ANALYSIS OF A DIDACTIC TEST FOR MEASURING FINANCIAL LITERACY

Vladimir Tirolski

Abstract. *Measuring students' financial literacy knowledge in recent decades has been more important and relevant than ever, as this knowledge is linked to the growing need to make informed financial decisions and understand how they will affect the future [1].*

This article presents an a posteriori analysis of a didactic test, compiled with the purpose of developing and studying the practical and applied knowledge of students on the topic of "Percent", after an experimental educational innovation. The methodology of conducting it is consistent with the principles of objectivity, reliability and validity of measurement. Diagnostic tools, such as the didactic test, are an integral part of the didactic experiment. They represent, in addition to checking the test itself, an objective mechanism for quantitative and qualitative analysis of the effectiveness of a new educational approach.

A consistent policy and methodology for teaching this literacy is needed to achieve sustainable well-being for everyone, starting from early childhood.

Key words: Financial Literacy, Didactic Test, Measuring, a Posteriori Analysis.

Introduction

Didactic tests are a basic tool for assessing knowledge and skills in the modern educational process and are an essential element of pedagogical diagnostics. Through assessment, students and teachers receive information about whether the goals of teaching have been achieved [2]. Didactic tests represent a certain system of judgments about the relationship and dependencies between the phenomena and processes measured by them [3]. The desire to achieve a comprehensive scientific justification of the test as a research method led to the development of two main theories of didactic tests – the classical one, also called normative, and the probabilistic one, also known as stochastic or criterion-based. Table 1 shows the similarities and differences between normative and criterion-based tests. Despite their inherent differences, as outlined in Table 1 [4, 5, 6], the two theoretical frameworks function as complementary approaches rather than as antagonistic concepts.

Table 1. Similarities and differences between normative and criterion-referenced tests

Characteristics	Normative tests	Criterion tests
Main goal	To compare the individual's performance with that of a reference group (norm).	To assess the individual's performance against predetermined standards or criteria.
Focus of assessment	Where does the individual stand in relation to others?	What does the individual know and can do?
Advantages	They allow comparison between individuals, useful for large data sets.	They clearly show what has been learned, and provide specific feedback on the learning.
Disadvantages	They may not reflect individual progress, they do not show specific gaps in knowledge.	Requires clearly defined criteria, may not be suitable for comparison of large groups.

The choice of a didactic test, including practical financial tasks, is dictated by the need to assess not only reproductive knowledge and algorithmic execution of mathematical operations, but also the ability of students to apply what they have learned in real situations.

The present study used a didactic test that integrates characteristics of both norm-referenced and criterion-referenced didactic tests, with the emphasis on criterion-referenced tests.

The following interrelated research objectives were set:

- Measuring financial literacy after differentiated training.

- Preparation of an a posteriori analysis of the didactic test for diagnosing students' practical and applied knowledge of financial literacy.

- Identify specific areas of difficulty and provide data for future adjustments in teaching methodology.

Methodology and description of the experiment

The development of the test tasks is based on the educational content defined by the State Educational Requirements. It is these normative documents, expressed in curricula, syllabi and methodological guidelines, that served as the basis for formulating the goals and objectives of the proposed test.

The testing of the test tasks took place in February – March 2025 among a total of 178 5th grade students from 4 schools in the country – 108 Sofia Secondary School, "Lyuben Karavelov" Plovdiv Secondary School, "Paisii Hilendarski" Plovdiv Secondary School and "Otets Paisii" Stamboliyski Secondary School.

