Giornale Italiano di Educazione alla Salute, Sport e Didattica Inclusiva / Italian Journal of Health Education, Sports and Inclusive Didactics - Anno 6 n. 1 - ISSN 2532-3296 ISBN 978-88-6022-436-1 - gennaio - marzo 2022 - CC BY-NC-ND 3.0 IT- https://doi.org/10.32043/gsd.v6i1.526

ANALYSIS OF THE RELATIONS BETWEEN STUDENTS' ABILITY AND MATHEMATICAL PROBLEM SOLVING STRATEGIES IN AN MST TEST

ANALISI DELLE RELAZIONI TRA L'ABILITÀ DEGLI STUDENTI E LE STRATEGIE DI RISOLUZIONE DEI PROBLEMI MATEMATICI IN UN TEST MST

Emanuela Botta[1]
Sapienza University of Rome
emanuela.botta@uniroma1.it

Stefania Pozio[2]
INVALSI
stefania.pozio@invalsi.it

Abstract

Incorrect responses to an item requiring solving a geometry problem were analysed as part of a research project on an adaptive computer-based test assessing mathematical ability (Botta, 2021a). The item bank was calibrated by using the Rasch model (1960) and the item revealed to be of a medium-high level of difficulty. Analyses of incorrect responses allowed to identify and hypothesize students' problem-solving strategies. The responses were categorised using a two-level procedure, related to the basic steps of problem-solving processes. Errors were fairly distant from the correct answer and showed different levels of reasoning consistency. ANOVA was conducted on the mean ability of students to reveal the differences between the categories and relate the students' abilities to the error. The results show that the ability is correlated to motivation to reach the solution and that there are significant differences in the mean ability between the different categories identified.

Nell'ambito di una ricerca su un test adattativo *computer based* costruito per stimare l'abilità matematica di studenti di grado 10 (Botta, 2021a), sono state analizzate le risposte errate a un item che riguardava un problema in ambito geometrico. La banca di item che costituiva il test è stata calibrata secondo il modello di Rasch (1960) e l'item in oggetto è risultato di una difficoltà medio-alta. L'analisi delle risposte errate ha permesso di individuare o ipotizzare le strategie risolutive degli studenti. Le risposte sono state categorizzate con una procedura a due livelli, legati ai passi fondamentali del processo risolutivo. Gli errori sono risultati più o meno distanti dalla risposta corretta, e hanno mostrato vari gradi di coerenza nel ragionamento. Sono poi state effettuate delle Anova sulle medie dell'abilità degli studenti per rilevare differenze fra le diverse categorie e mettere in relazione l'abilità degli studenti con l'errore commesso. Dai risultati si evince che l'abilità è correlata alla motivazione a procedere nella risoluzione.

Key-words

Adaptive Test, MST Test, mathematical ability, problem solving
Prove adattative, MST Test, abilità in matematica, risolvere problemi

[1] Author of Inroduction, Multilevel adaptive testing, Research methodology, Results, Conclusions
[2] Author of. Introduction, Framework, The question, Conclusions

Introduction

In solving a mathematics problem, we always get to a final result that, often, especially in school, is a numerical result. But behind this result, whether correct or incorrect, there is always a reasoning carried out by those who provide this result. Sometimes this reasoning is extremely straightforward, but it may happen that it is sometimes bizarre or even without any logic. The analysis of the errors made by students while solving a mathematics problem shows very interesting aspects and allows us to highlight a series of misconceptions or mental images which are very common among students. If the error analysis is carried out on a very large sample of students, it is possible to draw broader conclusions compared to what a single teacher can draw about the errors made by the students in his/her class. In some cases, this also allows inferences to be made about the type of behavior students have while solving a math problem. Some students try to get to the end of the task, while others are satisfied with a partial result.
This research consists of the analysis and categorization of the incorrect answers that grade 10 students gave to a mathematics question regarding a geometry problem. The categorization of the errors was made by hypothesizing the different reasoning that students might have performed to get to a given answer and the possible solving strategies they have adopted. The purpose of this research is to see if there are significant differences between the ability of students who made different types of errors.

