Assessing covariation as a form of conceptual understanding through comparative judgement

Sara Bagossi[1] · Federica Ferretti[2] · Ferdinando Arzarello[3]

Abstract

This paper focuses on the importance of covariational reasoning within the processes of mathematics teaching and learning. Despite the internationally recognized relevance of covariation, research shows that only a small percentage of students and teachers is able to adopt covariational reasoning and the majority of mathematics curricula do not contain explicit references to covariational skills. In particular, when covariational reasoning manifests as conceptual knowledge, it becomes challenging to assess, and the need for innovative methods of assessment emerges; there is a need for suitable assessment to highlight the characteristics of covariation and capture the various features that characterize conceptual understanding. Comparative judgement (CJ) is an innovative assessment method based on collective expert judgements of students' work rather than requiring scoring rubrics. Due to its holistic approach, CJ is particularly suitable for assessing complex mathematical competencies, and, as we shall see in this study, it proved to be appropriate for the covariation's assessment. In details, our study aims to investigate the perception and relevance attributed by mathematics teachers to covariation as a theoretical construct and the way CJ can help in the assessment of covariational reasoning skills underlying a less structured modelling task.

1 Rationale

Most mathematics curricula worldwide, including Italian syllabi, highlight the relevance of modelling in students' activities in order to develop up-to-date mathematical literacy. Among the areas of mathematical content in which mathematical literacy is applied, and which are particularly linked to modelling activities, PISA names

✉ Federica Ferretti
  federica.ferretti@unife.it

1 Ben-Gurion University of the Negev, Beersheba, Israel

2 University of Ferrara, Ferrara, Italy

3 University of Turin, Turin, Italy

change and relationships. It claims: "Being more literate about change and relationships involves understanding fundamental types of change and recognizing when they occur in order to use suitable mathematical models to describe and predict change" (OECD-PISA, 2022). One of the crucial aspects in such processes is the ability to develop covariational forms of reasoning, which can help to model "the change and the relationships with appropriate functions and equations, as well as creating, interpreting and translating among symbolic and graphical representations of relationships" (OECD-PISA, 2022).

However, notwithstanding the didactical relevance of covariational reasoning, there are some obstacles to adopting it as a regular and effective practice in Italian classrooms. Firstly, covariational reasoning is not explicitly addressed in the Italian National Curricula; this absence is reflected in school practices and the majority of mathematics textbooks, so most Italian teachers are not aware of covariation and therefore do not foster its use in their classrooms. Secondly, covariational reasoning, insofar as it represents a form of conceptual understanding, is difficult to assess (Bisson et al., 2016). These findings, illustrating the absence of guidelines addressing covariational reasoning in classroom mathematical modelling activities and on its assessment, led us to engage in this research aimed at sowing some seeds to counteract this situation.

The object of this study is the assessment of some open-ended written tasks concerning mathematical modelling of a real phenomenon. In the proposed situations, a conceptual understanding of covariation and the use of suitable reasoning skills can be of great aid in exploring and describing the interconnections between the mathematical model and the real phenomenon under investigation. We identified in comparative judgement (CJ) a valuable tool for assessing students' conceptual understanding involving covariational skills, and making its assessment easier for teachers less confident with covariation. Open-ended tests, the object of this study, were administered to students of a 10th grade class at the end of a teaching experiment conducted in 2019. Students were asked to prepare a written report on an activity regarding the well-known Galilei experiment (Galilei, 1638) of a ball rolling along an inclined plane to represent the law of motion in action.

The specific goals of our research are twofold: firstly, they investigate how CJ can help in assessment of an open-ended test concerning a modelling task in which covariational reasoning skills can significantly help its understanding. This assessment method, based on a holistic approach, is particularly suitable for assessing complex mathematical skills, and in this study, we will explore the potential of assessment through the CJ of open-ended tasks involving covariational reasoning. Secondly, they verify if teachers can (implicitly or explicitly) recognize features of covariational reasoning as important in grading students' works; as a by-product, we can also gain some insight into Italian teachers' reactions to the CJ method. Thus, in addition to performing a CJ session that adopts an online tool, a post-judging questionnaire was designed and distributed to investigate teachers' comparative processes and the features which influenced them.

The paper is organized as follows. After a review of the literature (Sect. 2), we present the main features of the covariation construct and highlight existing research gaps concerning its assessment (Sect. 2.1). The method of CJ is introduced in Sect. 2.4 and all details about the methodology and design of our research are provided in Sect. 3. Data obtained are reported in Sect. 4 and the contribution of this study is discussed in Sect. 5 to provide answers to our research question. Finally, the limitations of this research are outlined, together with further research purposes.

2 Theoretical background

Covariational reasoning is a mental activity involved in solving specific cognitively demanding tasks, which require the development of "multiple levels of sophistication" (Thompson & Carlson, 2017, p. 436). Indeed, it is an example of what many researchers call "conceptual knowledge" (Skemp, 1976), insofar as it involves not only a deep understanding of the principles and theories that govern a certain domain of knowledge (Rittle-Johnson et al., 2001), but also the relationships among the various sources of information (Hiebert & Lefevre, 1986); above all, it does not reduce this reasoning to the mere application of mechanical procedures. Covariational reasoning is proved to be essential in activities concerning mathematical modelling (Thompson, 2011), specifically when involving motion or dynamic situations.

One important didactical problem related to students' achievement of covariational reasoning in modelling activities lies in its concrete assessment and this issue involves three main difficulties.

First, covariational reasoning deals with a typical conceptual knowledge construct, which is more complex than a content area or procedure to be assessed because of the variety and complexity of students' reasoning. Bisson et al. (2016) highlight the difficulties of using standard assessment practices when conceptual understanding is under the lens because "conceptual understanding is an important but nebulous construct which experts can recognise examples of, but which is difficult to specify comprehensively and accurately in scoring rubrics" (p. 143).

Second, this opaque situation for assessing covariational reasoning manifests also with regard to activities using mathematical modelling. Studies on this topic have mainly focused on the theoretical definition of mathematical models (Niss, 1989; Niss et al., 2007) or on a valid definition of mathematical modelling competence (Niss & Højgaard, 2019), but there is not usually a systematic exploration of the issue of its assessment which should reflect not only the aims of applications and appropriate modelling (Blum, 2015) but also students' ability to reason covariationally.

The third issue concerns some institutional and didactical aspects. As pointed out by Thompson et al. (2017), covariational reasoning is not a regular feature in mathematics curricula, except for a few Eastern countries, and therefore is not explicitly considered in students' assessment surveys. In addition, teachers themselves struggle when teaching covariation and one of the reasons is that they are neither able to master covariational reasoning (Thompson et al., 2017) nor to include it in their school practices.

2.1 Covariation as a theoretical construct

The large body of literature concerning covariation as a theoretical construct began in the early 1990s (Confrey, 1991; Thompson & Thompson, 1992) and is rooted in quantitative reasoning (Thompson, 1993). Saldanha and Thompson (1998) defined covariational reasoning as the ability to hold in mind a global image of the values of two quantities as they vary simultaneously. Thus, it concerns coupling the two quantities and forming a multiplicative object, an idea derived from Piaget's notion of logical multiplication (1950). Covariational reasoning emerges when students are able to grasp "that there is an invariant relationship between their values that has the property that, in the person's conception, every value of one quantity determines exactly one value of the other" (Thompson & Carlson,

2017, p. 436). The final version of this theoretical construct, presented in Thompson and Carlson (2017), consists in a taxonomy of six levels describing the characteristics of a person's skill in covariational reasoning: no coordination, pre-coordination of values, gross coordination of values, coordination of values, chunky continuous covariation, and smooth continuous covariation.