This article describes a pedagogical experiment aimed at determining the level of financial literacy after the use of differentiated instruction and how it

affected the understanding of the concept of percentage by 5th grade students. For the purposes of the study, a didactic test containing 4 tasks was used. Referring to T. Rüter's typology [7] of questions and tasks in the didactic test, we can classify the first task as a semi-open task for transformation by completion, and the remaining 3 – as tasks with multiple-choice answers. The second type of task is also known as structured-response tasks [8]. Although this type of test task has some inevitable disadvantages, such as the probability of accidentally hitting the correct answer and the need for more time to develop them, their diagnostic benefit is significant. In the PISA 2022 [15] financial literacy measurement, this shortcoming is avoided by accepting each response as an individual behavior and way of thinking of the test subjects. Accordingly, there is no right or wrong answer. The purpose of our test covers the study of the understanding of the concept of "percent", which limits the variability of the responses obtained.

The tests were conducted in mathematics classes with a duration of 40 minutes. Fifth-grade students in each school were divided into two groups – experimental and control, with one class (experimental group) studying percentages through the prism of practical financial tasks, and the other (control group) studying the material through a traditional approach. The duration of training for both approaches was identical. Students in the experimental group are introduced to and study the concept of percentage through a system of 3 consecutive lessons, in which they consider real financial processes directly related to the use of percentage. Each lesson contains financial concepts, which are presented in 3 main groups:

- Lesson 1. The concepts of saving, discount, income, expense, budget. The focus is on the basic concepts of personal finance management and budgeting.

- Lesson 2. The concepts of deposit, credit, interest. Students are introduced to the role of banks and basic concepts such as deposit, credit, interest.

- Lesson 3. The concepts of profit, cost. The emphasis is placed on understanding the process of forming the prices of goods, by revealing concepts such as profit and cost.

The tasks in each lesson cover two basic concepts when working with percentages:

- Finding a percentage of a number

- Finding a number from a percentage

The following task was proposed on the test: To make one patty, the baker uses cheese for 0.78 BGN, pie crust for 0.24 BGN, butter for 0.28 BGN, and electricity for 0.26 BGN. If this patty is sold for 2.60 BGN, what percentage is the baker's profit? The other tasks are in the same context.

In addition to the lessons, students in the experimental group are introduced to a presentation with an animated explanation of key financial concepts.

Test analysis

For the purposes of this study, a posteriori analysis was chosen. It allows for a detailed study of the characteristics of the questions and tasks after the test, which is essential for improving future measurement tools. An essential part of the a posteriori analysis is the determination of extreme groups – strong and weak. To the study, a different approach to defining the groups was used, considering the sensitivity to teaching. Two different groups of students are studied, one (experimental) having undergone a differentiated course of study, and the other (control) – a traditional one. Table 2 presents a summary of the results obtained from the analysis of the test tasks.

Table 2. Summary of the results obtained from a posteriori analysis

Task No.	Group	Distribution of responses				Difficulty index	Discriminative power	Total number of students who answered correctly
		A	B	C	D			
1	Experimental	True	61	Wrong	28	64,04%	0,09	61
	Control	True	53	Wrong	36			53
2	Experimental	20	26	22	21	66,67%	0,19	21
	Control	33	33	19	4			4
3	Experimental	9	60	17	3	47,22%	0,15	60
	Control	9	47	20	13			47
4	Experimental	49	19	8	13	30,56%	0,35	49
	Control	18	19	34	18			18

Difficulty analysis

The difficulty of a test item is determined by the percentage of students who answered it correctly. The higher the number of students who answered the item correctly, the lower its difficulty, and vice versa. The difficulty index provides a standardized method for measuring this aspect. G. Linnert [14] proposes the following formula for its measurement:

$$P = 100\frac{N_R}{N},$$

where N_R is number of people from both groups who solved the task correctly, and N is the total number of participants.

The optimal tasks are those whose difficulty index falls in the range from 25% to 80%. The results obtained from the study of the difficulty index are summarized in Table 3.

