Errare humanum est! A socio-psychological approach to a “Climbing Mount Fuji” PISA question

Published Version:
Selleri, P., Carugati, F. (2018). Errare humanum est! A socio-psychological approach to a “Climbing Mount Fuji” PISA question. EUROPEAN JOURNAL OF PSYCHOLOGY OF EDUCATION, 33(3), 489-504 [10.1007/s10212-018-0373-1].

Availability:
This version is available at: https://hdl.handle.net/11585/676328 since: 2019-02-27

Published:
DOI: http://doi.org/10.1007/s10212-018-0373-1

Eur J Psychol Educ (2018) 33:489–504
https://doi.org/10.1007/s10212-018-0373-1

Errare humanum est! A socio-psychological approach to a "Climbing Mount Fuji" PISA question

Patrizia Selleri[1] · Felice Carugati[2]

Received: 22 June 2017 / Revised: 7 March 2018 / Accepted: 8 March 2018 /

Abstract There is a consensus that the items proposed by the Program for International Student Assessment (PISA) program allow us to focus on the outcomes of the processes of appropriation and transformation of learning tools at the end of compulsory schooling, particularly regarding the key competencies for lifelong learning and citizenship in digital societies. Taking into account these assumptions, this paper focuses on a fine-grained analysis of the dynamics of students' performance when they are confronted with a question from the mathematics domain in PISA 2012, through the example of the Climbing Mount Fuji item (question 1). In the context of the interaction dynamics (between a student and a research assistant), twelve 15-year-old students from Naples (Campania, Italy) were requested to think aloud when answering the question 1 of the item. Verbatim transcripts of the interactions are analyzed from the point of view of the PISA framework, the mathematical educational framework, and the socio-psychological approach based on didactic contract. The results show that the students involved in this task commit themselves in a complex reasoning, relying on mathematical requirements (e.g., different mathematical procedures) in an attempt to resolve ambiguities in the text, also referring to their everyday school life, activated by the didactic contract implied by the scenario of the question. The interweaving of PISA performances, mathematical procedures, and the socio-psychological approach to test assessment is discussed as a tool for a better understanding of teaching and learning activities.

Keywords Mathematics PISA framework · Mathematical education · Didactic contract · Socio-psychological approach

The Program for International Student Assessment (PISA) represents a long-lasting approach for the assessment of 15-year-old students in mathematical literacy. One of the unique and notable features of PISA is the articulation of mathematical literacy that serves as a conceptual foundation for the project, including a formal definition of mathematical literacy, the mathematical processes, which students use when using mathematical literacy, and the fundamental mathematical capabilities that underlie these processes. The framework describes how mathematical content knowledge is organized into content categories and outlines the content knowledge that is relevant to the assessment of 15-year-old students. For the purposes of the PISA 2012, it is important that the construct of mathematical literacy (which is used to denote the capacity of individuals to formulate, employ, and interpret mathematics in a variety of contexts) should not be perceived as synonymous with minimal, or low-level knowledge and skills. Rather, it is intended to describe the capacity of individuals to reason mathematically and use mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena (OECD 2013a).

The impact of PISA surveys has encouraged many countries to examine the strengths and weaknesses in their own educational systems, attempting to develop policies and social strategies aimed at eliminating identified limitations and introducing best practices.

The PISA designers worked with two related assumptions, which go hand in hand. The first assumption is that the generic competencies of mathematical literacy are universal and independent of cultural and linguistic contexts (Rochex 2006; Romagnoli and Selleri 2011). The second one is that these competencies are stable and consistent across students regardless of the contexts and tests in which they are measured. These two suppositions constitute a substantialist point of view and explain the methodological choice of using item response theory modeling (IRT, Goldstein 2015). The initial aim of such a model is to assess people's performance in relation to different tasks rather than in comparison with other people, unlike other classically standardized procedures. Another aim is to make sure that individual performance assessment is independent from the items' difficulty, because IRT presupposes one could make the same assessment of an individual's competencies regardless of the items that are shown to him/her. Moreover, for all the groups of individuals that pass a test, whatever the contexts in which these individuals live, and whatever the situations in which they are assessed, one would be able to make the same evaluation of the test difficulty (Bottani and Vrignaud 2005).

In a more fine-grained reflection, Hambleton (2005) warns that "non-equivalent tests, when they are assumed to be equivalent, can only lead to errors in interpretation and faulty conclusions" (p. 5). Hambleton identifies three clusters of sources of errors in the PISA framework: (a) cultural/language differences, including construct equivalence, test administration, test format, and timing; (b) technical issues, designs, methods (including sentence structure), and other aspects that might be difficult to translate correctly; the selection and training of translators, the process and piloting of translations; (c) the interpretation of results, including similarities in curricula, student motivation, and the socio-political context.

Conversely, the PISA framework has been particularly criticized for the issue of item construction (Bracey 2009). Yildirim and Berberoğlu's (2009) work, for example, on the translation from English to Turkish, highlighted the existence of uncertainties in the correct interpretation of certain items. These authors compared a subset of PISA 2003 mathematical items and their fairness for students from Turkey vs. students from the USA. They used three different methods for differential item functioning (DIF) detection (Allalouf et al. 1999; Edelen and Reeve 2007), and based on the agreement of these methods, five items (out of 21) were "flagged." Two of these items favored the US students and three items favored the Turkish students. According to these authors, the Turkish educational system places greater stress on algorithmic and computational skills than on higher order cognitive processes. Moreover, the two terms "mean" and "average" share one common word in the Turkish language and this is misleading in that the Turkish students might misinterpret "on average" to mean that they should calculate an arithmetic mean, though the items do not require any such calculation.