International studies in mathematics education have strongly underlined and supported (with evidence-based research) the importance of covariational reasoning for the understanding not only of many mathematical ideas (e.g., function, rate of change, variable, exponential growth) (Thompson & Carlson, 2017). Covariation is revealed as crucial for the conceptualization of dynamic situations (Carlson, 1998; Carlson et al., 2002) and the mathematical modelling of a physical phenomenon (Arzarello, 2019) because "the operations that compose covariational reasoning are the very operations that enable one to see invariant relationships among quantities in dynamic situations" (Thompson, 2011, p. 46), and so it is fundamental for launching into the steps of the so-called modelling cycle (Blum, 2015; Blum & Niss, 1991).

2.2 Relevance of covariation in the mathematics curricula

Despite the recognized importance of covariation, school practices and mathematics curricula also rarely focus on covariational reasoning in the USA, where covariation is an extremely relevant research topic (Thompson et al., 2017). Explicit attention to covariation can be found instead in Japan, South Korea, and Russia, and literature shows that also Germany has a long tradition supporting the covariational approach in functional thinking (Thompson et al., 2017; Vollrath, 1989). The Italian curricula for secondary schools instead (both for second cycle MIUR, 2010a, b, c and for first cycle MIUR, 2012) do not contain explicit references to covariation. Concerning the curricula for science-oriented secondary school, an implicit reference can be found in the following statement, which could be interpreted according to a covariational reasoning perspective: "An important topic of study will be the concept of rate of change of a process represented through a function[1]" (MIUR, 2010a, p. 340). From the early years of high school, functions are studied, especially as representations of real phenomena, and the Italian curricula stress the relevance of mathematical modelling activities in teaching practices. The words *model* and *modelling* appear nearly 12 times, under the headings *General guidelines*, *Change and relationships*, and *Uncertainty and data*. In particular, it states that modeling consists in "the possibility to represent the same class of phenomena through different approaches" (MIUR, 2010a, p. 337) and that the study of the language of functions is the "first step in introduction to the concept of a mathematical model" (p. 339). This lack of explicit references is also reflected in the fact that, apart from a minority of textbooks which in recent years have introduced covariational reasoning as a thinking tool (Paola & Impedovo, 2014), most Italian textbooks currently in use do not foster this approach when dealing with the modelling of classes of phenomena or conceptualization of dynamic problems. For instance, when teachers introduce the concept of function, they mainly adopt a static definition, e.g., that of Bourbaki (1939) as a law of correspondence between the elements of two sets. As literature suggests, the lack of a covariational approach may be one reason why students are

---

[1] The translation of the quotations here reported is made by the authors.

unable to interpret dynamic situations and to construct meaningful formulas suitable for representing one quantity as a function of another (Carlson, 1998).

2.3 Assessment of covariation

Given the peculiarity and complexity of covariation as a theoretical and cognitive construct, the variety of fields and topics where covariation can be applied, its occasional and only implicit presence in school curricula and practices, and, above all, the limited knowledge of teachers in this regard, any assessment of this form of reasoning turns out to be challenging, and a non-homogeneity can be detected both in terms of achievements and assessment practices. We here provide some examples of existing ways in which covariational reasoning has been assessed, and we use those findings to help identify the benefits of a comparative judgement method.

The six-level taxonomy developed by Thompson and Carlson (2017), constructed through individual interviews, aims at classifying the different ways in which one person can reason in a covariational way. Their framework levels should be interpreted as "descriptors of a class of behaviors" or as "characteristics of a person's capacity to reason covariationally" (Thompson & Carlson, 2017, p. 435) and there are several qualitative studies that have used these descriptions of mental actions as a coding scheme (e.g., Carlson et al., 2002). Another significant contribution is the diagnostic assessment of teachers' mathematical meanings called Mathematical Meanings for Teaching secondary mathematics (MMTsm; Thompson, 2015). This assessment has a very detailed rubric to score students' quantitative and covariational reasoning and has been used with both US and Korean teachers (Yoon & Thompson, 2020). In Thompson et al. (2017) on the other hand, the investigation of US teachers' attitudes to reason in a covariational manner was conducted through the development of a specific scoring rubric based on the features of graphs that teachers were asked to sketch. As Thompson et al. (2017) underline, the scoring categories were not defined theoretically because scorers would not have had the theoretical background to understand the rubric.

What seems to be missing, and proves even more challenging, is the assessment of covariation intended as a form of conceptual understanding. Covariational reasoning, intended as a form of conceptual knowledge "is defined by how it is perceived, understood and used by the relevant community of expert practitioners" (Bisson et al., 2016, p. 143), and in a strict rubric of items it risks reducing to a "rigid definition that fails to capture the full meaning and usage that exists in practice" (p. 143). Moreover, the limited pedagogical and mathematical knowledge of teachers about this theoretical construct makes it hard to elaborate suitable items of a scoring rubric whose meaning is shared by the expert mathematicians' community. Finally, "conceptual understanding is best assessed using open-ended and relatively unstructured tasks" (Bisson et al., 2016, p. 143) but assessing open-ended tests is difficult because of the variety and unpredictability of students' responses, and also the huge amount of time required to obtain both a valid and reliable outcome.

In recent years, an alternative approach to the assessment and enhancement of conceptual understanding has been proposed (Jones & Karadeniz, 2016) based on the technique of comparative judgement. In this study, instead of developing a specific content-based scoring rubric, we use the holistic method of CJ to assess some students' written productions involving covariational reasoning. Although CJ is usually adopted as a time-saving technique to assess a huge number of tests, we used it to assess just a small sample of tests and our main research question can be formulated as follows:

How can CJ, performed by expert mathematicians, be valuable in the assessment of open-ended tasks requiring covariational reasoning?

2.4 Comparative judgement: an assessment method

One explanatory factor of the difficulties that accompany assessment practices, in Italy as in other contexts, is linked to changes in the way of understanding the teaching and learning of mathematics nowadays. Teaching does not consist merely in the transmission of disciplinary concepts but must also enable students to autonomously build "significant learning" (in the sense of Ausubel, 1960). Learning becomes meaningful when it does not arise from the accumulation of notions and information, but when the learner becomes able to use these in order to tackle complex problems, identifying paths and tools that help him/her to act effectively and competently. Summing up, school education should enhance conceptual understanding rather than mere procedural knowledge, but all the issues concerning assessment are amplified when it comes to assessing complex forms of conceptual knowledge such as covariational reasoning. All of this has obviously led to the need to search for new methods and tools in the assessment field. It is within this perspective that the use of CJ in the educational field was born. The comparative judgment is an assessment method based on the idea that human beings are more reliable at making judgements between objects, through comparison, than at making value-laden judgements about the quality of an object. This non-traditional assessment method is based on expert judgment to whom are presented successive pairs of students' work and asked to decide, for each pair, which student has displayed the greatest proficiency (Jones et al., 2015). As detailed later, ties are not permitted, and the judge must choose one student's work in preference to the others. Comparative judgment offers an alternative to traditional scoring for the assessment of students' works; it is an approach that bases its validity directly on what is valued and expected by judges, rather than on what can be precisely established in scoring rubrics. CJ has been applied successfully in various educational assessment contexts (Tarricone & Newhouse, 2016) and, as we will argue in this study, it may be helpful also in the assessment of unstructured tasks involving forms of conceptual understanding as covariational reasoning.

