



A Tool for Comparing Mathematics Tasks from Paper-Based and Digital Environments

Alice Lemmo¹ 

Received: 13 September 2019 / Accepted: 8 August 2020 / Published online: 8 September 2020
© The Author(s) 2020

Abstract

Comparative studies on paper and pencil– and computer-based tests principally focus on statistical analysis of students’ performances. In educational assessment, comparing students’ performance (in terms of right or wrong results) does not imply a comparison of problem-solving processes followed by students. In this paper, we present a theoretical tool for task analysis that allows us to highlight how students’ problem-solving processes could change in switching from paper to computer format and how these changes could be affected by the use of one environment rather than another. In particular, the aim of our study lies in identifying a set of indexes to highlight possible consequences that specific changes in task formulation have, in terms of task comparability. Therefore, we propose an example of the use of the tool for comparing paper-based and computer-based tasks.

Keywords Comparative study · Computer-based assessment · Mathematics education · Task analysis · Task design

Introduction

This article stems from a wider Ph.D. research focused on comparing students’ problem-solving processes when tackling mathematics tasks in paper-based and computer-based environments (Lemmo, 2017).

The increasing use of tests administered in the digital environment allows research in mathematics education to develop new fields of study. On the one hand, research in computer-based tests is concerned with the validity of these tests, while, on the other hand, it focuses on their comparability with existing paper tests. Large-scale surveys were conducted to study these two issues; they involved students from different

✉ Alice Lemmo
alice.lemmo@gmail.com

¹ Dipartimento di Scienze Umane, Università degli studi dell’Aquila, Viale Nizza, 14, 67100 L’Aquila, AQ, Italy

educational levels, from primary to secondary school instruction (Drasgow, 2015; Way, Davis, & Fitzpatrick, 2006).

One of the first studies conducted on the topic involved the *National Assessment of Education Progress* (NAEP). Russell and Haney (1997) conducted a study to compare the effects of administering a test in two environments (paper and pencil vs computer) on performance (in terms of scores) of secondary school students. The findings revealed differences concerning the type of response: no substantial differences were identified in the case of multiple-choice items, while some such differences were found regarding open-ended items. Furthermore, research has shown that familiarity with the use of a keyboard allows students to obtain higher scores in digital mode than in paper-based tests (Russell, 1999; Russell & Plati, 2001). In general, the different studies carried out on the NAEP assessment system show how the performance of students completing computer-based tests is closely linked to their familiarity with the administrative environment. Similar research has been conducted using the *Oregon Statewide Reading and Mathematics tests* (Choi & Tinkler, 2002) and the *Texas statewide tests in mathematics, reading/English language arts, science and social studies* (Way et al., 2006). Differently from previous research, these studies show that the scores obtained in computer-based tests are higher than those with paper and pencil. Contrasting results have been observed with *Kansas Online Assessment* (Poggio, Glasnapp, Yang, & Poggio, 2005), *Florida State Assessment in high school reading and mathematics* (Nichols & Kirkpatrick, 2005) and *Virginia High School End of Course tests* (Fitzpatrick & Triscari, 2005). Significant differences in students' performance were found, with scores measured on the paper test resulting slightly higher than digital test results.

The research studies listed represent only a small part of those developed on the topic; from these and many others, it is clear that surveys mainly follow a statistical approach to comparison. In particular, they focus on measuring performance (correct or incorrect answers) and then on a quantitative analysis (Lottridge, Nicewander, Schulz, & Mitzel, 2008) that present conflicting findings (Organization for Economic Co-operation and Development [OECD], 2016).

In line with this statement, there is no shared perspective on whether computerised tests can be compared with those adopting a paper and pencil format. Such contradictions can be reasonable if these studies have been carried out on the assumption that comparing student performances can provide information about the problem-solving processes performed to provide the answers. On the other hand, it is important to question whether paper and digital tasks may be considered equivalent and, if so, whether it is possible to define an operational definition of equivalence.

Several authors mentioned above refer to *equivalent tasks*, opening up the question of the term "equivalence" to the reader. We question the idea that switching from paper to computer can be a neutral process; changes may occur due to the nature of the environment in which you are migrating (for example, the size of a sheet of paper is not always the same as the screen of a PC or a tablet, and the instructions provided to the user using paper and pencil can be different from those given to someone who uses software with particular digital tools). In light of this, when we consider the passage from one administrative environment to another, the matter of continuity, i.e. ensuring compliance in the information provided by statistical processing of results between the two versions of the task, is essential. A qualitative study could begin with the task analysis. Specifically, our goal is to identify a qualitative analysis methodology that,

starting from well-defined tasks, allows us to establish reasonable parameters of comparison. Analysis of students' behaviour must be based on accurate a priori analysis of the tasks (Margolinas, 2013) which is precisely the aim of this study. To reach this goal, we need to develop a theoretical framework that allows us to clarify when, and how, two tasks administered in different environments are comparable. We will do this by showing an example of task analysis in the Algebra content domain; however, our analysis criteria could be adopted in all mathematics content domains.

Two Approaches for the Construction of Computer-Based Tests

The issue of comparability of tests in the two environments could arise in the case of standardised assessments involving several countries. For example, according to the PISA 2015 survey, not all countries had the opportunity to use the computer-based testing method; for this reason, some countries chose to make continued use of paper and pencil environment. In order to maintain comparability and compatibility of the tests, it is necessary to identify which type of approach is used to design and implement the passage from one environment to another. Two types of approaches are described in the literature: *migratory* and *transformative* (Ripley, 2009). Ripley defines the migratory approach as the use of technological support as a tool of administration; it consists in a transition of tasks conceived in a paper format into a digital format. The transformative approach involves the transformation of original paper tests integrating new technological devices which support interactive tools (graphs, applets, etc.) that enhance new affordances. There are no specific studies comparing these two approaches; one possible and reasonable hypothesis is that migration could be a suitable approach for constructing what in the literature are called equivalent tasks.