Table 3. Difficulty index of tasks from the primary test

Group	Task 1		Task 2				Task 3				Task 4			
	True	Wrong	A	B	C	D	A	B	C	D	A	B	C	D
Experimental	61	28	20	26	22	21	9	60	17	3	49	19	8	13
Control	53	36	33	33	19	4	9	47	20	13	18	19	34	18
TOTAL Participated	178		178				178				178			
TOTAL Correctly decided	114		25				107				67			
Difficulty index in %	64,04		14,04				60,11				37,64			

Taks 2 has a difficulty coefficient of 14.04, which places it in the "Very difficult" category of the task. In mathematics education, 5th grade students calculate a percentage as a part of something that is explicitly mentioned in a condition. In this task, there are two consecutive calculations of a percentage, and the base from which it is found is hidden in the condition of the task. That is why the solution to the task requires not only an arithmetic calculation of a percentage of a number, but also the compilation of a mathematical model of the solution.

Tasks 1 and 3 are optimally difficult. Task 4 is difficult but achievable. We believe that the presence of such tasks is advisable for every test because they allow for a more accurate assessment of the discriminative power of a test.

Discriminate Power Analysis

Discriminative power indicates the ability of a task to distinguish between strong and weak students. A score close to 0 indicates that both good and weak students have equal chances to solve. Such a task should be dropped or revised, because both experimental and control tasks can be solved relatively easily. To measure this indicator, we used the formula of L. Gay [9]. Table 4 summarizes the results obtained.

Table 4. Discriminate Power Analysis

Group	Task 1		Task 2				Task 3				Task 4			
	True	Wrong	A	B	C	D	A	B	C	D	A	B	C	D
Experimental	61	28	20	26	22	21	9	60	17	3	49	19	8	13
Control	53	36	33	33	19	4	9	47	20	13	18	19	34	18
TOTAL Participated	178		178				178				178			
TOTAL Correctly decided	114		25				107				67			
Discriminatory power	0,09		0,19				0,15				0,35			

In addition, in tasks of the "Semi-open transformation tasks" type, binding to precise criteria for a correct or incorrect solution is not always achievable. In this case, students must fill in empty fields in a table according to a given model. The question remains controversial as to whether all fields must be correctly filled in for the task to be considered correctly solved, or whether a percentage of inaccuracy in filling in is acceptable.

Distractor analysis

A posteriori analysis allows us to evaluate the effectiveness of distractors and their influence on the validity and discriminatory ability of test items [10]. The effectiveness of distractors is necessary in multiple-choice tasks. A distractor is effective when its discriminative power is negative. To measure the effectiveness of distractors, we used the formula

$$D = (Re/Ne) - (Rc/Nc),$$

where R is number of students from a given group who chose the distractor, and N is the total number of participants in the given group.

Table 5. Distractor analysis

Group	Task 1		Task 2				Task 3				Task 4			
	True	Wrong	A	B	C	D	A	B	C	D	A	B	C	D
Experimental	61	28	20	26	22	21	9	60	17	3	49	19	8	13
Control	53	36	33	33	19	4	9	47	20	13	18	19	34	18
Total participants Experimental	89		89				89				89			
Total participants Control.	89		89				89				89			
Distractors	0,09	-0,09	-0,15	-0,08	0,03	0,19	0,00	0,15	-0,03	-0,11	0,35	0,00	-0,29	-0,06

A distractor is effective when its discriminative power is negative. The rule of thumb is used to arrange distractors – arrange distractors in ascending order.

Reliability

To measure the reliability of the didactic test, we used Cronbach's Alpha [11]. The result $\alpha = 0.828$ indicates internal consistency of the test and an acceptably high level of reliability. Reliability is a necessary but not sufficient condition for the validity of the test.

Validity

The validity of the test is verified through the posterior distribution of abilities, with the discriminatory power of distractors and the reliability of the

estimates being the main indicators of the quality of the measurement [12]. The validity indicator provides information about the extent to which the didactic test measures what it is intended to measure. For the purposes of the study, construct validity was measured. Using the item-by-item method, we analyzed the influence of each of the tasks on the overall construct of the test. The results obtained based on the Classical Test Theory (CTT) [13] and "Alpha if Item Deleted" show that the factor weight of each of the tasks is above 0.50, with the average factor weight of the test being 0.64.