It should be emphasized that the combined effect embedded in the overall PISA framework (i.e., substantialist postulate, psychometric model, and comparative or ranking perspective) leads logically to an attempt to reduce or eliminate every potential source of cultural bias in favor of certain countries or certain linguistic or cultural groups. The goal of these designers is therefore to hunt for anything that might blemish the objectivity and accuracy of the survey's comparative scales on which different countries and school systems are ranked. The belief that it is necessary to minimize cultural and linguistic biases is shared by scholars who attempt to do so and who claim that surveys such as PISA are almost unbiased, as well as by scholars who consider such a project as illusory and such a claim as deceptive.

Some socio-psychological considerations regarding the role of didactic contract

Something the PISA designers failed to consider is how to deepen the analysis and interpretation of data (including biases) rather than merely commenting on rankings. According to this position, the biases could be seen not as psychometric nuisances but, rather, as a heuristic call and an opportunity to analyze the specificities of different countries and different cultural and linguistic groups in depth (Bottani andVrignaud 2005; Murat and Rocher 2004). This does not necessarily imply that one must renounce the use of any scale of measurement, as Master (1988) maintains. However, it does suggest that designers of surveys need to develop ways of combining quantitative and qualitative modes of investigation so as not to reduce comparison to ranking. This approach would allow the development of ways of complementing the assessment of students' performances (through common criteria and common scales) with in-depth analyses of differences between countries or cultural linguistic areas beyond different ranks on score scales.

At the same time, despite the established fact that PISA scoring systems allow for the assessment of different levels of students' competencies, whether or not they truly encompass the rich variety of modalities of reasoning used by the students has been questioned (Rochex 2006) and this has prompted many researchers (Owens 2013) to investigate the features of the cognitive processes that guide students to the solution of items. With such a change of perspective, students, items, and meaning become the vertices of a triangle able to explain that a solution is not only an individual cognitive performance, but a setting where cognition and task representations are intertwined (Carugati and Selleri 2014).

A specific issue to be taken into consideration is that PISA designers assume that the way in which they define a context as societal[1] is a necessary and sufficient condition for triggering

[1] The outer-most box in Figure 1.1 (OECD 2013a, p. 25) shows that mathematical literacy takes place in the context of a challenge or problem that arises in the real world. In this framework, these challenges are characterized in two ways. The context categories, which will be described in detail later in this paper, identify the areas of life from which the problem arises. The context may be of a personal nature, involving problems or challenges that might face an individual or one's family or peer group. The problem might instead be set in a societal context (focusing on one's community—whether it be local, national, or global), an occupational context (centered on the world of work), or a scientific context (relating to the application of mathematics to the natural and technological world).

and assessing the competencies related to one's everyday life. Some examples of PISA "societal" units include context of slicing a pizza, litter, rock concerts, walking, and carpentry (OECD 2013a, pp. 50–55). In addition, "societal" is defined as follows:

> Problems classified in the societal context category focus on one's community (whether local, national or global). They may involve (but are not limited to) such things as voting systems, public transport, government, public policies, demographics, advertising, national statistics and economics. Although individuals are involved in all of these things in a personal way, in the societal context category the focus of problems is on the community perspective. […] In identifying contexts that may be relevant, it is critical to keep in mind that a purpose of the assessment is to gauge the use of mathematical content knowledge, processes, and capabilities that students have acquired by age 15 […] National project managers from countries participating in the PISA survey are involved in judging the degree of such relevance (ibidem, pp. 37–38)

We gladly acknowledge that PISA designers are sensitive to societal relevance and to students' interests and lives, but we propose to focus on the PISA units from the students' point of view.

This issue could be approached with reference to Goffman's (1967) papers and books which appeared in the mid-1970s, when they became very popular in intellectual circles and had a strong impact on the way scholars started thinking about even their everyday interactions with other people and institutions. In fact, it was Goffman who started identifying the various "contracts" that bind our interactions with other people in everyday and professional lives. He called them "frames" that can be played in different "keys," (e.g., as comedies or tragedies). Applied to the classroom context, several identifiable "frames," such as "lecturing," "questioning," "reprimand," and "praise" can be identified. The frame of "school questioning" is very different from the frame of "asking a question" in a non-didactic situation. In the latter, the person who asks the question does not normally know the answer (for instance, asking for information on the street; Selleri 2016). A student who comes to school for the first time may be quite astonished that the teacher is asking questions he/she most certainly knows the answers to! When the student does accept this new situation as normal, he/she has already understood the frame and became aware of the existence of a definite didactic contract that binds his/her own and the teacher's behavior, and he/she is ready to enter into the student's role (Carugati and Selleri 2004). Brousseau (1997/2002) developed Goffman's ideas by introducing the notion of "didactic contract"[2] with norms/rules of interpretation, proposed by the approach on the didactic of mathematics within the didactic contract.