Even though CJ may be innovative in the Italian context, internationally it has received increased attention in the last two decades. It constitutes an efficient alternative approach to traditional marking (Pollitt, 2012), works particularly well in absence of restrictive indications or scoring rubrics, and allows the assessment of forms of complex knowledge such as covariation, with a holistic approach (Jones et al., 2019). It is rooted in the psychological principle that people are more accurate in making a comparison between two objects than when expressing an isolated judgement (Laming, 1984; Thurstone, 1927). Some experts are asked to make pairwise comparisons of students' works, choosing the one they consider better according to a global construct. Results are fitted to the Bradley-Terry statistical model (1952) which returns a unique score for each student, i.e., a scaled rank order of the works. Research literature supports both the validity and reliability of comparative judgement. It is proven to be even more reliable than traditional marking in open-ended assessment not only in mathematics (Jones & Alcock, 2014; Steedle & Ferrara, 2016). This assessment technique has been used with several mathematical topics: problem-solving (Jones & Inglis, 2015), conceptual understanding (Bisson et al., 2016; Jones & Alcock, 2014; Jones & Karadeniz, 2016), mathematical proof (Davies et al., 2020), and also statistical knowledge (Bisson et al., 2016; Marshall et al., 2020). As far as we know, CJ has

never been used to assess tests requiring covariational reasoning. Typically, CJ is used to assess tests on a national scale in countries like the UK or New Zealand but is little-known in Italy. Its advantages reside in the non-problematic nature of the variety of possible students' responses, the non-necessity of a scoring rubric to grade the productions, and the resulting reduction in time required to assess a huge number of tests.

Although validation can never be fully demonstrated or completed (Marshall et al., 2020), suitable statistical methods allow us to verify the reliability and validity of CJ outcomes. All the details useful for the purposes of our research are provided in Sect. 4.1.

3 Methodology

3.1 The context of the CJ experimentation

The written productions analyzed in this research paper are the result of a larger teaching experiment (Steffe & Thompson, 2000) conducted in 2019 in a 10th grade class of a science-oriented school in the province of Turin (Italy). The experimentation took 3 weeks for a total duration of almost 16 h. The class was made up of 22 students who worked in small groups of $4-5$ students. All the group work sessions were followed by classroom discussion guided by the teacher. The main purpose of the activity was to mathematically explore the phenomenon of the motion of a ball running along an inclined plane, the so-called Galilei experiment. This goal was achieved through interaction with various artifacts providing different representations of the main aspects of the phenomenon analyzed: specifically, a video reproducing the experiment and some applets within the GeoGebra environment simulating the phenomenon and providing a table of finite differences. The final aim was to achieve a strong grasp of the law of motion of the ball, $s = k \cdot t^2$ where s is the distance traversed by the ball, t is the time, and k is a parameter depending on the angle of inclination of the plane. The design of the proposed activities allowed the students (i) to explore the relationship between time and distance thanks to the numerical relationships between time and covered distance provided in the video and the table of first finite differences provided in applet and (ii) to explore how the distance-time graph is affected by a change in the angle of inclination of the plane thanks to a simulation of the inclined plane contained in the GeoGebra applet and the possibility to change its inclination by dragging a blue point located at the end of it and so obtaining the related discrete distance-time graph. The screen of the GeoGebra applet is shown in Fig. 1.

During the working group sessions on the video and GeoGebra applets, students were provided with some worksheets containing open questions and they were asked to explore and elaborate some hypotheses on the law of the motion of the ball. These working group sessions were followed by classroom discussions. We remark that only covariational reasoning enables one to envision time and distance traversed varying simultaneously and continuously and to appreciate how the angle of inclination of the plane globally affects the distance $-$ time relationship. At the end of the project, students were asked to write a text in which they described (to students of another class) the activity completed and the law of the obtained motion, producing a form of theoretical support for the study. The task of the essay is reported below translated into English:

> Thinking back to the work carried out on the inclined plane, write to schoolmates of another class to outline the work itself and, specifically, the relationship that

Fig. 1 The GeoGebra applet interface

describes and explains mathematically the motion of the ball along the inclined plane. This report should be a theoretical support for you and your schoolmates.

The expedients of a narrative addressed to schoolmates from a different class and that of the theoretical support for study have been chosen to invite students to adopt a formalization approach and not to take details for granted. Students were accustomed to this kind of task and were familiar with the terminology used in the prompt.

3.1.1 Participants

The students, who are accustomed to written tasks in mathematics and physics, were involved in this activity without prior notice, had two hours of time to complete the task, and received no specific instruction on the formal structure of the essay. The texts were from two to four pages long; students could freely report graphs, formulas, and numerical tables in their written work.

3.1.2 Judges

The research group recruited 13 judges from pre-existing contacts. All judges were mathematicians and experts in the field: specifically, 2 of them were middle school teachers, 7 were high school teachers of mathematics and physics, and 4 were university professors specialized in mathematical physics. The group of judges was heterogeneous, to collect different points of view on this evaluation experience. Based on informal interviews, we can state that none of the judges had ever heard about the CJ assessment technique. The mathematical competence of the judges is necessary to ensure reliability of the outcomes (Bisson et al., 2016; Jones & Alcock, 2014), particularly as they were asked to identify the better text from a mathematical point of view, rather than that which displayed more covariational reasoning. All the judges volunteered for the case study and received no financial compensation. The students' teacher was not part of the group of the judges, and she elaborated a personal evaluation of students' texts as addressed in Sect. 4.1.

3.2 The research design

3.2.1 Phase prior to comparative judgement

A few weeks before starting the CJ session, judges were introduced to comparative judgement through e-mail correspondence which explained the peculiarity of this assessment approach and the main differences from traditional marking. On the first day of the project, the judges received an instruction sheet containing all the information related to the teaching experiment on the law of the inclined plane, how the activity was structured, its main objectives, and the task given to the students (see Sect. 3.1). Moreover, the instruction sheet showed specifically how the comparisons platform worked. The judges could keep the instruction sheet on hand throughout the whole CJ session. The judges did not undergo any trial CJ session because one of the research purposes was to capture their first impressions when facing this new kind of assessment approach.

3.2.2 Comparative judgement procedure

The 22 open-ended texts were anonymized for a matter of privacy, scanned, and uploaded to an online comparative judgement platform (www.nomoremarking.com), freely available for research purposes. When starting comparisons, judges saw a screen recalling the main aims of the Galilei teaching experiment and were asked to "choose the best mathematical text." Judges made 17 judgements each, for a total amount of 221 judgements, so that each text was compared at least 10 times according to literature standards (Jones & Alcock, 2014). The judges worked in their free time and had 20 days of time to complete their comparisons.

The CJ website displayed the two texts to be compared side by side and the judges just had to click the left or right button to express their preference. An example of the display screen is shown in Fig. 2.
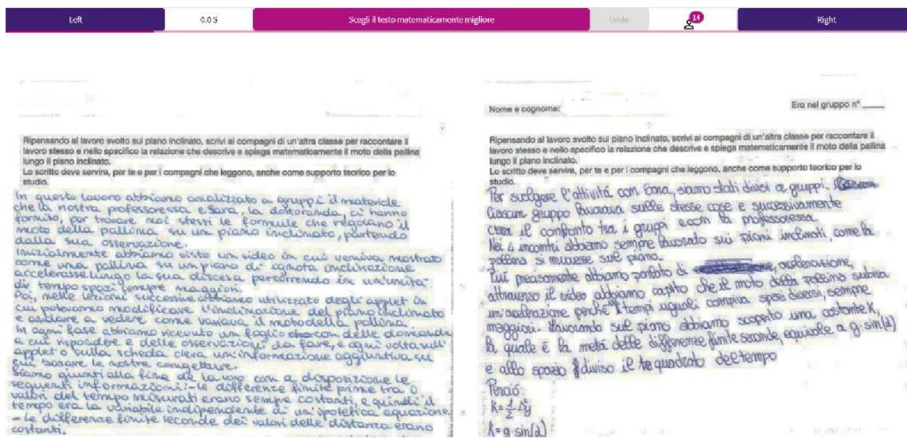
Fig. 2 Screen for the comparisons displayed on the online engine

3.2.3 Post-judging questionnaire

One week after completing the CJ procedure, the judges received by email new instructions to fill in an online questionnaire. The judges were sent two texts chosen by the researchers and based on the following criteria: the two texts had obtained a good and close final CJ score, but they also presented many dissimilarities. They were written by students of different gender, a boy and a girl, had different lengths (text 1: 2 pages; text 2: 3 pages) and one was clearly written but presented misspelling, while the other was untidier. A picture of the first page of two texts is contained in the (Appendix Figures 6,7, 8 and 9) along with a translation into English of the content.