Currently, there is a high number of tests aimed at the evaluation and certification of skills of all kinds; most of them, in particular those aimed at the assessment of school learning, are born in a paper environment and have always been administered in this way. Moreover, many assessment surveys, for example OECD PISA, re-propose some of the items administered in one cycle in the following cycle in order to make the data measured over several years comparable. Therefore, in view of the introduction of a digital version, there immediately emerges the issue of how to guarantee the possibility of such a comparison, usually indicated with the term *anchorage*. Given these requirements, the migratory approach seems to be the most suitable for the purpose of keeping the tests compatible in different environments in terms of both anchorage and affordance (Kaptelinin, 2013).

On the implications imposed by a migratory approach, the national centre concerned with American education processes (NCES¹) financed projects in order to highlight the strengths and weaknesses of this choice on the website of NAEP: *Mathematics Online* (MOL) (Bennett, Braswell, Oranje, Sandene, Kaplan, & Yan, 2008) and *Writing Online* (WOL) (Horkay, Bennett, Allen, Kaplan, & Yan, 2006). In these studies (Sandene, Horkay, Bennett, Allen, Braswell, Kaplan, & Oranje, 2005), several questions are addressed, regarding which type of paper item is best adapted to migration into the computer environment, how to make data obtained in one environment

¹ National Center for Education Statistics, <http://nces.ed.gov/>

comparable with that of another, which types of software and operating systems are better to use in order to make the administration accessible for schools and much more.

In line with the aims of this study, we examine whether the migratory approach may have effects on students' problem-solving processes when analysing tasks.

Definition of Variables for Task Comparison

Successful migration from one environment to another one depends on the intrinsic properties of the environments. Also, when using a migratory approach, the aim of defining criteria for comparability between two tasks first imposes the problem of defining a mathematical task. In general, there is no univocal definition of this term, as shown by Margolinas (2013) in an ICMI study on task design: "The word 'task' is used in different ways. In activity theory (Leont'ev, 1977), task means an operation undertaken within certain constraints and conditions (that is in a determinate situation, see Brousseau (1997). Some writers (Christiansen & Walter, 1986; Mason & Johnston-Wilder, 2006) express 'task' as being what students are asked to do. [...] Other traditions (e.g. Chevallard, 1999) distinguish between tasks, techniques, technology and theories, as a way to acknowledge the various aspects of a praxeology. We are also aware that 'task' sometimes denotes designed materials or environments which are intended to promote complex mathematical activity (e.g. Becker and Shimada (1997)), sometimes called 'rich tasks'." (Margolinas, 2013, p. 9).

In line with the aims of our work, we refer to a classical definition provided by Leont'ev (1977) of task as *activity*. According to the author, an activity is a unit of observation for which an individual acts towards a goal; this goal motivates the activity itself and defines its meaning and direction. Leont'ev has a hierarchical view of activity as a complex of *actions*. Each activity consists of a series of actions that are carried out by the individual to achieve a purpose related to the objective of the activity. Therefore, these actions are undertaken consciously and are part of the more general plan of the activity. In the same way, the actions are implemented through a series of *operations* which constitute the basic units of the actions.

We choose to define the task as an activity because this perspective allows us to focus on the processes and procedures (and therefore the actions and operations) that are activated to solve it. In this way, it is possible to go beyond the collection and analysis of the products, or of simple results, and to focus on the entire resolution process and how it is articulated from the student's first impact with the task until the final response is decided. The main motivation of the activities is to provide an answer to a request. This answer can be reached via procedures activated to respond on the basis of information, conditions and limits presented in the task and therefore in reference to the nature of the task in itself, which we define as the *content of the task*.

In addition, an important aspect to keep in mind is that the subject who designs and administers a task (usually the teacher or researcher) is not the same person who has to solve it (usually the student). In other words, the teacher/researcher who gives the student a task chooses a form of communication to be used, on the one hand to assign the task and on the other to allow the student to activate certain procedures/resolutive operations which he/she intends to use to resolve the task. We call this method of communication *format of the task*.

Finally, aspects related to the actions that can be activated by the student should not be overlooked, regarding the environment in which the task is presented. These aspects are also part of the peculiarities of the task, especially in relation to the task definition that we have chosen, i.e. in terms of activity and so of actions and operations engaged by students. We call this dimension *solution of the task*.

In this perspective, an analysis of a task is opened through three interconnected dimensions: content of the task, format of the task and solution of the task. Below, we will analyse each dimension; in particular, we will refer to national and international researches related to *mathematical word problems*.

In the literature, there are many different definitions of *mathematical word problems*; in particular, we refer to Gerofsky (1996). For the author, a mathematical word problem is a task presented through a written text and possibly integrated through mathematical symbolism. Often, word problems also involve narrative aspects because they describe real situations with characters who perform certain actions; for this reason, they are also often called *story problems* (Verschaffel, Greer, & De Corte, 2000). We choose a general definition in order to widen the analysis also for tasks that generally are not recognised as word problems even if they include a written text, symbols or images. This kind of task may change in the environment migration, and we want to create analysis criteria that take it into account.

