A8, the same graph was given in item B8, 43% of students knew how to describe in words the minimum, maximum, and inflection points from the graph. They were less successful since it was necessary to determine the inflection point. Also, 12 students stated the extremes points correctly but made a mistake on the inflection point (some thought that it must be on the x-axis). In item D8, only two students chose the correct alternative based on the first derivative calculation, while the rest chose a correct alternative by guessing or using the point values.

In items referred to the convexity/concavity of the function students achieved the worst results. Only 16% of students concluded on a given graph at which intervals the function was convex and at which concave. The most common mistake was that they looked to the point (D) on the x-axis and not to the inflection point (C) and that they substituted the terms of convexity and concavity. Only 18% of students concluded from the graph in which intervals second derivative of function is positive and negative. Only three students stated the connection between the second derivative and the convexity/concavity of a function. Only 21% of students took the correct graph based on the sign of the second derivative, but only four students determined the second derivative sign, which determined whether the function was convex/concave. Others guessed or gave incorrect explanations, most often observing whether the graph is below or above the x-axis (in terms of the sign of the function). For the algebraic representation of the function, only three students concluded at which intervals the function was convex and at which concave. They explained that their conclusion was based on the second derivative value, while others guessed the answer.

Table 6. Values of metric characteristics of the original CURF test

Test statistics	CURF values	Desired values
Difficulty index	Average of 0.49	[0.30, 0.90]
Discrimination index	Average of 0.44	≥0.30
Point-biserial coefficient	Average of 0.35	≥0.20
Reliability index	0.79	≥0.70
Ferguson's delta	0.97	≥0.90

Table 7. Evaluations of the CURF after excluding items A9, C9, D8, and D9

Test statistics	CURF values	Desired values
Difficulty index	Average of 0.49	[0.30, 0.90]
Discrimination index	Average of 0.49	≥0.30
Point-biserial coefficient	Average of 0.39	≥0.20
Reliability index	0.85	≥0.70
Ferguson's delta	0.98	≥0.90

DISCUSSION

Characteristics of the Items Bank

In this section, we discussed the obtained metric characteristics of the CURF test to determine if these values can be improved by excluding some items from the test. The metric characteristics of the CURF test were determined through statistical analysis. The results are shown in **Table 6**.

Based on the point-biserial correlation coefficient and the fact that a value of $r<0.2$ means a bad item, we concluded that items A9, C9, D8, and D9 could be considered bad. Items A9, D6, and D8 have a negative correlation with other items. Items A9, B6, C9, and D8 have a low discrimination index. Items A9, B6, C9, D8, and B9, are challenging, while A5, A7, and D1 are straightforward. Thus, items A9, C9, and D8 have the worst results in all the measurements implemented. If we exclude these three items, we get a reliability coefficient of 0.85 (**Table 7**). If we exclude item D9 (bad item, very difficult), then the reliability coefficient is also 0.85, which shows that the test has good reliability ($0.8\leq\alpha\leq0.9$) (George & Mallery, 2003).

However, when deciding to exclude an item, we must consider the content of the item because some items do not have to be highly consistent with others. They may be such that just a few students may solve or round them off, but they still have a place in questionnaire (Husremović, 2016). According to Urbina (2004, p. 130):

> "The degree to which a test is intentionally designed to contain heterogeneous questions cannot be treated as a source of error. The heterogeneity of the test or cognitive functions measured by the test and required to perform the test become a source of error only when the intention was to make the test homogeneous through all or most of the questions."

Two questions that are formulated differently are heterogeneous, and if they measure the same professional interest, they can be homogeneous (Husremović, 2016). As item D8 is related to items A8, B8, and C8 (measures the same concept through different representations), we decided to keep that item and exclude items A9, B9, C9, and D9 because they measure the understanding of the same concept (convexity/concavity). Now our item bank has 32 questions. The number of distractors after excluding A9, B9, C9 and D9 decreased to 96. Now, for the final test, we have 87.5% of functional distractors. The values of test statistics for the final CURF item bank (as presented in **Table 8**) are now slightly lower, but they are still appropriate.