Both texts first describe what they had seen on the video, then recall the GeoGebra applet and what they had done with it, and eventually summarize the meaning of the experiment using different representations.

Text 1 is more focused on verbal descriptions than formulas. The student's verbal descriptions provided evidence of their covariational reasoning; we detect the use of verbal markers like "always constant" and distinguishing dependent and independent variables. The student introduces a table, which condenses the covariation between t, s, first finite differences of s, and formulas that make explicit their mutual dependence. Text 1 privileges conciseness and tries to condense the message in not too many chunks, which in any case contain all the essence of covariation.

Text 2 instead uses more formulas and representations and gives more details about the thinking processes through which the covariational relationships between the different quantities have been achieved, also referring to the support given by the teacher for progressing into it. The student's use of general formulas, skipping references to explicit numerical values, provided evidence of their covariational reasoning: it is underlined mixing words, formulas, pictures, and using arrows to mark the mutual relationships among quantities. In this, text 2 is much richer than text 1 since it gives more details. Even if adopting different modalities and communicational registers, both the authors show a good mastery of covariational reasoning.

The first part of the questionnaire aimed at investigating which factors influenced the outcome of the CJ procedure. After asking judges to state which of the two texts they considered better in mathematical terms, we elaborated a four-point set of items representing an ordinal scale from 1 (little influence) to 4 (strong influence) and the judges were presented with some basic features of students' work. The 10 features listed below were of three main categories: *communicational aspects*, like the presentation clearness, the structure of the text, and the ability to synthesize $[3-4-6]$; *mathematical aspects*, like the use of formal notations (e.g. graphics, images), of a formal mathematical vocabulary, the modelling capability, and the presence of errors $[1-2-5-7-8]$; and *specific covariation features*, like the ability to describe verbally or symbolically the relationships between distance, time, and the angle (of inclination of the plane) $[9-10]$.

1. Presence of errors
2. Use of formal notations
3. Untidy presentation
4. Structure of the presentation

5. Use of graphics and images
6. Ability to synthesize
7. Use of a formal mathematical vocabulary
8. Modelling capability
9. Ability to describe *exhaustively* the relationships between distance, time, and the angle (of inclination of the plane)
10. Ability to describe *formally* the relationships between distance, time, and the angle

This list of features and also the following questions were mainly inspired by the work on judgement processes by Jones and Inglis (2015): we adapted them according to the characteristics displayed by our written texts and added some features specifically related to the content of the task proposed. The feature "modelling capability," explicitly refers to the modelling competence stressed in national curricula and intended as the ability to represent classes of real phenomena. Features 9 and 10 instead were designed to grasp specifically the relevance of covariation. The tasks proposed during the experimentation (see Sect. 4.1) allowed us not only to investigate the distance − time relationship, but also to explore in depth how the angle affects the distance − time graph (Bagossi, 2021). The specific covariation features refer to the ability to describe that relationship using the natural language (exhaustively), or to condense it into suitable formulas (formally). The two adverbs respectively refer to the verbal and symbolic semiotic registers according to Duval's frame (Duval, 2017).

The questionnaire concluded with three open-ended questions: (1) Please list any other features you think may have influenced your judgement when comparing texts. (2) Please comment on this overall experience and state your feelings during the comparative process. (3) In his theory of quantitative reasoning, P. Thompson states that a person reasons in a covariational manner when able to envision the values of two or more quantities as varying simultaneously. In which of the two texts do you think the presence of covariation between physical magnitudes can be better encaptured and why?

The judges again worked in their free time and had two weeks to complete their task.

4 Data analysis and results

4.1 Outcome of the CJ procedure

The CJ method allows positioning a set of complex objects on a unidimensional scale (Davies et al., 2020). The CJ website fitted the 221 judgements with the statistical Bradley-Terry model (1952) and using a maximum likelihood estimation procedure (Pollitt, 2012). It produced a final parameter estimate for each student ($M = 0$, $SD = 1.7$). The distribution of the class' scaled scores is shown in Fig. 3; the box shows the two texts involved in the post-judging questionnaire, whose values were 0.57 (text 1) and 0.47 (text 2).
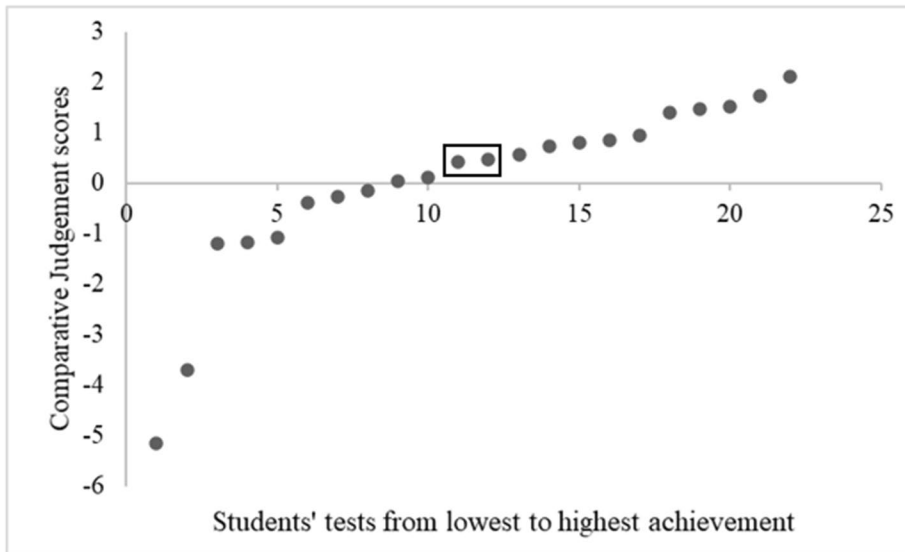
Fig. 3 Distribution of CJ scores as per number of students reported on the vertical axis ($N=22$). The box frames the two texts involved in the post-judging questionnaire

4.1.1 Reliability

The reliability of an assessment procedure refers to the consistency of its outcomes. Internal consistency can be measured with the Scale Separation Reliability (SSR), analogous to Cronbach's alpha. The outcome showed high internal consistency since $SSR \geq 0.7$ ($SSR = 0.81$).

Moreover, looking for "misfitting" judges, we computed an infit statistic for every judge and compared this data to a threshold value, i.e., two standard deviations above the mean of the infit (Marshall et al., 2020; Pollitt, 2012). Only one judge resulted slightly above the threshold value (0.03 above); so we did not consider it necessary to adapt the CJ scores by removing the misfitting judge.

4.1.2 Criterion validity

In order to evaluate the validity of the outcome of the CJ procedure, we considered the *criterion validity* by computing the Pearson Product−Moment correlations between the parameter estimate of the CJ procedure and some benchmark measures. We would expect positive correlations, meaning that those students who are more successful on these other covariation-related performances were also more successful in the assigned task.