1. Multilevel adaptive testing

The question whose incorrect answers we analyze is part of a multilevel adaptive test (Multistage Test, MST), designed and constructed as part of a consortium research between Sapienza University of Rome and INVALSI (Botta, 2021a). Both classical, item-by-item, and multilevel, module-based, adaptive tests differ from linear tests because students' ability is not assessed only once at the end of the test but is assessed multiple times, during the test (Weiss, 1985; Hambleton et al., 1991; Sireci, 2004). Measuring ability multiple times during the test allows each student to be administered questions with a level of difficulty appropriate to their skill level. This makes the assessment of skill level more accurate and reliable than that provided by a linear test (Weiss, 1985). Therefore, an MST test is characterized by a structure with stages and modules (Luecht et al. 2006). A module is a set of items whose difficulty is within a specific range, centered around a given ability value, and constructed in such a way as to comply with a set of constraints aimed, for example, at ensuring content validity. Each test is divided into levels, or stages, each of which consists of one or more modules. Modules of the same level are centered on different values of the difficulty parameter. The student will take one module for each level of the test. The module to be assigned to the student is chosen during the test based on the answers given so far and the resulting assessment of ability. The set of modules that a given student performs is called a pathway or path.
The design of the test can vary in relation to the number of levels, the number of modules that make up each level, the number of items per module and the paths that are allowed to be taken passing from one level to another.
The implementation of this type of test requires the construction of an item bank (Choppin, 1976) as an item that can significantly influence its structure, particularly in relation to the availability of items with given psychometric or content characteristics. The item bank is generally calibrated in relation to a one or multi-parameter Item Response Theory (IRT) model.

Giornale Italiano di Educazione alla Salute, Sport e Didattica Inclusiva / Italian Journal of Health Education, Sports and Inclusive Didactics - Anno 6 n. 1 - ISSN 2532-3296 ISBN 978-88-6022-436-1 - gennaio - marzo 2022 - CC BY-NC-ND 3.0 IT- https://doi.org/10.32043/gsd.v6i1.526

It has been agreed to divide the set of items into three classes of difficulty, appropriate to obtain three distinct profiles of ability level. The composition of the item bank constructed during the research was taken into account, in particular the availability of items with respect to the difficulty parameter.

The time available for administering the test, the level of precision of the desired ability assessment, and the proven efficiency of the model (Luecht and Nungester, 1998) were the main elements taken into account in choosing a 1 - 3 - 3 MST model, such as the one shown in Fig. 1, in which each pathway is composed of 46 items and balanced with respect to content.

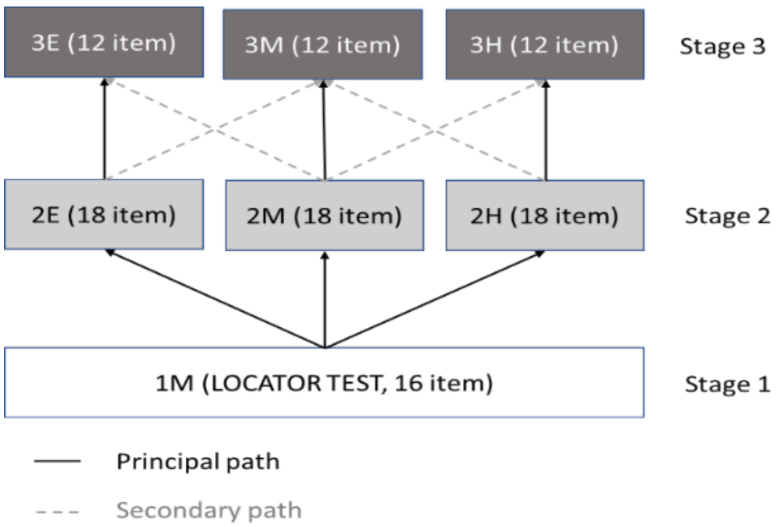


Figure 1. MST 1-3-3 model

The letters E, M, and H indicate the average difficulty of each module (E: easy, M: medium, H: hard).

In terms of navigation within the test, it was established that when passing from one level to another, it was not possible to move forward or backward by more than a single level of difficulty and that it was possible to navigate within each level by reviewing the items and modifying the answers, a function which is inhibited between one level and another.

The paths that can be taken within the test are indicated by arrows.