Success Rate Analysis on Cards

In order to address the third research question, we discussed the differences in success rates on question cards. It is crucial to observe the differences in students' abilities in transitions between the two representations. This research has two cards with the same representations but different transitions: card B–from graphic to algebraic and C–from algebraic to graphic. Students were more successful in items that required the transition from algebraic to graphic. The success rates differ in transitions in which the source

Table 8. Evaluations of the CURF after excluding items A9, B9, C9, and D9

Test statistics	CURF values	Desired values
Difficulty index	Average of 0.49	[0.30, 0.90]
Discrimination index	Average of 0.44	≥0.30
Point-biserial coefficient	Average of 0.38	≥0.20
Reliability index	0.84	≥0.70
Ferguson's delta	0.97	≥0.90

representation is the same, and the target representation is different. Students were more successful in transitions from graphic to verbal (card A) than from graphic to algebraic (card B) throughout the test. Also, if we look at cards C and D, where the source representation is the same, and it is algebraic, while the targets are graphic and verbal, we see that students were more successful in transitions from algebraic to graphic than from algebraic to verbal. Thus, they find it difficult to describe in words the algebraic notation of a concept without the help of a graph. Looking at each card separately, they achieved the best result on card A–transition from graphic to verbal. The worst success rate was shown on card D. This means that students have difficulties when they are required to move from algebraic representation to verbal.

Limitation of the Study

The main limitation of this study is related to the fact that the item bank CURF covers only the area of *real functions of one real variable* and can be used for assessing conceptual understandings of graduate students of grammar and technical schools and at universities. Another limitation of this study is that there is room for improvement of distractor since we detected 13 NF-Ds.

CONCLUSION

In this paper we described development of an item bank for measuring conceptual understanding of real functions. The content domain included all the content typically covered in introductory mathematics courses at the university level and at the final grade of high school (grammar and technical school). For purposes of ensuring content validity we implemented and analyzed expert surveys, whereas cognitive validity has been checked through analyses written student surveys. Final field-testing was performed for 36 items. The evaluation of the field-tested items and test as whole was implemented through descriptive statistics. Results of item and test analysis showed that 32 out of 36 items can be combined into a scale that reliably measures students' conceptual understandings of real functions.

Through this study we gained insight into students 'understandings and their misconceptions in the area of real function. Students can recognize the type of function (square, linear, cubic, and exponential), understand the concept of the function zero, sign of the function, concept of an even function, flow and extremes of the function. We found that students have misconceptions when it comes to understanding of concept of an odd function, limits and asymptotes of the function, application of the first derivative, inflection point and convexity/concavity of the function.

Students were most successful in transitions from graphic to verbal representations, while they were least successful in transitioning from algebraic to verbal representations. Thus, students can describe a graph in words, but it is difficult for them to use algebraic notation to describe a situation without a graph of the function. It may suggest that they were more used to describing graphs in words than in algebraic notation through their education. The success rates on the cards differ, which may designate that, in most cases, different representations are viewed as different concepts rather than different representations of the same concept. It means that there is no coordination between representations of the same concept.

The results of this research are significant because math teachers can use this test to check learning outcomes, discover misconceptions, and identify on which concepts focus to help students overcome comprehension difficulties. The conducted research represents the first phase of creating a CURF item bank. In our future research, we plan to conduct additional tests of the item bank CURF. The plan is to create a computer-assisted test and by using Rasch's model check additional characteristics of our item bank. Our goal is to further improve the item bank for better measuring conceptual understandings of freshman and graduate students of grammar and technical schools.

Author contributions: All authors were involved in concept, design, collection of data, interpretation, writing, and critically revising the article. All authors approve final version of the article.

Funding: The authors received no financial support for the research and/or authorship of this article.