The Content of the Task

Below, we analyse the content of the task and thus the task as an activity in itself, regardless of the format chosen to communicate it to the student. To simplify the analysis of the task content, we chose to consider this content divided into simpler components by referring to some studies developed on word problems. Johnson (1992) and Gerofsky (1996) describe word problems in terms of 3 main components:

Most word problems, whether from ancient or modern sources, and including ‘student-generated’ word problems, follow a three-component compositional structure:• A ‘set-up’ component, establishing the characters and location of the putative story. (This component is often not essential to the solution of the problem itself)

- An ‘information’ component, which gives the information needed to solve the problem (and sometimes extraneous information as a decoy for the unwary)
 - A question
- (Gerofsky, 1996, p. 2)

The *set-up* component regards the description of the task from a narrative point of view, i.e. regarding a timeline of events in which the characters presented have certain goals and perform specific actions. The literature shows different findings which suggest that variations in the set-up component in a word problem, or more generally a task, can influence the choices of students (Kulm, 1979; Pimm, 1995). However, a change in the set-up component is not consistent with the migratory approach because it would alter excessively the content of the task. For this reason, we choose to ignore variations of this component.

Information

The information component refers to what is generally called *data*, i.e. the set of information through which the student can determine an answer to the questions presented. This component generally has a strong influence on the cognitive dimension in relation to the mathematical skills and knowledge of the student. In fact, information is the collection of data with which the student operates in the various actions and operations that he/she activates to achieve the goal of the activity. For this reason, we hypothesise that even a small variation in the information could change the nature of the task as regards actions that can be activated.

This is not a new problem; in fact, research shows that familiarity with the numerical data presented in the information of a word problem can create difficulties and hinder the resolution process. Fischbein, Deri, Nello, and Marino' (1985) studies on this subject are well-known, especially those referring to word problems involving multiplication and division. Through several examples, they show that the complexity, in terms of percentage of correct/wrong answers, changes when considering tasks with variation in numerical data. In these various studies, Fischbein carried out a very extensive research on word problems, considering the different components and varying them all (either a limited number or one at a time). Starting from this point, Sbaragli (2008) experiments with new word problems minimising variations on the set-up to focus attention only on the information component. Her results confirm the evidence proposed by Fischbein: it is the numbers presented in the task and the relationships that the student identifies among them that hinder or facilitate recognition of the operation to be performed. We remember also that considering the information component does mean taking into account not only numerical information but also other information: for example, instructions for answering the question or other.

The aims of the aforementioned studies are different from ours; what interests us here is not understanding the causes of these difficulties but showing that the choice of information has a certain impact on the student's actions. This means that a possible variation in the information can cause even considerable differences in the actions the student takes to solve the task (Table 1). It is clear that the more varied the information, the less possible it is to consider two tasks as comparable.

Table 1 Description of the elements of the component information

Content of the task		
Component	Elements	Variables
Information	1. Presence of information	The information is presented: <ul style="list-style-type: none"> - In the set-up - In the question - In the pictures/images/graphics - Separated by the other components

Question

Most of the research found in the literature on word problems focuses mainly on the components set-up and information. Few studies focus on the question; most of them deal with the problem of its formulation in terms of text rather than the *question type*.

The reason can be traced back to scholastic tradition; in general, the typical assessment processes of Italian teaching practice, at least as far as mathematics is concerned, involve the administration of open-ended question tasks in which students must describe certain procedures, write calculations, justify their choices, etc. In Italy, there are few cases where students have to answer multiple-choice, univocal² or cloze³ questions. This is evident in the most common textbooks or when analysing the final examination at the end of secondary school. In recent years, however, the situation in Italy has started to change, and assessment teaching practice is now opening up towards other types of questions. This phenomenon can be traced back to the advent of standardised assessment tests that involve tasks with questions of various types: open-ended, univocal, multiple-choice, true/false or cloze questions.

Some studies on the issue of standardised tests exist: Kazemi (2001) studies students' performance in relation to a test, focusing on the type of question presented. Starting with multiple-choice questions and replacing them with open-ended questions, the author observed that the question type has a strong impact on answers provided by students, particularly on their perception of the difficulty of the task. The data collected by the author shows that open-ended questions have a higher percentage of missing answers than closed-answer questions.

The type of question may be an aspect of the question component that may change. For example, some open-ended questions that involve producing a drawing in a paper and pencil environment may be changed to closed-answer items, due to the inability to provide a kind of graphical response in a digital environment. In the particular case of comparative studies on two environments, paper-based and computer, the study on the type of questions is of considerable interest. In particular, Russell and Haney (1997) present the results of students' performance in the two environments in relation to the type of question. They show that these differences depend on the type of question administered. In particular, the authors did not record substantial differences in student performance in multiple-choice questions, while they found significant differences in open-ended questions. The authors in fact show that the computer environment requires longer response times for students who are unfamiliar with writing on a keyboard. In this perspective, it is possible to hypothesise that in the migration from one environment to another, the change in the type of question could result in differences at least from the point of view of performance (Table 2).

A change in the content of the question does not only concern possible changes in the type of question. In some cases, two questions may remain of the same type yet vary, depending on the request submitted.

² A task with a *univocal* question is a task in which solvers respond with a unique, unambiguous and straight answer; e.g. a number, a word, etc.

³ A task with a *cloze* question is a task consisting of a portion of text from which a few words have been removed. The student has to enter the missing words from a set of available terms.