Each student completes a single pathway consisting of the initial module, a level 2 module, and a level 3 module. As shown in the figure, there are two types of path, the principal paths, indicated by the continuous lines, for which in the transition from level 2 to level 3 the class of assignment of the subject's ability does not change and the secondary ones, for which instead, the class of assignment changes going from level 2 to level 3. In total, 7 different paths are possible, each of which results in the assessment of the student's ability within a given range of the ability continuum. For the optimization of the measurement properties, it has been chosen to maximize the information function of each module. The number of items in each module was defined to ensure real differentiation of paths within the test and to avoid the risk that the ability assessment was predetermined by the initial module. The test has been constructed from a bank of 497 items calibrated according to Rasch's (1960) model.

Table 1 shows the composition of each pathway and the corresponding reference ability range.

Path	Modules	Range of ability θ	Number of students	Lower bound	Upper bound	Center
Path 1	1M+2E+3E	Easy	1065	-2.243	-1.018	-1.631

Path	Modules	Range of ability θ	Number of students	Lower bound	Upper bound	Center
Path 2	1M+2E+3M	Medium	226	-1.018	+0.206	-0.406
Path 3	1M+2M+3E	Easy	162	-2.243	-1.018	-1.631
Path 4	1M+2M+3M	Medium	1023	-1.018	+0.206	-0.406
Path 5	1M+2M+3H	Hard	97	+0.206	+1.430	+0.818
Path 6	1M+2H+3M	Medium	371	-1.018	+0.206	-0.406
Path 7	1M+2H+3H	Hard	1188	+0.206	+1.430	+0.818

Table 1. MST Paths

It is important to notice that the advantage of using an adaptive test comes from the certainty that the proposed item is both difficult enough to be challenging and easy enough to be actually solvable by the student. In other words, it is highly likely that the student is working within their zone of proximal development. It is also noteworthy that in this situation all students have an estimated ability that falls in the middle range. Thus, it is particularly relevant to find any differences in ability with respect to the error made.

2. Theoretical framework

The conceptual framework that will be used to discuss the results of this research has several aspects. Firstly, as this research arises from error analysis, it is intended to emphasize that student errors in mathematics have a very high educational potential and there are many ways to use them to improve learning (Borasi, 1987). Mistakes can be a tool to highlight not only learning difficulties, but also differences in mathematics learning difficulties among different students. It is important to keep in mind that errors are not just a sign that something has gone wrong, and therefore, they must be remedied, because in the history of mathematics it is precisely from errors or wrong guesses that interesting theories have often originated. If a teacher views situations where students make mistakes as unproductive, they are likely to tend to provide them with instructions on how to solve them, rather than supporting them to become "productive protagonists" (Granberg, 2016). Furthermore, more robust learning occurs precisely when students are the true protagonists of problem solving, rather than mere doers (Jonsson et al., 2014).

In this research, we focused specifically on the errors made by students when calculating the perimeter of a rectangle. The task may seem trivial for students in grade 10, but the literature shows that at that level several students still confuse the concept of perimeter with that of area. From research of Chappell and Thompson (1999) carried out on secondary school students, it is very clear how these students confuse the concept of perimeter with that of area because they do not have a conceptual knowledge, but only a procedural knowledge. Conceptual knowledge (Hiebert and Lefevre, 2009) is a knowledge rich in relations, that is, it can be assimilated to a network of interconnected knowledge, a network in which the connecting relations are as important as the individual pieces of information. A unit of conceptual knowledge cannot be an isolated piece of information; by definition it is part of conceptual knowledge only if the owner recognizes its relation to other information. Thus, the development of conceptual knowledge is achieved only through the construction of relations among pieces of information.

On the other hand, Procedural knowledge is composed of two distinct parts. One part consists of formal language, or symbol representation system of mathematics. The other part consists of algorithms, or rules that allow mathematical tasks to be completed. It is precisely the knowledge of these rules, often memorized at school without real conceptual understanding, that leads

Giornale Italiano di Educazione alla Salute, Sport e Didattica Inclusiva / Italian Journal of Health Education, Sports and Inclusive Didactics - Anno 6 n. 1 - ISSN 2532-3296 ISBN 978-88-6022-436-1 - gennaio - marzo 2022 - CC BY-NC-ND 3.0 IT- https://doi.org/10.32043/gsd.v6i1.526

students to confuse area and perimeter. In particular, while attending school we generally learn the formula to calculate the area of the rectangle, not that of the perimeter, and it is the one that students spontaneously apply. In addition, students often confuse perimeter and area precisely because these topics are learned as a set of procedures. If their understanding is tied only to procedures, they may misunderstand these important measurement concepts. Whereas, if a precise meaning is tied to each of these concepts, confusion should no longer occur (Moyer, 2001).