Declaration of interest: Authors declare no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

APOSO. (2015). Zajednička jezgra nastavnih planova i programa za matematičko područje definisana na ishodima učenja [The common core mathematics curriculum based on learning outcomes]. *Agency for Pre-Primary, Primary and Secondary Education, Bosnia and Herzegovina, Council of Ministers*. https://aposo.gov.ba/sadrzaj/uploads/ZJNPP-matemati%C4%8Dko-podru%C4%8Dje-BOSANSKI.pdf

Bardelle, C., & Ferrari, P. L. (2011). Definitions and examples in elementary calculus: The case of monotonicity of functions. *ZDM, 43*(2), 233-246. https://doi.org/10.1007/s11858-010-0303-4

Bardini, C., Pierce, R., Vincent, J., & King, D. (2014). Undergraduate mathematics students' understanding of the concept of function. *Indonesian Mathematical Society Journal on Mathematics Education, 5*(2), 85-107. https://doi.org/10.22342/jme.5.2.1495.85-107

Bezuidenhout, J. (2001). Limits and continuity: Some conceptions of first-year students. *International Journal of Mathematical Education in Science and Technology, 32*(4), 487-500. https://doi.org/10.1080/00207390010022590

Bisson, M. J., Gilmore, C., Inglis, M., & Jones, I. (2016). Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education, 2*, 141-164. https://doi.org/10.1007/s40753-016-0024-3

Bjorner, J. B., Chang, C. H., Thissen, D., & Reeve, B. B. (2007). Developing tailored instruments: Item banking and computerized adaptive assessment. *Quality of Life Research*, *16*(1), 95-108. https://doi.org/10.1007/s11136-007-9168-6

Carlson, M., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment (PCA) instrument: A tool for assessing reasoning patterns, understandings and knowledge of precalculus level students. *Cognition and Instruction, 28*(2), 113-145. https://doi.org/10.1080/07370001003676587

Code, W., Piccolo, C., Kohler, D., & MacLean, M. (2014). Teaching methods comparison in a large calculus class. *ZDM, 46*, 589-601. https://doi.org/10.1007/s11858-014-0582-2

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, *34*(4), 344-377. https://doi.org/10.1016/j.dr.2014.10.001

Day, J., & Bonn, D. (2011). Development of the concise data processing assessment. *Physical Review Special Topics-Physics Education Research, 7*(1), 010114. https://doi.org/10.1103/PhysRevSTPER.7.010114

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics-Physics Education Research, 5*(2), 020103. https://doi.org/10.1103/PhysRevSTPER.5.020103

Doran, R. L. (1980). *Basic measurement and evaluation of science instruction*. National Science Teachers Association, 1742 Connecticut Ave., NW, Washington, DC 20009 (Stock No. 471-14764; no price quoted).

Dreyfus, T., & Eisenberg, T. (1982). Intuitive functional concepts: A baseline study on intuitions. *Journal for Research in Mathematics Education, 13*(5), 360-380. https://doi.org/10.5951/jresematheduc.13.5.0360

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement.* University of Iowa.

Elia, I., & Spyrou, P. (2006). How students conceive function: A triarchic conceptual-semiotic model of the understanding of a complex concept. *The Mathematics Enthusiast, 3*(2), 256-272. https://doi.org/10.54870/1551-3440.1053

Even, R. (1990). Subject matter knowledge for teaching and the case of functions. *Educational Studies in Mathematics*, *21*(6), 521-544. https://doi.org/10.1007/BF00315943

Even, R. (1998). Factors involved in linking representations of functions. *The Journal of Mathematical Behavior*, *17*, 105-121. https://doi.org/10.1016/S0732-3123(99)80063-7

Gagatsis, A., & Shiakalli, M. (2004). Ability to translate from one representation of the concept of function to another and mathematical problem solving. *Educational Psychology, 24*(5), 645-657. https://doi.org/10.1080/0144341042000262953

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference.* Allyn& Bacon.

Habre, S., & Abboud, M. (2006). Students' conceptual understanding of a function and its derivative in an experimental calculus course. *The Journal of Mathematical Behavior, 25*(1), 57-72. https://doi.org/10.1016/j.jmathb.2005.11.004

Hiebert, J. (Ed.). (2013). *Conceptual and procedural knowledge: The case of mathematics*. Routledge. https://doi.org/10.4324/9780203063538

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Erlbaum.