We correlated the CJ outcome with the marks assigned by their teacher to a physics test concerning problems on accelerated motion and the motion of bodies along the inclined plane ($r = 0.57$) (Fig. 4). The test was specifically designed by their teacher to enhance covariational reasoning. We also correlated the parameter estimate with the marks assigned by their mathematics teacher, who has a solid background in covariation, to her own students' written productions specifically focusing on covariational skills in the assessment phase ($r = 0.51$). These results, while modest, are in alignment with the results of other research in literature on conceptual understanding in secondary and tertiary mathematics
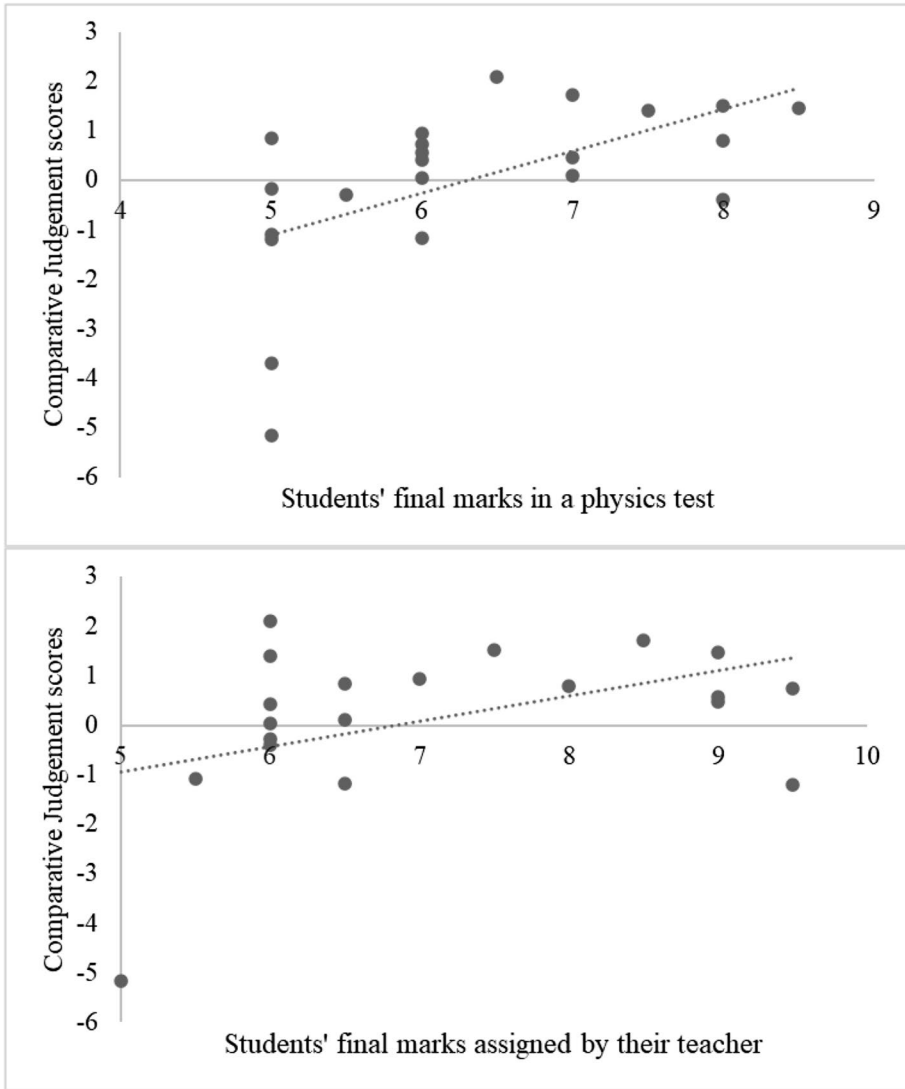
Fig. 4 Scatter plots of the relationship between CJ scores, those on a physics test on the same topic, and those assigned by their teacher on the same test

(Bisson et al., 2016), which reported correlation coefficients between 0.35 and 0.56. Finally, we correlated CJ scores with the final course scores (i.e., the evaluation obtained at the end of school year) in mathematics ($r = 0.48$): this positive correlation can be interpreted as a sign that covariational reasoning is a transversal competence in mathematics.

4.2 How much does covariation matter?

The last open-ended question (3) of the questionnaire asked the judges to state which of the two texts displayed a stronger covariational reasoning and why. Seven judges preferred text

Fig. 5 Table of finite differences present in text 1

2 and 4 preferred text 1 but 2 judges expressed a preference without justifying it. Moreover, one explicitly stated that he didn't know [J12], and another said neither of the two because he considered "the two texts very confusing" [J4]. Globally, 10 out of 13 answers were in agreement over which text was mathematically better.

Judges who preferred text 1 justified their choice by saying:

[J3] It clearly expresses how "the ratio between space and time is constant and this ratio is equal to the half of acceleration"
[J10] It "highlights a greater understanding of the subject in its overall aspect"
[J13] "It addresses finite differences," referring to a table made by the student (Fig. 5) in which he reported distances and first finite differences of distance in function of time and the parameter k, i.e., providing a general expression and not numerical values.

Instead, those judges who expressed a preference for text 2 provided the following reasons:

[J1] "In the supported conclusion, the covariation between the two magnitudes involved is perfectly grasped and the formula simply assumes the role of symbolic expression of that covariation, so that it is not even mentioned anymore." In fact, the student concluded with a sentence summarizing her considerations: "The distances traversed by the ball are directly proportional to the squares of times"
[J2] "In several steps the relationships of dependence of the mutual growth are underlined"
[J6] Another answer, more difficult to interpret, attributed her decision to the "more schematic structure of the discourse" of the study.
[J7] "Not only did they report formulas, but they also tried to explain variations. They also referred to dependent and independent variables"
[J8] "It expresses which magnitudes depend on the others"
[J11] "The conclusive section summed up the different hypotheses and discussed each one in depth, eventually arriving at formulas in correct terms of the law of 'falling' bodies on an inclined plane (s as a function of time t) free of friction."

The conceptual aspects of covariation are highlighted, considering the role that some judges award to formulas in the statements above. J1 who chose text 2 states that "the formula just assumes the role of symbolic expression of that covariation" [J1] and is "condensed" into the sentence "the distances traversed by the ball are directly proportional to the squares of times." So, one crucial point for the judge is that formulas address conceptual knowledge, while formulas alone are not always enough; this is stressed also in another statement: "Not only did they report formulas, but they also tried to explain variations"

[J7]. In another stream of thought, J13 who chose text 1 considers formulas positive when judging, if they allow the formulation of "general expressions" and not simply numerical ones, relying on a conceptual representation such as that of finite differences. In this case, students use their existing knowledge (the finite differences method) to elaborate methods for solving the problem. In both cases, covariation is thought of as an "implicit or explicit understanding of the principles that govern a domain and of the interrelations between units of knowledge in a domain" (Rittle-Johnson et al., 2001, p. 347), namely of a key feature of conceptual knowledge.