Another situation highlighted in the literature (Ryan, 2007) and confirmed by this research is the one when students need to deal with a problem that requires two steps to get to the solution (as in this case, calculating the side and then the perimeter). Generally, they tend to perform only the first of them and this may be due to lack of concentration, cognitive load, or difficulty in reading.

The original research also found that in the last level of the test, a drop in performance due to cognitive strain or ego depletion may occur in the secondary paths (Botta, 2021b; Kahneman, 2012).

3. The task

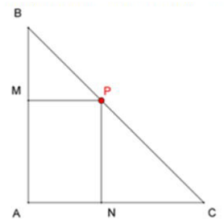
The question we consider in this article requires us to analyze a geometric figure of which some characteristics are given to allow us to derive the data necessary to calculate the perimeter of a rectangle. The question belongs to the content area Space and Figures and to the cognitive dimension Problem Solving. It is a closed constructed response item, where the correct answer corresponds to a number, in this case 24. As shown in Fig. 2, the item is formulated by using different registers (words, figures, symbols) and it is divided into an initial stimulus and a subsequent task. The stimulus is provided primarily in textual and figural form, although numbers (12, 5) and letters (P, BC, etc.) also appear. The initial stimulus describes the figure, an isosceles right triangle whose side measurements are given. However, the numerical data present in the text are not given in the figure. Several specific mathematical terms are used in the textual portion, some of which, such as isosceles, catheti, and hypotenuse, are commonplace for a grade 10 student. The term projection, on the other hand, is certainly familiar to most students, but not all students may remember the geometric meaning. Although the presence of the term in the text of the question may be disorienting, it should be noted that the problem can be solved even without knowing the meaning of projection; in fact, if AMPN is a rectangle the segment PN is perpendicular to AC and the segment PM perpendicular to AB.

The psychometric characteristics of the item are: proportion of correct responses $p = 0.41$, discrimination (biserial point correlation) $R = 0.39$, difficulty $b = -0.009$. In the 3M module, the mean difficulty of the items is -0.402 and the range of difficulty varies between -0.907 and 0.166. The item in question is the penultimate, thus the eleventh in ascending order of difficulty. The item was also present in the Invalsi tests, where it was found to have level 4 difficulty. It was released in 2020 among the examples of questions describing grade 10 levels: to answer the question it is necessary that the student recognizes and uses properties related to the similarity of triangles.

The student can follow different strategies to solve the problem. For example, starting from the information that M and N are the projections of P on the sides and that AMPN is a rectangle, the student can deduce that MP//AC and PN//AB. From here, by the theorem of Thales according to which a bundle of parallel lines intersecting two transversals determines on them directly proportional classes of segments, they can deduce the following similarities of the

triangles ABC, NPC and PMB; since also AB = AC, by hypothesis on the isosceles triangle ABC also PN = NC and BM = MP = 5. So PN = AB - BM = 7; finally, they must calculate the perimeter of the rectangle (5 + 5 + 7 + 7) cm = 24 cm.
This is one of the possible correct procedures to solve the task, but the students who provided the correct answer did not necessarily carry out this one. Some students may not have taken into account the similarity of the triangles to justify the equality of the segments MP and BM or PN and NC, but relied on simple analysis of the figure where "by eye" there is a perception that these pairs of segments are equal. Unfortunately, we do not have this information because the students were not required to report the process they followed to find the solution to the problem.

The isosceles right triangle *ABC* has sides that measure 12 cm. *P* is a point that belongs to the hypotenuse *BC*. *M* and *N* are the projections of *P* onto *AB* and *AC*. The segment *MP* measures 5 cm.



How long is the perimeter of the rectangle *ANPM*?