All kinds of items and word problems involving numbers (Säljö and Wyndhamn 1987) trigger complex dynamics for classroom routine (Selleri and Carugati 1999; Selleri 2016), such as "discursive routines" (Jungwirth 1993). For example, the teacher proposes content of a mathematical nature, the students interpret it as an opportunity to understand the teacher's goals. In order to view mathematical meanings as a matter of negotiation, it is helpful to take into account the ambiguity of mathematical objects (such as averages, means, arithmetical operations) in classroom activities. If one looks carefully at micro-processes in the classroom,

[2] The didactic contract (Brousseau 1997/2002, pp. 31–32) implies the determination, which is neither written nor clearly stated, of the respective roles of the student and the teacher in the classroom and in relation to knowledge. Brousseau also said that the didactic contract is a "relationship which explicitly determines for a small part, but especially implicitly, what each partner, the teacher and the students, is responsible for managing and in one way or another will be responsible to the other."

the abovementioned topics appear to be ambiguous and in need for interpretation (Resnick 1987). A point of convergence is reached if the teaching activity is, for instance, reconstructed by the student as a questioning routine, where the teacher organizes all the required types of behavior and, in particular, provides the students with verbal and/or non-verbal suggestions (referred to as "suggestive hints") in order to guide them in constructing the correct answer. In other words, the student responds correctly to the teacher's implicit suggestion if, and only if, the student interprets (aligns him/herself) with the teacher's intended suggestion.

Since the mid-1980s, a large amount of research (for instance, Schubauer-Leoni 1986) has confirmed that students are genuinely committed to giving serious answers even to silly (absurd) mathematical questions/problems (Brissiaud 1988; Verschaffel et al. 2000). In these cases, students try to reconsider the mathematical contents in terms of school mathematics. Therefore, when authors write that students dealing with word problems (particularly multiple-choice ones) try to give meaning to a problem, in fact students activate the didactic contract they are engaged in when they learn the student's role during the school years (Carugati and Selleri 2004). Teachers expect students to learn, guided by the educational curriculum and in compliance with the classroom rules. Students expect teachers to teach them the disciplinary content by proposing activities that can be done starting from what students listen to, read, and study. The case of a multiple-choice solution is an example of the presence of a major norm of the didactic contract, according to which a student expects the data of a task or the content of a query (especially in the case of mathematical problems) to be pertinent, necessary, and sufficient to formulate the solution, and the solution is available to the student within the task. In case of questions with five proposed options (i.e., multiple-choice questions), students are allowed a 1/5 chance of being right without making any effort, simply by chance.

One complementary question also concerns the students' different sensitivities to the influence of the didactic contract according to their performance in mathematics. In fact, we know that in some mathematical problems, the higher the level of schooling and the better students are in mathematics, the less chance they have of arriving at correct solutions. For instance, these students can venture into unrealistic proportional calculations in order to meet the requirements of what they believe their teachers' expectations to be, while the medium/lower performers try to find the subjectively best answer within their school toolkit (Säljö and Wyndhamn 1993).

At this point, we could claim that from the students' perspective, PISA units/items can be interpreted as the umpteenth school scenario of a school community. In fact, while the PISA designers have made an effort to implement societal contexts in order to focus on problems relating to the community, they seem to have forgotten (or neutralized) any potential influence of school scenarios, frames, routines, and didactic contracts embedded in the context.

Research questions

According to the abovementioned claims and premises, we are interested in a fine-grained understanding of the dynamics of the students' performance when they are confronted with an item from the mathematics domain at PISA 2012 through the example of the Climbing Mount Fuji item.

We are not interested neither in supporting nor criticizing the PISA units/items per se, but we propose a complementary, if not innovative, way (Owens 2013) of applying PISA mathematical units/items in order to illustrate certain important issues in detail. First of all, how can students go about solving standardized problems, i.e., how have students mastered these mathematical competencies? What strategies do they use to solve problems? What specific kind of errors do they make, according to the mathematics education framework and the didactic of mathematics? We refer to the perennial work of Polya (1945); see also Schoenfeld 1992; Clements et al. 2013; Daroczy et al. 2015). Inspired by this work on the didactics of mathematics, students are expected to understand the problem, then devise a plan, carry it out, and refer back to it (Polya 1945, pp. 5–14).

Moreover, inspired by the socio-psychological approach of the didactic contract, we are interested in the meaning students attribute to problems, what meaning can be drawn from the errors made, and whether indices of the activity of the didactic contract can be found. To sum up, according to PISA, and for the purposes of transparency, we are interested in showing the characteristics of both the correct and incorrect responses made by the students. According to the framework of mathematics education, which approaches do students choose in order to reach the solution? According to the framework of the didactic contract, what are the clues the students try to grasp in order to give meaning to the question: how do students negotiate between mathematical algorithms (evoked by numbers) and the didactic contract evoked by the school community for a mathematical problem?