Table 2 Description of the elements of the component question (part 1)

Content of the task		
Component	Elements	Variables
Question	1. Question type	Closed questions: <ul style="list-style-type: none"> - Multiple-choice questions: student has to choose the correct answer between a set of answers; - True/false questions: student has to recognise correct or wrong statements; - Univocal questions: where the student must provide a brief answer or a numerical answer; - Others Open-ended questions: questions in which the student is asked to: <ul style="list-style-type: none"> - Describe a problem-solving process through the explication of calculations; - Describe a pre-established mathematical content; - Explain certain choices or opinions; - Others

For example, let us consider the task: “What is the result of 23×5 ?” The request of this task is to explain a result and therefore a number that is obtained as a product of two factors. This is an open-ended question just like the one proposed in this second task: “How do you find the result of 23×5 ?” In both tasks, the students are asked to activate actions related to the application of the multiplication algorithm, but, while in the first task they are asked to identify the product, then to write a number, in the second one they are asked to describe and/or explain the procedure to determine this number, and then to produce an argument. In the two examples presented, we notice that there is a variation in relation to what is asked for in the question, or what we call the *request type*. This is an important aspect in the study of the comparability of task content, given that what is requested in the question is closely connected with the reason for the activity, on which the actions and operations that are activated depend. In this perspective, a change in the request necessarily provokes different actions and therefore differences in the selection of the problem-solving process that can be activated. For this reason, in addition to the type of request, we add a further related element (Table 3).

Format of the Task

As regards the format of the task, in literature, there are several studies on the topic with particular reference to word problems. Verschaffel et al. (2000) study several

Table 3 Description of the elements of the component question (part 2)

Content of the task		
Component	Elements	Variables
Question	2. Request type	<ul style="list-style-type: none"> - The question asks you to determine a numerical value; - The question asks for argumentation or, more generally, for a text; - Other

difficulties related to the solution of word problems; in particular, the authors point out that many of these difficulties can be encountered in the understanding of the verbal text in which the task is presented, rather than in the implementation of the solution. This phenomenon has been known for several years, De Corte and Verschaffel (1985) show that the difficulties observed in relation to the process of solving word problems can be caused by inadequate interpretation of the text.

Choosing the task format may have some influence on the student that will solve it. In this perspective, it is possible that small changes in the format may influence the choices of the student concerning the approach to be adopted. With regard to Italian mathematics standardised tests, Franchini, Lemmo, and Sbaragli (2017) show that many wrong answers are related to difficulties in understanding the text of the item, in particular to the linguistic interpretation. In fact, most research on this topic refers to a particular element of the format of the task that is generally referred to as *text of the task*.

Text of the Task

The text of the task refers to the communication modality chosen by whoever presents the task to whoever solves it: defining the modality means first producing a *text* through which to communicate it. The term text refers to any linguistic production, even oral, of variable length (Ferrari, 2004). In a general perspective, a text is a semantic unit that is independent of the medium in which the meaning is expressed, or rather represented (Halliday & Hasan, 1985). In other words, the text of the task is a *system of signs* that convey the communication of this task between those who present it and those who solve it. It is therefore an organised set of *signs* that can belong to different *semiotic systems* (Duval, 1993). We share Peirce's perspective that a *sign* is "Something which stands to somebody for something in some respect or capacity" (Peirce, 1974, p. 228). In other words, the author interprets the signs as a means to represent something to someone, and, for this reason, we can consider a text as a representation of the task with which the student interacts.

The choice of the format in which the task is presented is linked to the production of a text representing it. In this perspective, the formulation of the text and therefore the choice of the system of signs that composes it is decisive in the process of solving a task.

Word problems, and in particular the texts of the tasks that we are going to consider in our study, are expressed through a verbal language that can be enriched by some element of symbolic, graphic or other language. When comparing tasks, therefore, a comparison of semiotic systems that have been used to construct the text in all the dimensions that we analyse (content, format and solution of the task) is of interest.

In considering the task components, variables related to the text of the task cannot be considered independently of the others. For this reason, they occur in all the components we analyse: content, format and solution of the task (Table 4).

Layout of the Task

Examining the format of the task calls for consideration of all communication variables, and not just those related to the text of the task. In other words, we also consider

Table 4 Description of the elements of the dimension text of the task

Elements	Variables
Systems of signs used to represent the component	<ul style="list-style-type: none"> - Verbal; - Graphic; - Iconographic; - Other

choices related to the spatial relationship of the various components that make up the task with respect to the environment in which it is presented (the sheet of paper or the computer screen). In particular, we refer to the choices related to the *layout of the task*, i.e. aspects related to the drafting characteristics of the text, the reciprocal position of the various components of the text, the choice of certain fonts and the presence or absence of spaces.

Specifically, we define as *stylistic choices* everything that refers to the selection of font of the characters and their size, the presence or absence of underlined and bold words, etc. This aspect, which may seem negligible, can have a decisive impact on students' choices in the process of solving tasks. For example, Zan (2012) stresses that pupils often show an attitude of *selective reading* of the text during word problem-solving. In other words, sometimes the students identify only the numerical data presented and then search in the verbal text for some keywords that can guide them in the choice of operations to be carried out to reach the solution. In this perspective, it is possible that particular stylistic choices, such as underlined or bold text, may influence the student to follow this particular approach.

In addition, it is important to take into account the reciprocal position of the components of the text of the task, what we call the *structure of the task*. Generally, when a task is expressed through a text, it consists of a verbal section, sometimes followed by images or graphs, with the question appearing only at the end. Sometimes, figures or other representations of the task data can be placed within the written text that frames it; however, they can also be placed on either side of the sheet. These variations could influence students' choices or at least impact an order in the reading.