Answer: ______ cm

Figure 2. The task

4. Methodology

4.1 The sample

The administration of the MST field test took place between September and December 2018 and involved a sample of 4195 students representative of the population of pupils attending the third grade of secondary school in Italy in the 2018-19 school year, in vocational and literature secondary schools or high schools. It was chosen to administer the test to pupils at the beginning of the third year, rather than at the end of the second, to observe skills and knowledge acquired effectively in the previous two years.
The sample was selected using a two-stage method. In the first phase, a judgment sample of schools was identified in each of the three major geographic areas of Italy, North, Center, South and Islands. In the second phase, a sample of classes was randomly selected in each school, at least two per school. Students who, for technical reasons, could not take or finish the test were then eliminated. Analyses were performed on an actual sample of 4132 students, including 278 students with special educational needs or disabilities. The average ability of the sample was found to be equal to $\theta_{\mu} = -0{,}406$, on the Rasch scale on which the test was calibrated.
The question that is the focus of this research belongs to the 3M module and was administered to 1620 students.

After data cleaning, the actual sample consisted of 1617 students of whom 273 provided the correct answers, 438 did not answer, 51 provided meaningless answers, and 855 provided incorrect answers.

4.2 Error categorization

As explained above, the purpose of the research is to understand if there are significant differences between students' abilities in relation to the type of error made. Since it is a closed constructed item, it was not possible to identify with certainty the origin of each error since there was no automatic feedback or guidance from the teacher during the test (Granberg, 2016), nor the possibility for students to write down the process they followed to get to the given answer. Therefore, categorization of the incorrect answers was made based on the available literature and the authors' expertise.

At the beginning, each author has analyzed each wrong answer trying to identify possible procedures that could have led to its genesis. Subsequently, when the procedures identified by each author were not the same, they tried to establish which of the two proposed ones was the most probable. Finally, once agreement was reached, the final categorization of the incorrect answers was carried out by taking into account the two main steps in solving the proposed problem: finding the measure of the side PN and calculating the perimeter of the rectangle ANPM.

If in the first phase the correct measurement of the side PN was found, in the second phase there are three distinct possibilities: the incorrect determination of the measure of the perimeter of the rectangle, the determination of the measure of the area of the rectangle, instead of its perimeter, the choice not to proceed further and the choice to stop and report only the correct measure of the side. Whereas, if, in the first phase the measurement of side PN was found incorrectly, there are five different ways to proceed: determining the measure of the perimeter of the rectangle incorrectly, inconsistently with the measure assigned to the side, determining the measure of the perimeter of the rectangle consistently with the measure identified for the side, determining the area of the rectangle instead of its perimeter, and moreover inconsistently with the measure of the side, determining the area of the rectangle, instead of the perimeter, but consistently with the measure of the side, and finally choosing not to proceed further by stopping to report only the missing side measure. Some answers, 194, could not be placed in any category.