Current study

Large regional differences were visible across the PISA 2012 mathematics scores when observing performance of students in Italy. In some of them, 15-year-olds in the northwest of the country scored higher than the PISA average (494 points), but in Campania (i.e., the southern region of which Naples is the capital), the students scored 453 points. However, there were huge differences between students attending different types of upper secondary educational tracks (in Italy, students from 14 to 19 years). Students attending classical grammar schools obtained a slightly higher score (500 points) than the international score, while students from technical schools obtained much lower results (between 430 and 440 points). The difference between grammar and technical schools is equivalent to almost two years of schooling, and the scores of students attending other vocational tracks were even lower (between 370 and 380 points). Compared to students from the technical schools, it is as if they had been educated 1.5 years less, and compared to those at grammar schools, they appear to have had 3 years of schooling less.[3]

Methodology

Participants

According to the research questions, a qualitative approach was chosen, based on conversational analysis (Sachs et al. 1974). Twelve 15-year-old students (8 boys and 4 girls) from a

[3] Across the OECD countries, a more socio-economically advantaged students score 38 points higher in mathematics—the equivalent of nearly one year of schooling—than a less advantaged students (OECD 2013b, p. 13).

vocational school involved in a public project (Second Chance Programme[4]) aimed at keeping young school dropouts from extremely deprived backgrounds in the Spanish Quarter of Naples in education were given a question from a mathematics unit that appeared in PISA 2012. The task was presented to each participant in a separate room by a trained research assistant (RA). Students were advised to think aloud (Ericsson and Simon 1998) as to facilitate a deeper understanding of their thought processes while solving the question. The RA has been trained to avoid the implementation of the discursive structure, characteristic of mathematical discourse in instruction (Sinclair and Coultard 1975). The interactions between students and the research assistant were only audiotaped and transcribed verbatim. Video recording was not allowed by the school.

Instrument

The question 1, from the PISA Unit Climbing Mount Fuji (Fig. 1), was chosen for two reasons.

According to the PISA 2012 mathematics framework (OECD 2013a, p. 295), this question received a low score (average 464 points) and was classified at level 2 (out of 6). The intent of the question is described in Table 1.

The choice of such a low scoring question was made in order to offer students a question within their potential ability. In fact, this question was considered to be of medium difficulty, with some 46% of students in the PISA 2012 main survey administration identifying the correct response C. The two most popular wrong choices were (E) (which is obtained by using 27 days instead of 31 + 27 days) with 19% of responses and (A) (a place value error) with 12% of responses (OECD 2013a, p. 49).

Specifically, question 1 requires the calculation of the average number of people per day. The text is simple and clear, requiring a low demand to the students. The strategy does involve finding the number of days from the data provided and using this to find the average number of people. This multiple-step solution requires some monitoring, which is also a step for devising a strategy. The mathematizing demand is low, because the mathematical quantities required are explicitly provided in the question (number of people per day). Demand for a representation capability is similarly low (only numerical information and text are involved). The technical knowledge required includes knowing how to find an average, being able to calculate a number of days from data, being able to perform division (with or without a calculator depending on the country's assessment policy), and rounding the result up or down as appropriate. There is also little demand for reasoning and argument. The rationale of the question is as follows:

> Interpret the text to understand the task; formulate a strategy; define a time period in required units (days), and combine information to devise a method to calculate a daily average; perform the calculation. In terms of processes behind, question 1 requires to formulate the content, defined as quantity; in a societal context (OECD 2013a, p. 38; 48).

A second reason for choosing this question was that it was formulated as a multiple-choice (i.e., five options were provided to the students), which meant that even low-performing students had a 1/5 chance of being right, without any effort, by chance.

[4] More information about the project may be found in the following: https://www.coe.int/t/dg3/socialpolicies/socialcohesiondev/trends_EN.asp; Series of publications Trends in social cohesion, No. 9

Mount Fuji is a famous dormant volcano in Japan.
Mount Fuji is only open to the public for climbing from 1st July to 27th August each year.
About 200,000 people climb Mount Fuji during this time period.
On average, about how many people climb Mount Fuji each day?
A 340
B 710
C 3400
D 7100
E 7400

Fig. 1 Question 1 in the Climbing Mount Fuji item

Results

We will present the results, detailing them from the viewpoint of the PISA framework, the mathematical educational framework analysis of correct and incorrect solutions, and the framework of the didactic contract.

The viewpoint of the PISA framework

Eight students, seven boys (b) and one girl (g), provided a final correct solution (i.e., (C): 3400 people climbing each day). One student gave a response (B) (i.e., 710) and students 4, 6, and 12 considered the response (E) (i.e., 7400) to be true. The time needed in providing a solution varied, with the range from 1 minute to 9 minutes and 15 seconds (Table 2). Further analyses could not establish any link between the time needed to provide with a solution and whether the solution was correct or incorrect.

The viewpoint of the mathematical educational framework analysis

As a complementary way for a deeper understanding of the thinking activity of the students who took part in the study, the following analyses focus on the procedures (Procedures = P) of finding solutions (Polya 1945; Clements et al. 2013; Daroczy et al. 2015).