The literature does not offer many studies related to the structure of the task. Some research studies show that differences in the task format can have an influence on the problem-solving process. For example, Thevenot and colleagues (Thevenot, Barrouillet, & Fayol, 2004) show that the relative position of some parts of the text can condition the responses of young students; in particular, the researchers note that positioning the question before useful information can facilitate students in adopting the correct solving process.

It suggests that variables relating to the structure of the task should also be monitored from the point of view of comparability (Table 5).

Solution of the Task

We have defined the task as an activity; this choice requires us to dwell on the actions and operations that can be activated by the student to complete the task and therefore

Table 5 Description of the elements of the dimension format of the task

Format of the task		
Component	Elements	Variables
Layout of the task	1. Stylistic choices	The font and style of the text may show: <ul style="list-style-type: none"> - Underlined words; - Bold words; - Large or small font size; - Other
	2. Structure of the task	This element refers to the relative positions of the various components of the task (set-up, information, question); in particular, they can be presented: <ul style="list-style-type: none"> - Consecutively; - Put side by side; - Separated by images, spaces, etc.; - Other

also the way in which he/she provides the answer to the task; we call this last dimension the *solution of the task*. This is a crucial aspect when considering the comparability of two tasks administered in different environments, such as paper and computers.

With regard to actions and operations that can be activated by the student, there are cases in which (in the paper version) the student has the opportunity to carry out a manipulation of the data to calculate the solution, directly on the sheet of paper. This condition may be missing in the digital version where the user can only look at the screen and read the text without any possibility of manipulation in the environment itself (except for the action of clicking on the answer to give when there is a multiple-choice question). This difference could be limited by inserting a text box in the task environment. However, in the paper environment, the student has a different freedom of expression than in the case of the text box: in paper and pencil mode, the student has the ability to produce sketches, set and perform calculations, and write texts in both natural and symbolic language; these actions are not always feasible in a simple text box where you can only introduce the characters of the keyboard or otherwise allowed by the writing medium available, and according to an organisation predefined by the software used. In this case, the text box is the space where the student can provide the answer and is not always designed to support all the actions that the student activates to find the answer.

Similarly, the variation in the way the student can respond may be significant. For example, in the case of paper and pencils, the student has the opportunity to communicate through different languages both in the resolution and in the response phases: graphic, verbal and iconographic; on the contrary, the digital environment is reduced to the limits of the peripherals that are made available to the user, usually keyboard and mouse. These devices do not always allow the use of different forms of language; for example, when using a keyboard, it is not always possible to use a symbolic verbal language or iconography.

In Table 6, we present the element linked with the dimension “solution of the task”.

Table 6 Description of the elements of the dimension “solution of the task”

Component	Elements	Variables
Solution of the task	1. Actions and operations that the student can activate to solve the task	The student activates actions and operations in a communication mode that provides for different semiotic systems, for example: <ul style="list-style-type: none"> - Verbal (written or oral); - Graphic; - Symbolic; - Iconographic; - Other
	2. Modalities of communicating the response	The modalities of communicating the response are supported by different semiotic registers, for example: <ul style="list-style-type: none"> - Verbal (written or oral); - Graphic; - Symbolic; - Iconographic; - Other

The Task Comparison Grid

In the previous paragraphs, we presented a system of criteria for the analysis of tasks. These criteria are constructed by determining and discussing different elements and variables defined on the basis of research studies that have different origins, i.e. research studies on word problems that are independent of each other.

Most of the studies and examples presented above are related to the content domain: number and algebra. There are two reasons for this choice: the first is that most of the research on this topic refers to this content domain; the second is that the example of analysis that we will propose in the following belongs to it. However, we believe that the task comparison grid is suitable for a broader task analysis and can be adapted to all content domains of mathematics.

The task comparison grid, which consists of merging the tables presented above, is therefore a general overview of possible variables to consider when comparing tasks. In other words, it is a system of criteria that we have built from the literature on word problems, based on each table presented above.

The comparison task grid is our reference point for identifying possible variations that can intervene in the migration process in order to provide information that allows us to define a priori the comparability of tasks, in reference to which and how many indexes change. Starting from this, our goal is to make these variations explicit and then relate them to how pupils' behaviour can vary or not.

From this perspective, the grid is built according to each of the three dimensions we have considered: content of the task, format of the task and solution of the task (Table 7).

With regard to the format of the task, we have observed that one of the key elements is the text. It is clear that the text is presented to communicate the task to the student, which means that it is a representation of the content of the task and therefore of every single component that we identified. This definition, however, relates only to the

Table 7 The task comparison grid

Indexes	Elements	Variables
Content of the task	1. Presence of information	The information is presented: <ul style="list-style-type: none"> - In the set-up; - In the question; - In the pictures/images/graphics; - Separated by the other components
Question	2. Systems of signs used to represent the information component	<ul style="list-style-type: none"> - Verbal; - Graphic; - Iconographic; - Other
	1. Question type	Closed questions: <ul style="list-style-type: none"> - Multiple-choice questions: student has to choose the correct answer between a set of answers; - True/false questions: student has to recognise correct or wrong statements; - Univocal questions: where the student must provide a brief answer or a numerical answer; - Others Open-ended questions: questions in which the student is asked to <ul style="list-style-type: none"> - Describe a problem-solving process through the explication of calculations; - Describe a pre-established mathematical content; - Explain certain choices or opinions; - Others
	2. Request type	<ul style="list-style-type: none"> - The question asks you to determine a numerical value; - The question asks for argumentation or, more generally, for a text; - Other
	3. Systems of signs used to represent the question component	<ul style="list-style-type: none"> - Verbal; - Graphic;