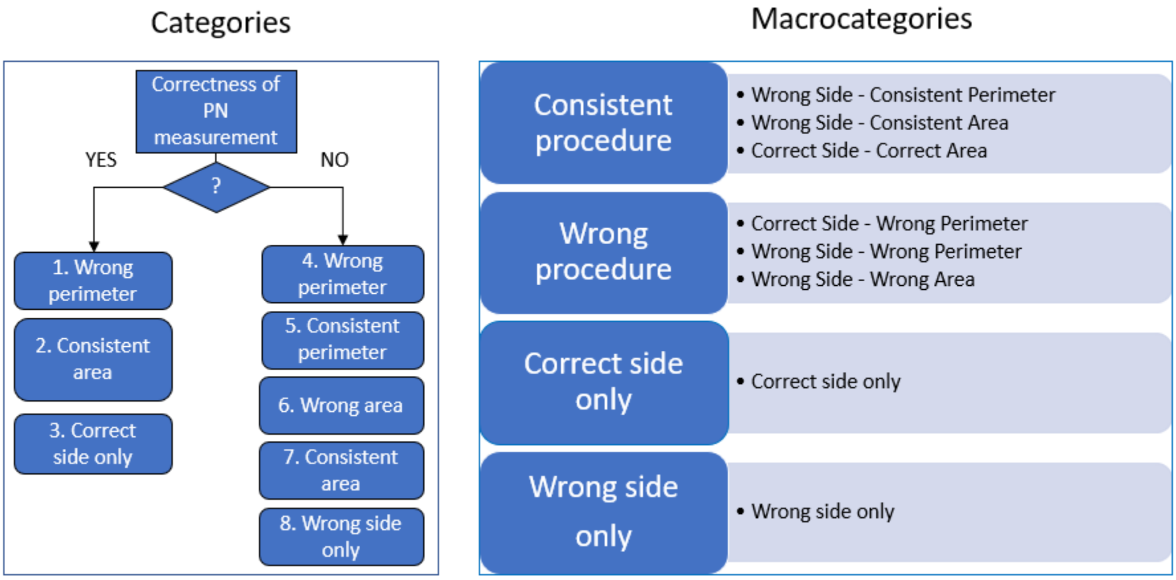


Figure 3. Categories and macrocategories

A further phase of comparison and analysis of the categories identified led us to hypothesize a higher level categorization and to define four macro-categories related to the hypothesis that there could be differences in ability among students who decided to carry out the procedure, until reaching a possible solution anyway, and those who decided not to proceed further. The four macro-categories identified are: one for students who carry on with a procedure consistent with the measure identified for the missing side whether it is aimed at finding the area or the perimeter, one for students who carry on with an incorrect or inconsistent procedure, one for students who stop at finding the correct side, and finally one for students who stop when finding the incorrect side, as illustrated in Fig. 3.

5. Results

For a complete knowledge of the sample in terms of ability and the path taken to find the answer, analyses were performed regarding the distribution of students in each specific path with respect to the type of answer given (correct, incorrect, non-response) and their average ability (Tab. 2) and a subsequent one-way Anova analysis, with multiple comparisons using Bonferroni's method.

	Correct answer		No answer		Wrong answer	
	Number of students	Mean Ability	Number of students	Mean Ability	Number of students	Mean Ability
Path 2	37	-0.623	70	-0.836	113	-0.851
Path 4	178	-0.310	274	-0.658	532	-0.644
Path 6	58	-0.063	94	-0.357	210	-0.302
Total	273	-0.300	438	-0.622	855	-0.587

Table 2. Students' distribution by path and type of answer

After the preliminary analyses and categorization of responses, two one-way Anova analyses with multiple comparisons using Bonferroni's method were conducted to test the hypothesis that there were significant differences between the average abilities of students falling into each category or macro-category.

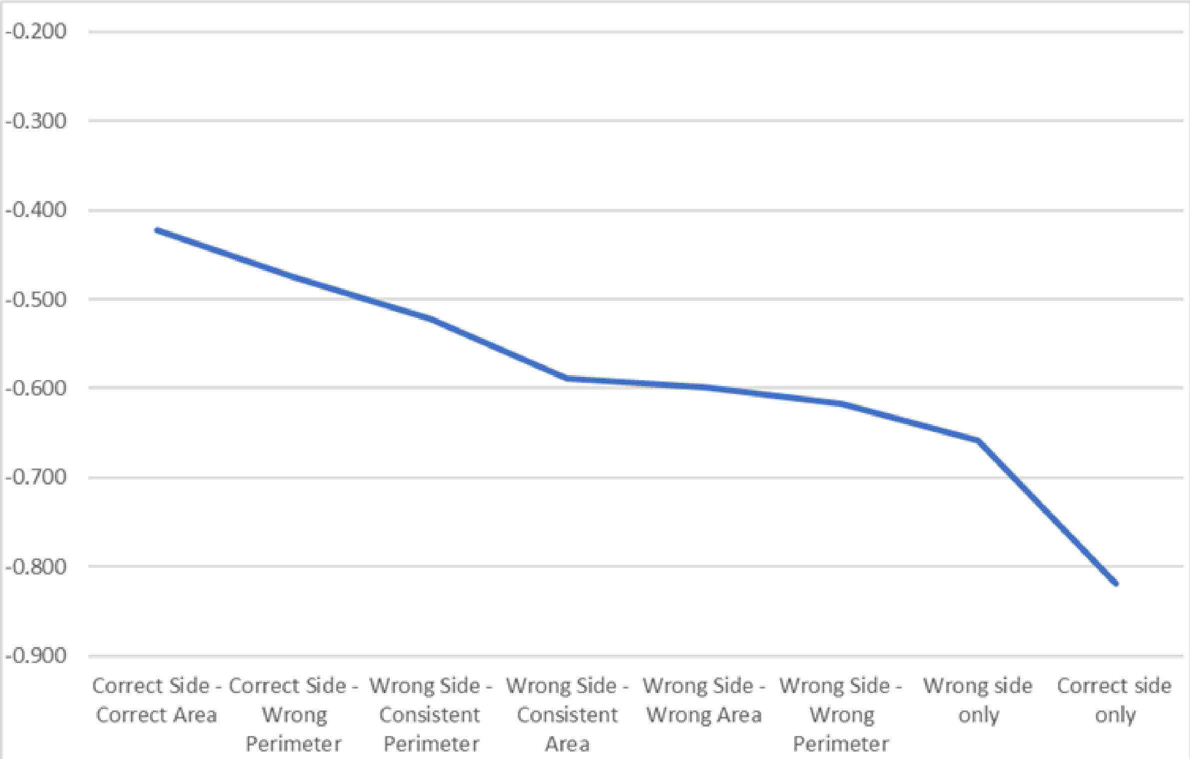
Preliminary analyses revealed that the mean abilities of students who did not answer, provided an incorrect answer, or wrote meaningless statements related to the question were all significantly lower than the mean of those who provided a correct answer ($p < 0.001$). However, it should be noted that even among students who answered correctly there are students who are at or near the lower end of the ability range $\theta = -1.018$. This might lead us to believe that they calculated the perimeter of the rectangle by using an incorrect process to find the missing side, for example without any reflection on the similarity between the triangles but just by manipulating the available data. Moreover, in all groups, the differences between the averages of students of different paths are statistically significant, which confirms the effective functioning of the adaptability of the model and, therefore, its ability to guide students towards the most appropriate path for the assessment of their ability and the accuracy of the latter.
The analyses performed to compare the average abilities of students in the different categories were carried out by eliminating all responses that could not be categorized.
From the Anova and multiple comparisons performed on this sample, it was found that there are significant differences ($p < 0.005$) between some categories, in particular, as shown in Fig. 4, students who simply identify the missing side measure either correct or incorrect still have lower average ability than those who go on to calculate the area of the rectangle instead of its perimeter, or perimeter or area consistent with the value assigned to the side.