Conclusion

This experiment clearly demonstrates the potential of targeted and innovative learning to improve financial literacy in adolescents. Integrating practical tasks and visual aids, such as animated presentations, significantly improves the acquisition of complex financial concepts and problem-solving skills.

The quality of the test is of utmost importance for the correct diagnosis and correct analysis of the obtained results. A posteriori analysis is a reliable tool for revealing the weak points in the structure or content of the didactic test. The results emphasize the need for a wider introduction of such methodologies into the school curriculum to prepare students for the challenges of the modern financial world.

References

[1] M. Haupt, Measuring financial literacy. The role of knowledge, skills, and attitudes, *The Routledge Handbook of Financial Literacy*, 2022, pp. 79–95, `https://doi.org/10.4324/9781003025221`

[2] F. Di Paola, K. Hoy, *Principal Improving Instruction*, Pearson Education, Inc. Boston, 2008, ISBN: 0-205-49102-2

[3] G. Bijkov, *Theory and methodology of didactic tests*, Prosveta, Sofia, 1992, ISBN 954-01-0107-7

[4] W. Popham, T. Husek, Implications of criterion-referenced measurement, *Journal of educational measurement*, 1969, 6, pp. 1–9, `https://doi.org/10.1111/j.1745-3984.1969.tb00654.x`

[5] I. Ivanov, *Pedagogical diagnostics*, University Publishing House "Bishop Konstantin Preslavski", Shumen, 2006, ISBN:-10:954-577-393-6

[6] F. Stoyanova, *Criterion-Oriented Didactic Tests*, Daniela Ubenova, Sofia, 2008, ISBN: 978-954-791-100-0

[7] T. Rütter, *Formen der Testaufgabe, Handbuch der pädagogischen Diagnostik*, (KJ (Hrsg.), Ed.), Düsseldorf, 1978, ISBN: 3-590-14354-1

[8] R. Thorndike, E. Hagen, Measurement and evaluation in psychology and education, *American journal of psychiatry*, 1969, p. 705, DOI: NLM 0210127

[9] L. Gay, *Educational Evaluation & Measurement: Competencies for Analysis and Application*, Charles E. Merrill Publishing Company, New York, 1980, ISBN: 9780675081436

[10] K. Mendes Vieira et al., Digital Financial Knowledge Scale: Insights from a Developing Economy, *International Journal of Financial Studie*s, 2024, `https://doi.org/10.3390/ijfs12040120`

[11] L. Cronbach, Coefficient alpha and the internal structure of tests, *Psychometrik*, 1951, 16 (3), pp. 297–334, `doi.org/10.1007/BF02310555`

[12] B. Tabak et al., Modeling Financial Literacy Using Multilevel Item Response Theory and the COVID-19 Pandemic, *Journal of Behavioral and Experimental Financ*e, 2022, `doi:http://dx.doi.org/10.2139/ssrn.43 68359`

[13] F. Lord, M. Novick, *Statistical Theories of Mental Test Scores*, Addison-Wesley,1968, Download on 10.11.2025 from `https://psycnet.apa.org/record/1968-35040-000`

[14] G. Lienert, *Testaufbau und Testanalyse (3 ed.)*, Weinheim: Beltz, 1969, Download on 05.11.2025 from `http://d-nb.info/457433566/04`

[15] FINANCIAL LITERACY QUESTIONNAIRE for PISA 2022, Download on 15.12.2025 from `https://www.oecd-ilibrary.org/content/dam/oecd/en/data/datasets/pisa/pisa-2022-datasets/questionnaires/FINANCIAL%20LITERACY%20questionnaire%20PISA%202022.pdf`

Vladimir Tirolski
Paisii Hilendarski University of Plovdiv,
Faculty of Mathematics and Informatics,
236 Bulgaria Blvd., 4027 Plovdiv, Bulgaria
Corresponding author: `vladimirtirolski@uni-plovdiv.bg`