Table 7 (continued)

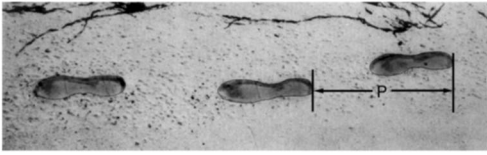
Indexes	Elements	Variables
Format of the task	Layout of the task	<ul style="list-style-type: none"> - Iconographic; - Other
	1. Stylistic choices	<p>The font and style of the text may show:</p> <ul style="list-style-type: none"> - Underlined words; - Bold words; - Large or small font size; - Other
Solution of the task	2. Structure of the task	<p>This element refers to the relative positions of the various components of the task (set-up, information, question); in particular, they can be presented:</p> <ul style="list-style-type: none"> - Consecutively; - Put side by side; - Separated by images, spaces, etc.; - Other
	1. Actions and operations that the student can activate to solve the task	<p>The student activates actions and operations in a communication mode that provides for different semiotic systems, for example:</p> <ul style="list-style-type: none"> - Verbal (written or oral); - Graphic; - Symbolic - Iconographic; - Other
2. Modalities of communicating the response	<p>The modalities of communicating the response are supported by different semiotic registers, for example:</p> <ul style="list-style-type: none"> - Verbal (written or oral); - Graphic; - Symbolic; - Iconographic; - Other 	

PISA 2015
? < >

Walking
Question 1/2

Refer to the information about measuring a person's walking pacelength. Type your answer to the question below.

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

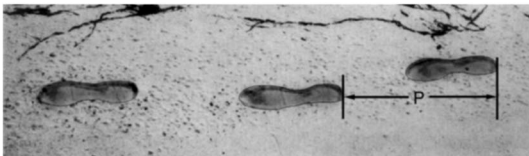


The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, gives an approximate relationship between n and P where,

n = number of steps per minute, and
 P = pacelength in metres.

WALKING



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula $\frac{n}{P} = 140$ gives an approximate relationship between n and P where
 n = number of steps per minute, and
 P = pacelength in metres.

Question 1:

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Fig. 1 Paper and digital versions of the item "walking"

content of the task and therefore disregards how each individual component is represented in the text that communicates it.

In constructing the task comparison grid, we decided to develop a projection of the three components so as to consider for each of them, on the one hand, the aspects related to the content and, on the other, those related to the text referring to the content. In the task comparison grid, the format of the task will refer only to the indexes linked to stylistic choices and to the structure of the task in general, while the text is declined for each single component.

Example of Use of the Task Comparison Grid

In order to better clarify task comparability in two environments, we propose below an example of the use of the task comparison grid. The task belongs to a mathematical standardised test; it is presented in the mathematics framework of OECD PISA 2015 (OECD, 2016). Since 2015, the PISA survey has been computer-based, and in 2021, mathematics will be the main field of measurement.

Let us consider the well-known task “walking” presented in the Draft of the Mathematical Reference Framework of the OECD PISA survey (Fig. 1).

In the passage from paper and pencil to computer environment, there are no changes in the text and there is no difference in the linguistic dimension of this text. In this particular task, the text related to the information component is not clearly identifiable with respect to the other components in any one environment. In fact, it is mixed with the text of the set-up component and the question; the data are represented through the verbal, symbolic and graphic language on one side in the set-up and on the other through a verbal and symbolic language within the application through a subordinate sentence. In fact, within the text that represents the set-up, we can observe the algebraic formula that indicates the relationship between the number of steps per minute, the pace length and a diagram that clarifies what is meant by pace length. In the question, the number of steps per minute is written.

There are no differences between the paper and computerised versions in terms of the text of this component, but a slight change can be observed with regard to the content; in fact, a text written in the verbal language has been added to the digital version, in which instructions on how to read the text of the task and answer the question are presented to the student.

This enriches the information component and increases the text length of the task related to that component. In both cases, the question is open-ended and communicated through verbal text. In this sense, there are no changes to the component, in terms of either format or content.

As regards the solution of the task, in the paper version, the student has a different expressive freedom compared with the case of the text box: in paper and pencil mode, the student has the possibility to produce sketches, set and perform calculations, and write texts in both natural and symbolic languages; none of which actions is allowed in a simple text box where he/she can only enter characters present on the keyboard or permitted by the available writing medium, and according to an organisation predefined by the software used. Even assuming that the student is familiar with the writing tools available (for example, the keyboard), it is reasonable to suppose that this change would result in significant differences in the solution process. It is clear that these differences have an impact both on the user’s possibilities for implementing actions and operations, and on how to respond.

Undoubtedly, there are substantial differences in the layout of the task in the two environments. In the paper version, the components follow each other vertically in order: explanatory image, set-up and information, question. In the digital format, however, the space of the task can be considered divided into two columns: on the right-hand side, the question is proposed with some information, while the left offers the image, the set-up and the remaining information. In the digital environment, the change in the task’s structure seems

Table 8 The task comparison grid applied to compare paper and digital versions of the task “walking”. The third column summarises the variations observed previously