Table 1 Characteristics of the Climbing Mount Fuji item (OECD 2013a; p. 38)

Description	Identify an average daily rate given a total number and a specific time period (data provided)
Mathematical content area	Quantity
Context	Societal
Process	Formulate

Table 2 Students' characteristics, final solutions, procedures, errors, and duration time

Student	Gender/age	Final solutions (correct: 3400)	Procedures/errors	Duration time
1	b/16	C - 3400	P1 (E1, E2)	2′:45″
2	b/15	C - 3400	P1 (E1, E2)	2′:00″
3	b/14	C - 3400	P2	5′:00″
4	b/15	E - 7400	P1	4′:25″
5	g/15	B - 710	No procedure	3′:30″
6	g/15	E - 7400	P1 (E2)	1′:20″
7	b/15	C - 3400	P1	1′:10″
8	g/15	C - 3400	P1	1′:0″
9	b/14	C - 3400	P2	1′:20″
10	b/16	C - 3400	P1	1′:25″
11	b/15	C - 3400	P1	2′:20″
12	g/14	E - 7400	No procedure (E2, E3)	9′:15″

Note: "b" stands for boy, "g" for girl; P1–P2, for particular solving procedures (E1–E3 for errors) described further in the results

The immediate procedure (P1) The following procedure is the canonical one, a usual didactical approach to this kind of mathematical problems. For this solution, students would use mathematical knowledge and the ability to calculate the following:

- Counting days: July (27) + August (31) = 58 or keeping in mind that July and August have both 31 days, but from August 27 to August 31, there are 4 days missing; therefore, (31 × 2) − 4 = 58 (counting with hands or silently or with calculator)
- Dividing the number of people who have climbed Mount Fuji during the period by the number of days in the period: 200,000: 58 = 3448.275
- Check solution (C) = 3400 (the number closest to 3448.275)

  Four students provided the correct solution (200,000/58), according to the immediate procedure (P1: students 7, 8, 10, and 11). Example 1 illustrates this procedure.

Example 1. Illustration of the P1 procedure (student 10)

> RA: How did you think ...?
> Student: I thought about it ... the division… of 200000 divided the total opening 58 days, and ... so… 3400…

A backward procedure (P2) Using the commutativity of the product (3 × 2 = 2 × 3), which is a property that does not imply division (12: 4 ≠ 4: 12), the student can, through a series of attempts, seek the correct solution through multiplying the total number of days (58) by multiple proposed choices (A, B, C, D, E), going to the number that approaches 200,000. Actual steps would include the following:

- Counting days: July (27) + August (31) = 58 or keeping in mind that July and August have both 31 days, but from August 27 to August 31, there are 4 days missing; therefore, (31 × 2) − 4 = 58 (counting with hands or silently or with calculator)

– Multiplying the number of days by the proposed choices (A, B, C, D, E) and going to the number that approaches 200,000. More specifically, the calculations would include the following:

$$58 \times (A)\ 340 = 19,720$$
$$58 \times (B)\ 710 = 41,180$$
$$58 \times (C)\ 3400 = 197,200$$
$$58 \times (D)\ 7100 = 411,800$$
$$58 \times (E)\ 7400 = 429,200$$

The correct solution is (C) = 3400; because of the rounding procedure, the closest number to 200,000 is 197,200.

The students 3 and 9 have reached the correct solution, avoiding the use of divisions, probably considered more difficult than multiplication (Example 2).

Example 2. Illustration of the P2 procedure (student 3)

Student: .. for example, I do 58 x 710 ...
RA: ... ah ... and you are seeing if ...
Student: … and I see whether or not the number is the right number, you know?

In the following section, we focus on the possible errors (E) students may have come to.

Miscalculation of days (E1) For the 15-year-old students, the number of days of every month should be known, coupled with recall strategies. Even in the case of miscalculation in finding the correct number of days from the data provided (e.g., 56, 57, 59, 60 days), the formulation of the PISA options (i.e., On average, about how many people climb mount Fuji each day?) requires, in all cases, the procedure of rounding the results. The results of the possible divisions are following:

$$200,000 : 56 = 3571$$
$$200,000 : 57 = 3508$$
$$200,000 : 58 = 3448\ (58 \text{ is the correct sum of days})$$
$$200,000 : 59 = 3389$$
$$200,000 : 60 = 3333$$

The correct result employing rounding would therefore be 3400 (Example 3).

Example 3. Illustration of the E1 error (student 2)

Student:... I could do ... e.g. July is a month of 30 days ...
RA: ... how do you think ?
Student: ...ah! 31 ... from July to August ...26 days are missing …I could do 31 + 26; so 57

Focusing on the numbers in the text (E2)

Many students approach exercises quickly, focusing on the numbers in the text to combine them in "as if" plausible way. The numbers in the text are 27; 200,000; and 1; then the consequence is to divide 200,000: 27 = 7407, and then multiplying by 1; a result sustained by the presence of the solution (E) = 7400 (Example 4).

Example 4. Illustration of the E2 error (student 4)

Student: ... I did 200000 divided by 27 and then multiplied by 1

Ambiguities in the text (about/on average) (E3) The use of words and expressions as "about" or "on average" can create reservations for the students, because they introduce an element of uncertainty on the mathematical ground, usually an accurate one. In the framework of mathematics education, the reference to the average usually relates to a precise, exact value. Thus, the students do not find problems that sound like "A circumference with a radius of about 5 cm." In the same way, the arithmetic mean often used to assess students is given by the sum of marks, divided by the number of evaluations; a student who obtains five different evaluations: 5, 7, 4, 8, 6, immediately knows to have a mean of (5 + 7 + 4 + 8 + 6)/5 = 6 (Example 5).

Example 5. Illustration of the E3 error (student 12)

Student: .. then, here is written on average ... so divided, … then… it's not approximate ... in my opinion will be this one…7400

The viewpoint of the didactic contract framework

We proposed to the students a context aimed at creating an alternative to the classroom routines, but the majority of the 15-year-old students has conscientiously followed the classroom routines. During the twelve discursive interactions, the asymmetric relationship between the RA and the students appears to be the basis of interaction, aligned with the expectation offered by the scenario of the didactic contract (Selleri and Carugati 2013).