Anova Significant differences Row - Column	Correct Side - Wrong Perimeter	Correct Side - Correct Area	Correct side only	Wrong Side - Wrong Perimeter	Wrong Side - Consistent Perimeter	Wrong Side - Wrong Areaa	Wrong Side - Consistent Area	Wrong side only
Correct Side - Wrong Perimeter								
Correct Side - Correct Area			0.396					0.236
Correct side only		**-0.396**			**-0.297**		**-0.230**	
Wrong Side - Wrong Perimeter								
Wrong Side - Consistent Perimeter			**0.297**					0.136
Wrong Side - Wrong Area								
Wrong Side - Consistent Area			**0.230**					
Wrong side only		**-0.236**			**-0.136**			

Figure 4. Anova by categories

From Graph.1 we can also observe how the different categories are arranged in a descending order with respect to the average ability of the students who belong to them. The highest average ability is recorded for students who, once they have found the correct measure of the missing side, proceed to calculate either the area of the rectangle or its perimeter but in an incorrect way, for example they calculate the semi-perimeter or forget one of the sides of the rectangle. The lowest level of ability is for students who merely identify the measurement of the missing side correctly or incorrectly. It should be noted that, although there is a difference in ability between the last two groups of students, it is not statistically significant.

Giornale Italiano di Educazione alla Salute, Sport e Didattica Inclusiva / Italian Journal of Health Education, Sports and Inclusive Didactics - Anno 6 n. 1 - ISSN 2532-3296 ISBN 978-88-6022-436-1 - gennaio - marzo 2022 - CC BY-NC-ND 3.0 IT- https://doi.org/10.32043/gsd.v6i1.526



Graph 1. Categories and student's ability

Finally, the analysis conducted across macro-categories shows substantially similar results. As shown in Tab. 3, this further analysis confirms that students who carry on with the procedure, whether consistent with the initial error or not, have significantly higher ability than those who stop at identifying the side measure. The multiple comparisons table shows this result in detail ($p < 0.005$).