Indexes	Elements	Changes
Content of the task		
Information	1. Presence of information	In both tasks, information is presented in the set-up and question components. In the digital environment, an explanatory paragraph on information is added before the question. This information does not appear in the paper environment.
	2. Systems of signs used to represent the information component	In both tasks, information is represented with verbal, graphical and symbolic systems of signs. In the digital environment, the explanatory paragraph increases the length of the verbal text.
Question	1. Question type	No
	2. Request type	No
	3. Systems of signs used to represent the question component	No
Format of the task		
Layout of the task	1. Stylistic choices	No
	2. Structure of the task	In the paper version, the components follow each other vertically in order: explanatory image, set-up and information, question. In the digital format, the space of the task is divided into two columns: on the right, the question is proposed with some information; on the left, there are the image and the set-up component.
Solution of the task		
Solution of the task	1. Actions and operations that the student can activate to solve the task	In the paper version, the students have a large expressive freedom, e.g. produce sketches, set and perform calculations, and manipulate algebraic formulas. In the digital version, the student cannot manipulate any formulas or image.
	2. Modalities of communicating the response	In the paper version, the student has the possibility to produce an answer enriched by sketches, calculations and symbols. In the digital format, there is a text box where student can only enter characters present on the keyboard or permitted by the available writing medium, and according to an organisation predefined by the software used.

to complicate the reading of the task: in fact, it passes from a linear reading in the paper environment to a reading that envisages two distinct columns in the computerised environment. In this second case, the reader needs to coordinate the interpretation of different parts in which the verbal text is divided.

Finally, in both tasks, the same picture is presented; nevertheless, it is possible to notice that in the digital version, the picture is presented on the screen with all the strengths and limitations of the software that supports it. For example, it might be difficult (or impossible) to analyse the image through common and simple manipulation actions such as turning the paper or complete the picture by drawing lines, highlighting points, etc.; these actions are possible only in a paper and pencil environment.

The comparative analysis can be summarised with the task comparison grid (Table 8).

Conclusion

The aim of our research was to identify parameters that would allow a qualitative analysis of tasks designed in two administration environments. To achieve our goal, we presented a system of criteria constructed by determining and discussing different elements and variables defined in different and independent research projects. The task comparison grid summarises such criteria and presents a general overview of the possible variables to consider when comparing tasks. It allows the identification of possible variations that may intervene in the migration process in order to provide information that allows us to define a priori the comparability of tasks. Starting from this point, we presented an example of task analysis.

As shown in the example, this grid allows the highlighting of particular differences related to different aspects of the task that depend strictly on the passage from one environment to another. The migration process could at first glance appear accurate, but deeper analysis shows otherwise. Initially, the minor highlighted differences might appear superfluous; however, they are crucial in order to analyse and interpret students' behaviour in the problem-solving process. The literature described in the first part of the paper indicates that each difference observed in the example may affect the student's approach. For instance, the change in the format of the task could simplify the text comprehension if it is presented in a linear way; on the contrary, the reading could be difficult if the verbal description is fragmented in several parts. These little differences reserve important results for assessment, especially if the purpose of the migration process is to ensure continuity between the paper mode and computer-based administration.

Many comparative studies presented in the literature assume that the tasks administered in the two environments are equivalent. However, our analysis shows that this view needs to be re-assessed. The equivalence between performances (in terms of right or wrong results) does not imply an equivalence between the processes adopted by students. Therefore, the analysis of the results collected in the two environments is probably not equivalent in terms of educational assessment. The answers produced by students in the digital environment seem hardly comparable with what they do on paper. Thus, there is a substantial difference in terms of assessment, which cannot be ignored, especially by national or international large-scale assessment authorities. In addition, the change in the solution of the task is crucial

because it strictly depends on the intrinsic feature of the environment and on the familiarity that the student has with the tool available.

We recall also that there are cases where it is impossible to translate a task from a paper to digital format through the migratory approach; for example, in the event of tasks that require the use of physical tools and measuring instruments such as a ruler, compass or other. One special case involves tasks with the goal of assessing students' drawing abilities. In this case, the item may require the student to draw a figure starting from a given skeleton model or measurement, or from written verbal instructions. In these cases, it is possible to introduce an ad hoc software or applet that simulates the use of drawing tools. However, the issue of students' familiarity with these software or applets arises (Bennett, Persky, Weiss, & Jenkins, 2010). In the case of lack of familiarity with the use of the instrument, the digital device could prove generally useless; students who are required to use digital tools may be disadvantaged in comparison with students who use paper and pencil and physical tools. This example highlights a very serious and complex issue. Further research is needed to define criteria of control that allow authorities to check and compare all the small differences that occur in the migration process.

As regards possible educational implications, the task comparison grid can open up different perspectives. First, the correlation between the variations detected by the grid and the subsequent study of the students' solving behaviour can be a diagnostic tool for the teacher. Indeed, the use of the grid can allow the isolation of particular variables of a task to identify where students have difficulties in the process of solving that task. These difficulties may not necessarily depend on strictly mathematical aspects but on transversal issues linked, for example, to the comprehension of the task text or choice of set-up. In the same way, the isolation of variables may be considered critical because link to particular difficulties can be the basis for the teacher to construct specific didactic situations in which tasks are presented that allow students to work on these variables.

Starting from these differences, it will be interesting to identify which of them can lead students to adopt particular behaviours and how these can depend on the administrative environment. Obviously, this does not necessarily mean that the performance of learners will be different in the two environments (this should be shown through ad hoc research), but this tool is significant because it allows a prior comparison of two tasks in order to interpret the behaviour of learners later. In a larger study, these differences were then studied with reference to the behaviour of students in solving the tasks (Lemmo, 2017).