Hitt, F. (1998). Difficulties in the articulation of different representations linked to the concept of function. *The Journal of Mathematical Behavior, 17*, 123-134. https://doi.org/10.1016/S0732-3123(99)80064-9

Hornsby, E. J., & Cole, M. (1986). Rational functions: Ignored too long in the high school curriculum. *Mathematics Teacher, 79*(9), 691-697. https://doi.org/10.5951/MT.79.9.0691

Husremović, Dž. (2016). *Osnove psihometrije za studente psihologije* [*The basics of psychometrics for psychology students*]. Faculty of Philosophy, University of Sarajevo.

Kalchman, M., & Koedinger, K. R. (2005). Teaching and learning functions. *How Students Learn: History, Mathematics, and Science in the Classroom*, 351-393.

Kline, P. (1986*). A handbook of test construction: Introduction to psychometric design*. Routledge.

Liu, X. (2010). *Using and developing measurement instruments in science education: A Rasch modeling approach*. IAP.

Lloyd, G., Beckmann, S., Zbiek, R. M., & Cooney, T. (2010). *Developing essential understanding of functions for teaching mathematics in grades 9-12*. National Council of Teachers of Mathematics. 1906 Association Drive, Reston, VA 20191-1502.

Mešić, V., Neumann, K., Aviani, I., Hasović, E., Boone, W. J., Erceg, N., Grubelink, V., Sušac, A., Glamočić, S. Dž, Karuza., M., Vidak, A., Alihodžić, A., & Repnik, R. (2019). Measuring students' conceptual understanding of wave optics: A Rasch modelling approach. *Physical Review Physics Education Research, 15*(1), 010115. https://doi.org/10.1103/PhysRevPhysEducRes.15.010115

National Research Council, Division of Behavioral and Social Sciences and Education, Center for Education, Mathematics Learning Study Committee. (2001). *Adding it up: Helping children learn mathematics* (J. Kilpatrick, J. Swafford, & B. Findell, Eds.). National Academy Press.

NCTM. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.

Nitsch, R., Fredebohm, A., Bruder, R., Kelava, A., Naccarella, D., Leuders, T., & Wirtz, M. (2015). Students' competencies in working with functions in secondary mathematics education–Empirical examination of a competence structure model. *International Journal of Science and Mathematics Education, 13*(3), 657-682. https://doi.org/10.1007/s10763-013-9496-7

Orton, A. (1983). Students' understanding of differentiation. *Educational Studies in Mathematics, 14*, 235-250. https://doi.org/10.1007/BF00410540

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. C. Cohen, & A. Dowker (Eds.), *Oxford handbook of numerical cognition* (pp. 1102-1118). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014

Sánchez, V., & Llinares, S. (2003). Four student teachers' pedagogical reasoning on functions. *Journal of Mathematics Teacher Education, 6*(1), 5-25. https://doi.org/10.1023/A:1022123615355

Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental Psychology, 46*(1), 178-192. https://doi.org/10.1037/a0016701

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1-36. https://doi.org/10.1007/BF00302715

Sierpinska, A. (1992). On understanding the notion of function. In E. Dubinsky, & G. Harel (Eds.), *The concept of function: Aspects of epistemology and pedagogy* (pp. 23-58). Mathematical Association of America.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, *77*(1), 20-26.

Szydlik, J. (2000). Mathematical beliefs and conceptual understanding of the limit of a function. *Journal for Research in Mathematics Education, 31(3),* 258-276. https://doi.org/10.2307/749807

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics, 12*, 151-169. https://doi.org/10.1007/BF00305619

Tarrant, M., Ware, J., & Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. *BMC Medical Education, 9*(1), 1-8. https://doi.org/10.1186/1472-6920-9-40

Urbina, S. (2004). *Essentials of psychological testing*. John C Wiley & Sons.

Zazkis, D. (2014). Proof-scripts as a lens for exploring students' understanding of odd/even functions. *The Journal of Mathematical Behavior, 35*, 31-43. https://doi.org/10.1016/j.jmathb.2014.04.001