For example, the student 7 asks for an explicit question to the RA and uses the answer for finding the solution (Example 6).

Example 6. Illustration of an explicit asking question (student 7)

Student: … then I have to do ... July… how many days.. isn't?
RA: … how do you think?
Student: 31+ 27 … can I get the calculator?
RA:. Yes, go ahead …

The following students look for an explicit confirmation that question is similar to the classroom questions, given that classroom tasks have only one response, adequate to student's level (Example 7).

Example 7. Illustration of students seeking confirmation

RA:... so… you have calculated the days, for now, you are now seeing ...
Student 1: But there is a right answer? (provides a correct answer)
[..]
RA: How are you going to proceed?
Student 4: ... I have to do the division, right?... one of these results, ok? (provides an incorrect answer)

As a further example of the existence of the didactic contract, we report the thinking aloud of student 5. She reads the text of the question according to the format normally required in the classroom routines (i.e., reading the text of a problem carefully), but this routine is not enough for activating a solution. The answer is given as a result of a poor quality in mathematical reasoning and is a wrong one, (B) 710 (Example 8).

Example 8. Reading aloud (student 5)

Student: So ... why ... counting that days are not ... so there are enough, though anyway ... more or less is a month… a month and let's say a little ... and half, go and I'm about 200000 people are coming up ... all these days ... a lot of people do not come up every day, because of the days there are ... one month and a half … it's not that many and only get 200000 so I did .. 710
RA:So… you have practically considered the answers ...
Student: The last three are too much … the first is too little…

Mismatching between what is said and what is done

It may happen that a student introduces verbally the following operation:

days=200000

but in fact, the student actually tries to do a reverse division. At that point, he/she announces verbally the sum of days of July plus August, but finally the division is:

200000=26 ðAugust onlyÞ and is wrong:

Example 9. Mismatch between actual procedure and wording (student 6)

Student:… then, practically doing the calculations of July ... that is, I added the days of July ... until August 27
RA: Ok.
Student: ...then I did ... well… the sum… one month plus 26 days… divided by 200000 and I got… 7400; well, but it was approximate, because ... it was periodic…

Discussion

The integrated use of the three frameworks has allowed to differentiate between the six students who obtain a final correct solution according to the PISA requirements (i.e., without errors), and the two students who have made calculation errors (i.e., E1 and E2), before reaching the final correct solution.

Several mathematical procedures and several errors are shown: miscalculations of days, mismatching between what is said and what is done, focusing on the numbers in the question. These results are easily expected by students with complicated and troubled school experiences. Nevertheless, at least in the practice and reasoning of one student, we find an echo of mathematical concepts (e.g., the periodic numbers and the approximation—student 6).

Another observation is linked to students' referencing to "about," "average," and the "rounding procedure" (e.g., students 5 and 10), notions that are ambiguities in the text of the question, but the students try to take them into account during the process of finding a plausible solution. We suppose that these ambiguities have been introduced on purpose by the designers, and we claim that they trigger supplementary difficulties in the test comprehension, but our students implement them and they try to follow diligently the requirements of the didactic contract, despite their rather limited mathematics resources.

Moreover, the PISA designers underline that students have to engage in rounding off the results appropriately (italics by authors). According to the provided key, the correct solution of the item is (C) 3400, but the correct result of the division (with 58 days) is 3448.275 and the division (with 57 days) is 3508.71. Therefore, from the students' point of view, based upon the everyday life, even the most appropriate, plausible answer (given the words "about" and "average") may actually be considered rather ambiguous. In addition, in real-life situations, it is somewhat atypical to use the order of magnitude as an answer. In fact, the difference between the two previous results is only 60 people, nothing compared to 200,000 people!

Concluding remarks

One main limitation does mark this qualitative study. Our interest in the fine-grained analysis of the protocols allowed us to describe the flow of thinking aloud of low-performing students while they are coping with the Climbing Mount Fuji item question, but the size of the sample does not allow us to generalize the findings beyond this subpopulation of the students that were involved in the study. Nevertheless, the results do illustrate a complex interweaving of the text comprehension, the scenario evoked by the question, the mathematical procedures adopted by the students, and the final solutions they provide.

Our way of approaching this texture of elements is in contrast with the claims of PISA designers, when they conceptualize eventual cultural and linguistic specificities in terms of biases and potential technical and methodological dysfunctions. As a result, they try to eliminate every element linked to the question specificities that are not taken into consideration when they assess students and design the model of competencies.

As a contrast, our study demonstrates that students anchor, in fact, their thinking processes on the representation of the situation in terms of didactic contract (e.g., numbers, time, months). The influence of the school scenario seems to be difficult to expunge even when high level of methodological rigor is at work, like the one applied in PISA. Paradoxically enough, just the goal of implementing contexts as societal has the consequence of activating what is perceived by students as genuinely societal: school scenario. Students (particularly low-performing ones) show high sensibility to the details of the scenario of the question: they interpret them as signals to execute a school assignment in the framework of a didactic contract. From the methodological point of view, students express aloud procedures of solution that researchers could benefit knowing about in order to deeply understand the dynamics of teaching-learning activities.