With reference to the basic features listed in Sect. 3.2, we did not get significant different results considering the whole group of judges. They become more relevant considering the subgroup of 4 judges who chose text 1 and the subgroup of 7 judges who chose text 2. The first subgroup attributed more importance to the presence of errors and the modelling capability; subgroup 2 valued more the structure of the presentation, its tidiness and the use of graphics and images. Globally we can observe that those who chose text 1 gave more importance to non-strictly mathematical features and those who chose text 2 gave more importance to mathematical features. Both subgroups considered positively relevant those features addressing specifically the relationship between the quantities involved. All the means and standard deviations are reported in Table 1. Hence, the answers of the judges confirm that their choices had been deeply influenced by the communicational registers adopted, and particularly by the more or less relevant role of representations in expressing covariation. The communicational and representational registers used by the students were

Table 1 Table containing the features of the survey for two subgroups of judges, those who chose text and text 2, divided into three sections: non-mathematical features (white), mathematical features (grey), covariational features (dark grey)

Quantitative questions (1=little influence, 4=strong influence)	Text 1	Text 2
Untidy presentation	M=2 SD=0.71	M=2.71 SD=0.70
Structure of the presentation	M=1.75 SD=0.43	M=3.14 SD=0.83
Ability to synthesize	M=2.75 SD=1.09	M=2.85 SD=0.64
Presence of errors	M=3 SD=0.71	M=1.86 SD=0.83
Use of graphics and images	M=1.75 SD=0.83	M=2.86 SD=0.64
Use of formal notations	M=2.75 SD=0.83	M=2.57 SD=0.49
Use of a formal mathematical vocabulary	M=3 SD=0.71	M=2.86 SD=0.64
Modelling capability	M=3.25 SD=0.83	M=2.85 SD=0.64
Ability to describe exhaustively the relationships between distance, time, and the angle	M=2.75 SD=0.83	M=2.85 SD=0.64
Ability to describe formally the relationships between distance, time, and the angle	M=3 SD=1	M=2.85 SD=0.64

explicitly present in almost all judges' sentences. Moreover, there was a high frequency of the specific covariation features [9, 10] in the answers: both were quoted as relevant by all the judges. Eleven judges attributed equal importance to both features; in the case of the other two judges, one gave more weight to the discursive aspect and the other to the formal aspect, so confirming that covariation can be expressed within different communicational registers with equal success.

4.3 Opinions on CJ as an assessment technique

The comments of the judges on this experience express different opinions. Five judges stated the experience was "positive" or at least "interesting." It was defined "optimal from the point of view of accurate assessment because it diverts attention from the error of the single student" [J1]; "direct comparison between the two texts enables (from the beginning) different points of view without valuing or penalizing too much a single student" [J2]. Another judge said that "the assessment is quite fast, and the comparison facilitates the judgement" [J13]. The median time of judgement for the judges ranked from 2 to 11 min for a median time of 297 s (nearly 5 min per comparison). These time data are an overestimate because the timer of the online platform does not stop when judges move away from the computer leaving the screen open. The time required for the comparisons in our experimentation does not differ much from the estimate reported in Marshall et al. (2020) and, in our opinion, it is reasonable when taking into consideration the length of texts, their nature, and difficulties linked to deciphering handwriting. However, based on our experience, the time is less than that usually required to grade a single text by a traditional marking method.

A negative aspect underlined even by those judges who positively evaluated the experience laid in "bad handwriting" [J5], "errors of misspelling" [J4], and "lack of linguistic correctness" [J3].

Those judges who considered the experience as "complicated" and "hard" supported their statement by saying that "most of the texts presented both strengths and weaknesses and it was not always evident which text was the best" [J3]. The texts showed "different modalities of reporting, some students used mainly natural language, others formal, but eventually reached the same conclusion" [J8]. The presence of "descriptive elements related to an experience of the students" [J10] made the assessment "challenging" compared with other forms like "tests in the state examination" [J10].

Another judge stated that to "really focus on the mathematical aspects of the content" [J12], texts should be "visually comparable" [J12]. Only two judges clearly stated that they "prefer" or "need an evaluation grid" [J9-J10].

5 Discussion

Data emerging from our analysis seem to positively support the assumption that the CJ technique may be helpful in the assessment of an unstructured test involving covariational reasoning. We provide arguments for this claim. The positive correlation between the CJ scores and the three benchmark measures provided in Sect. 4.1 (finals scores in mathematics, grades given by their teacher, and the tests on accelerated motion) is a sign that (i) judges really focused on mathematical aspects when comparing students' texts; (ii) if there had not been covariational reasoning, students' productions would be

of low mathematical level. In the post-judging survey, 10 of the 13 judges claimed that the text displaying a greater ability to reason in a covariational manner was also the best mathematical text, at least regarding the two texts analyzed during the online survey. The method of CJ, due to its global approach, seems to have provided a reliable assessment of the texts even though judges were not specifically looking for covariation. Indeed, covariational reasoning results to be fundamental to succeed in a mathematical modelling task like the one illustrated here.

Our investigation allowed us to obtain implicitly some insights into our judges' perception of the construct of covariation and the relevance attributed to it. Judges identified the relationship among the magnitudes involved in the real phenomenon as strongly relevant for their judgements. The answers to the open-ended question clearly reveal that a full grasp of this relationship does not necessarily translate into elaboration of a formula condensing all quantities involved, but certainly results in a complete description of the relationships of dependence and mutual growth. One judge interestingly referred to the table of finite differences, a powerful tool but barely used in Italian school practices, which can be considered a forerunner instrument to the study of rate of change and can surely help students in grasping how variables vary and co-vary.

While the validity of CJ as an assessment method in general, and specifically concerning conceptual understanding (Bisson et al., 2016; Jones & Alcock, 2014; Jones & Karadeniz, 2016), is not in question, it is an unknown and rarely used tool in Italy since in our country there is a long tradition of the use of scoring rubrics in the assessment field. Hence, even though it was not a primary purpose of research, we were curious about gaining some insight into how Italian teachers could approach CJ. Although none of them had ever experimented with CJ, no judge expressed any discomfort in using it and only two judges clearly referred to the lack of a scoring rubric. In most cases, the positive factors of this assessment method were highlighted. Most of the judges involved in our research recognized that CJ allows a more holistic approach to assessment of conceptual understanding. Some judges stated that CJ avoids awarding too much importance to single errors that are often irrelevant in mathematical solution processes.

This study also presents many limitations. First of all, the physics test prepared by the teacher was not validated in class and this could affect the feasibility of the correlation. Regarding participants in the study, the sample of students involved is small, especially compared to other studies about CJ (e.g., 250 responses for Jones and Inglis (2015); 200 responses for Steedle and Ferrara (2016)). Moreover, these students were accustomed to this kind of unstructured tasks, which is not so common in Italy, and their teacher has a strong background concerning covariation, introducing this reasoning in her lessons. Although the number of judges involved in this experimentation is in agreement with the standard literature on CJ (Jones & Alcock, 2014), it is still a small number with which to generalize the questionnaire conclusions: they simply provide interesting insights into which elements capture attention when reading an open-ended text, especially concerning a mathematical topic framed in a real context, and offer some preliminary information on the attitude of Italian teachers toward covariation. We hope our study may act as evidence that larger investigation on the assessment of covariation as a form of conceptual understanding is required and of value. In particular, the last three items in the list in Sect. 3.2.3, which concern specifically modelling and covariation, have shown crucial in our analysis: possibly, new researches using CJ along these features could be fruitful for a comparison of results in investigations outside Italy about teachers' perception and awareness of covariation and of its relationships with students' modelling competences. We conclude underlining that we chose to undertake this line

of research adopting comparative judgement since it allows us to tackle the challenges and difficulties we outlined in the literature review, but this does not exclude the fact that other methods may be valid or that more teachers' training on covariational reasoning is needed.

Appendix

Fig. 6 Original version of text 1, page 1

Thinking back to the work carried out on the inclined plane, write to schoolmates of another class to outline the work itself and, specifically, the relationship that describes and explains mathematically the motion of the ball along the inclined plane. This report should be a theoretical support for you and your schoolmates.