Macrocategories		**Difference in mean of ability**	**Std. Error**	**Sig.**	**Confidence interval 95%**	
					Lower bound	**Upper bound**
Consistent procedure	Wrong procedure	0.058	0.036	0.669	-0.038	0.153
	Correct Side Only	0.274^*	*0.068*	*0.000*	*0.094*	*0.453*
	Wrong Side Only	0.114^*	*0.038*	*0.017*	*0.013*	*0.214*
Wrong procedure	Consistent procedure	-0.058	0.036	0.669	-0.153	0.038
	Correct Side Only	0.216^*	*0.072*	*0.017*	*0.025*	*0.407*
	Wrong Side Only	0.056	0.045	1.000	-0.063	0.175

Table 3. Anova by Macrocategories

Conclusions

This research highlights several important aspects. First of all, it confirms how essential is the analysis of errors students make in solving a mathematics problem in order to go and investigate different issues related to the students who carry out such a problem: from the type of reasoning they carried out in trying to answer the question, to the relation between the students' level of ability and the type of answers. In agreement with what has been stated in the

literature, it is necessary to overcome, even in the educational field, the idea that errors have a limited utility. By observing the work of mathematicians, it is possible to understand how erroneous conjectures, unjustified conclusions and distorted results are all necessary steps to find the solution to a problem. The finding that the ability of students who carry on with the solving process to the end is greater than the one of students who stop at a partial solution is an implicit confirmation of the importance of changing the conception of error in education.
Furthermore, the results presented confirm how important it is for teachers to pay more attention to processes rather than findings because for the same error the underlying processes can be very different and indicative of the students' way of reasoning. The link between type of error and student ability can provide a useful indicator in this sense.
Consider, for example, that among the students who answered correctly there are some who are close to the lower limit of the reference range in terms of ability. Therefore, it is likely that they used the wrong process to find the missing side without reflecting on the similarity between the triangles.
The adaptive testing model used for the research, by providing a more accurate measure of students' abilities and giving them a concrete opportunity to answer a significant number of questions (Botta, 2021a), can also provide teachers with the means to act on individual students or show them ways to intervene with targeted planning.

References

Borasi, R. (1987). Exploring mathematics through the analysis of errors. For the learning of Mathematics, 7(3), 2-8.

Botta, E. (2021) a. Sperimentazione di un modello adattativo multilivello per la stima delle abilità in matematica nelle rilevazioni su larga scala, Roma: Nuova Cultura

Botta, E. (2021) b. Percorsi secondari di una prova adattativa multilivello e valutazione formativa. Excellence and Innovation in Learning and Teaching-Open Access, 6(2).

Chappell, M. F., & Thompson, D. R. (1999). Take Time for Action: Perimeter or Area? Which Measure Is It?. Mathematics Teaching in the Middle School, 5(1), 20-23.

Granberg, C. (2016). Discovering and addressing errors during mathematics problem-solving—A productive struggle? Journal of Mathematical Behavior, 42, 33-48.

Hambleton, R. K., Zaal, J. N., e Pieters, P. (1991). Computerized adaptive testing: Theory, applications, and standards. In R. K. Hambleton e J. N. Zaal (Eds.), Advances in educational and psychological testing, Norwell, MA: Kluwer, pp. 341–366

Jonsson, B., Norquist, M., Liljekvist, Y., & Lithner, J. (2014). Learning mathematics through algorithmic and creative reasoning. Journal of Mathematical Behavior, 36, 20-32

Kahneman, D. (2012). Pensieri lenti e veloci. Edizioni Mondadori, Milano, 2012

Luecht, R., Brumfield, T., e Breithaupt, K. (2006). A testlet assembly design for adaptive multistage tests. Applied Measurement in Education, 19(3), pp. 189-202.

Luecht, R. M., e Nungester, R. (1998). Some practical examples of computer-adaptive sequential testing. Journal of Educational Measurement, 35, 229-249

Moyer, S.P. (2001). Using representations to explore perimeter and area. Teaching Children Mathematics, vol 8(1), pag. 52 – 59

Ryan, J. & Williams, J. (2007). Children’s mathematics 4_15: Learning from errors and misconceptions. Maidenhead, England: Open University Press

Rasch G. (1960), Probabilistic Models for Some Intelligence and Attainment Tests. Danish Institute for Educational Research, Copenhagen

Sireci, S. G. (2004). Computerized-adaptive testing: An introduction. In J. Wall e G. Walz (Eds). Measuring up: Assessment issues for teachers, counselors, and administrators, Greensboro, NC: CAPS Press, pp. 685-694

Weiss, D. J. (1985). Adaptive testing by computer. Journal of consulting and clinical psychology, 53(6), p. 774