When researchers give students a PISA question, as in our case, the content of the black box of the PISA mathematical framework (i.e., the three mathematic processes—formulate, employ, interpret; the four mathematic content categories—quantity, uncertainty and data, change and relationships, and space and shape and the socio-psychological embedded dynamics) should be deeply analyzed, beyond the final solutions, even if correct answers are provided. As we have shown, final solutions could hide different difficulties of understanding the problems, devising a plan, carrying it out, managing mathematical procedures, and referring back to them. Thus, this socio-psychological approach is manifold. It does offer a complementary look of the PISA toolkits.

With regard to the requirements of the academic research, the approach can be viewed as a vehicle for interpreting learning dynamics, a method for improving mathematics education, and as a "theoretical" tool that may improve teachers' awareness as they work particularly with low-performing students. Finally, it seems to be a valuable framework for research on the fine-grained understanding of mathematical teaching-learning activities.

References

Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. Journal of Educational Measurement, 36(3), 185–198.

Bottani, N., & Vrignaud, P. (2005). La France et les évaluations internationales. (n.16, Janvier, pp. 172.). Paris: Haut Conseil de l'évaluation de l'école.

Bracey, G. W. (2009). PISA: not leaning hard on U.S. economy. Phi Delta Kappa, 90(6), 450–451.

Brissiaud, R. (1988). De l'âge du capitaine a l'âge du berger. Quel contrôle de la validité d'un énoncé de problème au CE2? Revue Française de Pédagogie, 82, 23–31.

Brousseau, G. (1997/2002). Theory of didactical situations in mathematics. New York: Kluwer Academic Publishers.

Carugati, F., & Selleri, P. (2004). Intelligence, educational practices and school reform: organisations change, representations persist. European Journal of School Psychology, 2(1–2), 149–167.

Carugati, F., & Selleri, P. (2014). Piaget, Vygotskij and the European approach in social psychology of education: a space for virtuous dialogue? In T. Zittoun & A. Iannaccone (Eds.), Activities of thinking in social spaces (pp. 43–61). New York: Nova Science Publishers, Inc..

Clements, M. A., Bishop, A. J., Keitel, C., Kilpatrick, J., & Leung, F. K. S. (Eds.). (2013). Third international handbook of mathematics education. New York: Springer.

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. Frontiers in Psychology, 6, 348. https://doi.org/10.3389/fpsyg.2015.00348.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modelling to questionnaire development, evaluation, and refinement. Quality of Life Research, 16, 5–18.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. Mind, Culture, and Activity, 5(3), 178–186.

Goffman, E. (1967). Interaction ritual: essays on face-to-face interaction. Oxford: Aldine.

Goldstein, H. (2015). Rasch measurement: a response to Payanides, Robinson and Tymms. Response on Rasch model. British Educational Research Journal, 41(1), 176–179.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton & P. F. Meranda (Eds.), Adapting educational and psychological tests for cross-cultural assessment (pp. 3–38). Mahwah: Erlbaum.

Jungwirth, H. (1993). Routines in classroom discourse. An ethnomethodological approach. European Journal of Psychology of Education, 8(4), 375–387.

Master, G. N. (1988). Item discrimination: when more is worse. Journal of Educational Measurement, 25(1), 15–29.

Murat, F., & Rocher, T. (2004). The methods used for international assessments of educational competences. In J. H. Moskowitz & M. Stephens (Eds.), Comparing learning outcomes: international assessment and education policy (pp. 190–214). New York: Routgers Falmer.

OECD. (2013a). PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy. Paris: OECD Publishing.

OECD. (2013b). PISA 2012 results: excellence through equity: giving every student the chance to succeed (volume II). Paris: OECD Publishing.

Owens, T. (2013). Thinking beyond league tables: a review of key PISA research questions. In H. D. Meyer & A. Benavot (Eds.), PISA, power and policy: the emergence of global educational governance (pp. 27–49). Southampton: Symposium Books.

Polya, G. (1945). How to solve it. Princeton: University Press.

Resnick, L. B. (1987). Learning in school and out. Educational Researcher, 16(9), 13–20.

Rochex, J.-Y. (2006). Social, methodological and theoretical issues regarding assessment: lessons from a secondary analysis of PISA 2000 literary tests. Review of Research in Education, 30, 163–212.

Romagnoli, S., & Selleri, P. (2011). La competenza in cerca d'autore [Competence in search for author]. RicercAzione, 2, 231–236.

Sachs, H., Schegloff, A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696–735.

Säljö, R., & Wyndhamn, J. (1987). The formal setting as context for cognitive activities. An empirical study of arithmetic operations under conflicting premises for communication. European Journal of Psychology of Education, 2(2), 233–245.

Säljö, R., & Wyndhamn, J. (1993). Solving everyday problems in the formal setting: an empirical study of the school as context for thought. In S. Chaiklin & J. Lave (Eds.), Understanding practice: perspectives on activity and context (pp. 327–342). Cambridge: Cambridge University Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning. A project of the National Council of Teachers of Mathematics (pp. 355–358). New York: Macmillan Publishing Company.

Schubauer-Leoni, M.-L. (1986). Le contrat didactique: un cadre interprétatif pour comprendre les savoirs manifestés par les élèves en mathématiques [The didactic contract: an interpretative framework for the understanding of the mathematical knowledge displayed by pupils]. European Journal of Psychology of Education, 1(2), 139–153.