In these schooldays, we have carried out a class activity in collaboration with our math and physics teacher, S.B., and a PhD student in Math, S.. This activity divided into four meetings during which we "discovered" a new topic, that is accelleration and relative distance.

In the first meeting, during which, as in all the others, we were video-taped and recorded, we watched a video in which the motion of a ball along an inclined plane was illustrated: the ball travelled different distances in the same period of time, that is a ripple of a pendulum: they were indicated by a sequence of odd consecutive numbers. That was due to acceleration, that is the increasing of the speed of the ball, of its motion along the inclined plane, in a certain amount of time. The acceleration can be found ~~fr~~ applying the following formula: P_x/m. P_x is the component of the weight force parallel to the inclined plane, that can be found doing ~~the inclination angle~~ the sin of the inclination angle of the plane, α, times the total weight force, ~~hence P~~ which decomposes in m [mass] · g [constant]. The divisor instead is the mass of the ball. The final formula will be: $\frac{m \cdot g \cdot sin(\alpha)}{m}$. The two masses are simplified and $g \cdot sin(\alpha)$ remains.

During the study we noticed a formula, that is $\frac{s\,[distance]}{t^2\,[\,squared\;time]}$.

Going on with the study and with the explanation, we discovered that this formula denotes ~~denotes the distance traversed by the ball according to time~~ K//, a value that then we will use in a while. In the activity we worked mainly with GeoGebra, a software that allows to construct a plane with ~~the time~~ a graph and a table related to it. We could make the ball start from the top of the plane and this illustrated the motion.

**Fig. 7** Translation into English of text 1, page 1

Ripensando al lavoro svolto sul piano inclinato, scrivi ai compagni di un'altra classe per raccontare il lavoro stesso e nello specifico la relazione che descrive e spiega matematicamente il moto della pallina lungo il piano inclinato.

Lo scritto deve servire, per te e per i compagni che leggono, anche come supporto teorico per lo studio.

Attraverso il progetto, intrapreso il 29 settembre in cui coinvolger ci sono state una dottoranda, Sara e la nostra professoressa di matematica e fisica, Silvia Beltramino, abbiamo avuto l'opportunità di apprendere molteplici nozioni, sul moto di una pallina su un piano inclinato.

Il primo incontro si è sviluppato nell'arco di un ora... abbiamo inizialmente preso visione di un filmato (durato poco più di 1 min) e successivamente (divisi in gruppi) abbiamo descritto secondo fondamentalmente 2 aspetti. (quello visivo e quello comunicativo)

Ci è stata consegnata una scheda, dove appunto abbiamo raccontato "fisicamente" quello che a aveva visivamente colpito (quindi in pratica il piano inclinato, il pendolo posizionato sul lato opposto di α ed il tempo indicato in un cerchio blu in alto a sinistra... aspetti poco rilevanti), ma soprattutto "cos" che di fisico stava a significare il moto della pallina.

Sono molte le ipotesi che in gruppo abbiamo discusso, qualcuna sicuramente mi sfugge... ma principalmente eravamo arrivati alla conclusione che il piano era suddiviso in porzioni di misure pari a (1;3;5;7...) (una successione di numeri dispari) che la pallina percorreva durante 1 oscillazione del pendolo...

1 1
2 3
3 5
4 7
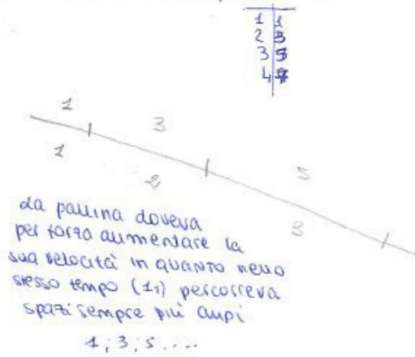
Circa all'inizio del video era poi comparsa una formula)

$$\frac{s}{t^2}$$; questa era

implicita in quanto appunto non contiene l'elemento che si voleva andare a trovare. Ricordo che tra le successive osservazioni avevamo dedotto si trattasse dell'accelerazione (ci è venuto poi smentito in seguito, in quanto ē = alla distanza)

da pallina doveva per forza aumentare la sua velocità in quanto nello stesso tempo (1) percorreva spazi sempre più ampi

1, 3, 5, ...

Fig. 8 Original version of text 2, page 1

Thinking back to the work carried out on the inclined plane, write to schoolmates of another class to outline the work itself and, specifically, the relationship that describes and explains mathematically the motion of the ball along the inclined plane. This report should be a theoretical support for you and your schoolmates

Through the project undertaken on September 29th, in which we were involved by a PhD student, S., and our math and physics teacher, S.B., we had the opportunity to learn many notions on the motion of the ball along an inclined plane.

The first meeting developed within an hour… we initially watched a video (it lasted just over a minute) and later (divided in groups) we described it according to two basic aspects (the visual one and the communicative one).

We were given a worksheet, where precisely we "physically" described what had visually impressed us (that is practically the inclined plane, the pendulum positioned on the opposite side of α and the time denoted in a blue circle on the top left…not so relevant aspects), but above all "what" the motion of the ball meant within physics.

Many are the hypotheses that we discussed in our group, I'm definitely missing something… but we mainly arrived at the conclusion that the plane was divided in portions of measures equal to (1; 3; 5; 7…) (a sequence of odd numbers) that the ball covered during one swing of the pendulum….
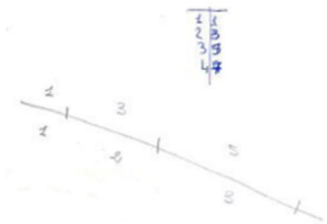
Nearly at the beginning of the video a formula appeared $\frac{s}{t^2}$; this was implicit since does not contain the element that we wanted to find. I remember that among the successive observations we deduced it was the acceleration (that was later denied, since it is equal to the distance).

The ball increased for sure its speed since in the same time (1 s) it traversed always larger distances 1; 3; 5…

Fig. 9 Translation into English of text 2, page 1

References

Arzarello, F. (2019). La covariación instrumentada: Un fenómeno de mediación semiótica y epistemológica, [Instrumented covariation: A phenomenon of semiotic and epistemological mediation]. *Cuadernos de Investigación y Formación en Educación Matemática*. Año 14. Número 18. pp 11–29. Retrieved from https://revistas.ucr.ac.cr/index.php/cifem/article/view/39888/40429

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*(5), 267–272. https://doi.org/10.1037/h0046669

Bagossi, S. (2021, April 8–12). *Toward second order covariation: Comparing two case studies on the modelling of a physical phenomenon*, [Paper presentation]. American Educational Research Association Annual Meeting.

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

Blum, W. (2015). Quality teaching of mathematical modelling: What do we know, what can we do?. In *The proceedings of the 12th international congress on mathematical education* (pp. 73–96). Springer, Cham. https://doi.org/10.1007/978-3-319-12688-3_9

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects—State, trends and issues in mathematics instruction. *Educational Studies in Mathematics, 22*(1), 37–68. https://doi.org/10.1007/BF00302716

Bourbaki, N. (1939). *Elements de mathématique: Structures fondamentales de l'analyse. Livre 1. Théorie des ensembles (fascicule de résultats)*. [Elements of mathematics: Fundamental structures of analysis. Book 1: Set theory (results leaflet)]. Hermann.

Bradley, R., & Terry, M. (1952). Rank analysis of incomplete block designs: I The Method of Paired Comparisons. *Biometrika, 39*(3), 324–345. https://doi.org/10.2307/2334029

Carlson, M. P. (1998). A cross-sectional investigation of the development of the function concept. *Research in collegiate mathematics education. III. CBMS issues in mathematics education*, 114–162.