Selleri, P. (2016). La comunicazione in classe [Communication in classrooms]. Roma: Carocci.

Selleri, P., & Carugati, F. (1999). "Ecoutez-moi les enfants". De la conversation à l'étude des routines scolaires ["Listen to me, kids!". From the conversation to the study of school routines]. In M. Gilly, J.-P. Roux, & A. Trognon (Eds.), Apprendre dans l'interaction: analyse des médiations sémiotiques [Learning through interaction: analysis of semiotic mediations] (pp. 279–300). Nancy: Presses Universitaires.

Selleri, P., & Carugati, F. (2013). Taking care of children and pupils: agreements and disagreements in parents' and teachers' social representations. In G. Marsico, K. Komatsu, & A. Iannaccone (Eds.), Crossing boundaries: intercontestual dynamics between family and school (pp. 229–269). Charlotte: IAP: Information Age Publishing Inc.

Sinclair, J., & Coultard, M. (1975). Toward an analysis of discourse: the English used by teachers and pupils. Oxford: Oxford University Press.

Verschaffel, L., Greer, B., & de Corte, E. (2000). Making sense of word problems. Educational Studies in Mathematics, 42(2), 211–213.

Yildirim, H. H., & Berberoğlu, G. (2009). Judgmental and statistical DIF analyses of the PISA 2003 mathematics literacy items. International Journal of Testing, 9(2), 108–121.

Patrizia Selleri. Department of Psychology, Alma Mater Studiorum University of Bologna, Via Berti Pichat, 5, 40100 Bologna, Italy. E-mail: patrizia.selleri@unibo.it

Current themes of research:

Social interaction and cognitive development. Classroom discourse. Social representations of development, education, learning and teaching. Dynamics of diffusion of scientific knowledge and common sense. Quality of life.

Most relevant publications in the field of Psychology of Education:

Selleri, P. (2016). La comunicazione in classe. Roma: Carocci (new edition)

Carugati, F. & Selleri, P. (2014). Social development and the development of social representations: two sides of the same coin? In A. Antonietti, E. Confalonieri & A. Marchetti (Eds.) Metarepresentation and narrative in educational settings: where cognitive and social development meet (pp.15–54). New York: Cambridge University Press.

Carugati, F. & Selleri, P. (2014) Piaget, Vygotskij and European approach in social psychology of education: a space for virtous dialogue? In T. Zittoun & A. Iannaccone (Eds.) Activities of thinking in social spaces (pp. 43–61). New York: Nova Science Publishers, Inc.

Selleri, P. & Carugati, F. (2013). Taking care of children and pupils: agreements and disagreements in parents' and teachers' social representations. In G. Marsico, K. Komatsu, A. Iannaccone (Eds.) Crossing boundaries: intercontestual dynamics between family and school (pp. 229–260). Charlotte, N.C., Publisher: IAP: Information Age Publishing Inc.

Scalone, L., Tomasetto, C., Matteucci, M.C., Selleri, P., Broccoli, S., Pacelli, B., Cavrini G. (2011). Assessing quality of life in children and adolescents: development and validation of the Italian version of the EQ-5D-Y. Italian Journal of Public Health, 8, 331–341.

Felice Carugati. Alma Mater Studiorum University of Bologna (retired). Bologna, Italy. E-mail: felice.carugati@unibo.it

Current themes of research:

Socialization processes in institutions and the families. Social and personal identity. Social interaction and cognitive development. Social representations of development, education, learning and teaching. Dynamics of diffusion of scientific knowledge and common sense. Social dynamics in assessment and evaluation of teaching, learning and social institutions. Dissemination of information and communication technologies in educational contexts, distance education in higher education.

Most relevant publications in the field of Psychology of Education:

Carugati, F. & Perret-Clermont, A-N. (2015). Learning and instruction: social-cognitive perspectives. In J. D. Wright (editor-in-chief), International encyclopedia of the social & behavioral sciences, 2nd edition, Vol 13. (pp. 670–676) .Oxford: Elsevier.

Carugati, F. & Selleri, P. (2014). Social development and the development of social representations: two sides of the same coin? In A. Antonietti, E. Confalonieri & A. Marchetti (Eds.) Metarepresentation and narrative in educational settings: where cognitive and social development meet (pp.15–54). New York: Cambridge University Press.

Carugati, F. & Selleri, P. (2014) Piaget, Vygotskij and European approach in social psychology of education: a space for virtous dialogue? In T. Zittoun & A. Iannaccone (Eds.) Activities of thinking in social spaces (pp. 43–61). New York: Nova Science Publishers, Inc.

Miguel, I., Pires Valentim, J., & Carugati, F. (2013). Social representations of the development of intelligence, parental values and parenting styles: a theoretical model for analysis. European Journal of Psychology of Education. https://doi.org/10.1007/s10212-012-0160-3.

Carugati, F. (2013). Commentary: Crossing boundaries between everyday and research contexts: extending the conversation. In G. Marsico, K. Komatsu, & A. Iannaccone (Eds.). Crossing boundaries: intercontestual dynamics between family and school (pp.171–183). Charlotte, N.C. Publisher: IAP: Information Age Publishing Inc., Charlotte, N.C.