Carlson, M. P., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education, 33*, 352–378. https://doi.org/10.2307/4149958

Confrey, J. (1991). The concept of exponential functions: A student's perspective. In *Epistemological foundations of mathematical experience* (pp. 124–159). Springer. https://doi.org/10.1007/978-1-4612-3178-3_8

Davies, B., Alcock, L., & Jones, I. (2020). Comparative judgement, proof summaries and proof comprehension. *Educational Studies in Mathematics, 105*, 181–197. https://doi.org/10.1007/s10649-020-09984-x

Duval, R. (2017). *Understanding the mathematical way of thinking – The registers of semiotic representation*. Springer. https://doi.org/10.1007/978-3-319-56910-9

Galilei, G. (1638). *Discorsi e dimostrazioni matematiche intorno a due nuove scienze attinenti la meccanica e i movimenti locali,* [Discourses and Mathematical Demonstrations Relating to Two New

Sciences]. Cierre, Simeoni Arti Grafiche. ISBN 9788895351049 (Translated into English by Henry Crew and Alfonso de Salvio. Macmillan, 1914: Downloadable from https://oll.libertyfund.org/title/galilei-dialogues-concerning-two-new-sciences).

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 113–133). Routledge. https://doi.org/10.4324/9780203063538

Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education, 39*(10), 1774–1787. https://doi.org/10.1080/03075079.2013.821974

Jones, I., & Inglis, M. (2015). The problem of assessing problem solving: Can comparative judgement help? *Educational Studies in Mathematics, 89*, 337–355. https://doi.org/10.1007/s10649-015-9607-1

Jones, I., & Karadeniz, I. (2016). An alternative approach to assessing achievement. *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education* – Szeged, Hungary. https://dspace.lboro.ac.uk/2134/21090

Jones, I., Bisson, M., Gilmore, C., & Inglis, M. (2019). Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal, 45*(3), 662–680. https://doi.org/10.1002/berj.3519

Jones, I., Swan, M., & Pollitt, A. (2015). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education, 13*(1), 151–177. https://doi.org/10.1007/s10763-013-9497-6

Laming, D. (1984). The relativity of 'absolute' judgements. *British Journal of Mathematical and Statistical Psychology*, 37(2), 152–183. https://doi.org/10.1111/j.2044-8317.1984.tb00798.x

Marshall, N., Shaw, K., Hunter, J., & Jones, I. (2020). Assessment by comparative judgement: An application to secondary statistics and English in New Zealand. *New Zealand Journal of Educational Studies, 55*, 49–71. https://doi.org/10.1007/s40841-020-00163-3

MIUR (2010a). Indicazioni nazionali per i licei. [National indications for high schools]. Author. Retrieved from: http://www.indire.it/lucabas/lkmw_file/licei2010a/indicazioni_nuovo_impaginato/_decreto_indicazioni_nazionali.pdf

MIUR (2010b). Linee guida per il passaggio al nuovo orientamento. Istituti tecnici. [Guidelines for the transition to the new orientation. Technical institutes]. Author. Retrieved from: http://www.indire.it/lucabas/lkmw_file/nuovi_tecnici/INDIC/_LINEE_GUIDA_TECNICI_.pdf

MIUR (2010c). Linee guida per il passaggio al nuovo orientamento. Istituti professionali. [Guidelines for the transition to new orientation. Professional institutes]. Author. Retrieved from: http://www.indire.it/lucabas/lkmw_file/nuovi_professionali/linee_guida/_LINEE_GUIDA_ISTITUTI_PROFESSIONALI_.pdf

MIUR (2012). Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. [National indications for kindergarten and first cycle of education]. Author. Retrieved from: http://www.indicazioninazionali.it/wp-content/uploads/2018/08/Indicazioni_Annali_Definitivo.pdf

Niss, M. (1989). Aims and scope of applications and modelling in mathematics curricula. In: W. Blum et al. (Eds), *Applications and modelling in learning and teaching mathematics*, 22–31.

Niss, M., Blum, W., & Galbraith, P. (2007). Introduction. In: W. Blum et al. (Eds), *Modelling and Applications in Mathematics Education. New ICMI Study Series* (vol 10, pp. 3–32). Springer. https://doi.org/10.1007/978-0-387-29822-1_1

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics, 102*(1), 9–28. https://doi.org/10.1007/s10649-019-09903-9

OECD-PISA (2022). *Mathematics framework* (draft2). Downloadable from: https://pisa2022-maths.oecd.org/ca/index.html; https://pisa2022-maths.oecd.org/files/PISA%202022%20Mathematics%20Framework%20Draft.pdf

Paola, D., & Impedovo, M. (2014). *Matematica dappertutto*. [Mathematics everywhere]. Zanichelli, ISBN: 9788808263940

Piaget, J. (1950). *Introduction à l'épistémologie génétique. Tome I: La pensée mathématique*. [Introduction to Genetic Epistemology. Volume I: Mathematical Thought]. Presses universitaires de France.

Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice, 19*(3), 281–300. https://doi.org/10.1080/0969594X.2012.665354

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346

Saldanha, L., & Thompson, P. W. (1998). Re-thinking co-variation from a quantitative perspective: Simultaneous continuous variation. In S. B. Berenson & W. N. Coulombe (Eds.), *Proceedings of the Annual Meeting of the Psychology of Mathematics Education – North America* (Vol. 1, pp. 298–304). North Carolina State University.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*(1), 20–26. https://doi.org/10.4324/9780203396391-18

Steedle, J. T., & Ferrara, S. (2016). Evaluating comparative judgment as an approach to essay scoring. *Applied Measurement in Education, 29*, 211–223. https://doi.org/10.1080/08957347.2016.1171769

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In R. A. Lesh & A. E. Kelly (Eds.), *Research design in mathematics and science education* (pp. 267–306). Lawrence Erlbaum Associates.

Tarricone, P., & Newhouse, C. P. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International Journal of Educational Technology in Higher Education, 13*(1), 1–11. https://doi.org/10.1186/s41239-016-0018-x

Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *Educational Studies in Mathematics, 25*(3), 165–208. https://doi.org/10.1007/BF01273861

Thompson, P. W. (2011). Quantitative reasoning and mathematical modeling. In L. L. Hatfield, S. Chamberlain & S. Belbase (Eds.), *New perspectives and directions for collaborative research in mathematics education. WISDOMe Mongraphs* (Vol. 1, pp. 33–57). University of Wyoming.

Thompson, P. W. (2015). Researching mathematical meanings for teaching. In L. D. English & D. Kirshner (Eds.), *Third handbook of international research in mathematics education* (pp. 968–1002). Taylor & Francis.

Thompson, P. W., & Thompson, A. G. (1992). *Images of rate*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. Retrieved from http://bit.ly/17oJe30

Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 421– 456). National Council of Teachers of Mathematics

Thompson, P. W., Hatfield, N. J., Yoon, H., Joshua, S., & Byerley, C. (2017). Covariational reasoning among US and South Korean secondary mathematics teachers. *The Journal of Mathematical Behavior, 48*, 95–111. https://doi.org/10.1016/j.jmathb.2017.08.001

Thurstone, L. (1927). A law of comparative judgement. *Psychology Review*, *34*(4), 273–286. https://doi.org/10.1037/h0070288

Vollrath, H. J. (1989). Funktionales denken [Functional thinking]. *Journal Für Mathematik-Didaktik, 10*(1), 3–37. https://doi.org/10.1007/BF03338719

Yoon, H., & Thompson, P. W. (2020). Secondary teachers' meanings for function notation in the United States and South Korea. *The Journal of Mathematical Behavior, 60*, 100804. https://doi.org/10.1016/j.jmathb.2020.100804