Komatsu and Jones (2019) highlight that an approach to enriching mathematical learning is to enhance task design, because, as ICMI Study 22 confirmed (Watson & Ohtani, 2015), tasks are the bedrock of day-to-day mathematics lessons and thus have a significant effect on student experience of mathematics and their understanding of the nature of mathematical activity. In the words of Sierpiska (2004, p. 10), "the design, analysis and empirical testing of mathematical tasks, whether for the purposes of research or teaching ... [is] one of the most important responsibilities of mathematics education". For these reasons, the verification of these differences is a general problem that could also be presented in other contexts, independently of a migration process in a new environment.

Acknowledgements I am deeply grateful to Professor Maria Alessandra Mariotti for her constant guidance and for all suggestions and ideas she shared with me. I am very grateful to Professor Giorgio Bolondi for his support and useful discussions.

Funding Open access funding provided by Università degli Studi dell'Aquila within the CRUI-CARE Agreement. This research is supported by a grant from Ministero dell'Istruzione, dell'Università e della Ricerca (IT) (PON-AIM1849353 - 3).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Becker, J. P., & Shimada, S. (1997). *The open-ended approach: A new proposal for teaching mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment*, 6(9).
- Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2010). Measuring problem solving with technology: A demonstration study for NAEP. *The Journal of Technology, Learning and Assessment*, 8(8).
- Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Choi, S. W., & Tinkler, T. (2002). *Evaluating comparability of paper-and-pencil and computer-based assessment in a K-12 setting*. New Orleans, LA: Annual meeting of the National Council on Measurement in Education.
- Christiansen, B., & Walter, G. (1986). Task and activity. In B. Christiansen, A.-G. Howson & M. Otte (Eds.), *Perspectives on mathematics education: Papers submitted by members of the Bacomet Group* (pp. 243-307). Dordrecht, The Netherlands: D. Reide.
- De Corte, E., & Verschaffel, L. (1985). Beginning first graders' initial representation of arithmetic word problems. *The Journal of Mathematical Behavior*, 4, 3-21.
- Dragow, F. (Ed.). (2015). *Technology and testing: Improving educational and psychological measurement*. Abingdon, England: Routledge.
- Duval, R. (1993). Registres de représentation sémiotique et fonctionnement cognitif de la pensée [Registries of semiotic representation and cognitive functioning of thought]. *Annales de didactique et de sciences cognitives*, 5, 37-65.
- Ferrari, P. (2004). *Matematica e linguaggio: Quadro teorico e idee per la didattica* [Mathematics and language: Theoretical framework and ideas for teaching]. Bologna, Italy: Pitagora.
- Fischbein, E., Deri, M., Nello, M. S., & Marino, M. (1985). The role of implicit models in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education*, 16(1), 3-17.
- Fitzpatrick, S., & Triscari, R. (2005, April). *Comparability studies of the Virginia computer-delivered tests*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Franchini, E., Lemmo, A., & Sbaragli, S. (2017). Il ruolo della comprensione del testo nel processo di matematizzazione e modellizzazione [The role of text comprehension in the process of mathematization and modeling]. *Didattica della matematica. Dalle ricerche alle pratiche d'aula*, 2017(1), 38-63.
- Gerofsky, S. (1996). A linguistic and narrative view of word problems in mathematics education. *For The Learning of Mathematics*, 16(2), 36-45.
- Halliday, M. A., & Hasan, R. (1985). *Language, context and text: Aspects of language in a social-semiotic perspective*. Geelong, Australia: Deakin University.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment*, 5(2).
- Johnson, M. (1992). *How to solve word problems in algebra*. New York, NY: McGraw-Hill.
- Kaptelinin, V. (2013). Affordances. In M. Soegaard, R. F. Dam, & M. A. Soegaard (Eds.), *The encyclopedia of human-computer interaction* (2nd ed.). Aarhus, Denmark: The Interaction Design Foundation.

- Kazemi, E. (2001). Exploring test performance in mathematics: The questions children's answers raise. *The Journal of Mathematical Behavior*, 21, 203–224.
- Komatsu, K., & Jones, K. (2019). Task design principles for heuristic refutation in dynamic geometry environments. *International Journal of Science and Mathematics Education*, 17(4), 801–824.
- Kulm, G. (1979). The classification of problem-solving research variables. In G. A. Goldin & C. E. McClintock (Eds.), *Task variables in mathematical problem solving* (pp. 1–22). Columbus, OH: ERIC Clearing house for Science, Mathematics and Environmental Education.
- Lemmo, A. (2017). *Dal formato cartaceo al formato digitale: Uno studio qualitativo di test di Matematica* [From paper to digital format: A qualitative study of Mathematics tests] (PhD Thesis). Università degli studi di Palermo. Retrieved from <http://hdl.handle.net/10447/220968>.
- Leont'ev, A. N. (1977). *Attività, coscienza, personalità* [Activity, consciousness, personality]. Firenze, Italy: Giunti.
- Lottridge, S., Nicewander, A., Schulz, M., & Mitzel, H. (2008). *Comparability of paper-based and computer-based tests: A review of the methodology*. Monterey, CA: Pacific Metrics Corporation.
- Margolinas, C. (2013) Task Design in Mathematics Education. *Proceedings of ICMI Study 22*. ICMI Study 22, Jul 2014, Oxford, United Kingdom.978-2-7466-6554-5. hal-00834054v3
- Mason, J., & Johnston-Wilder, S. (2006). *Designing and using mathematical tasks*. York, England: QED Press.
- Nichols, P., & Kirkpatrick, R. (2005). *Comparability of the computer-administered tests with existing paper-and-pencil tests in reading and mathematics tests*. Montreal, CA: Annual Meeting of the American Educational Research Association.
- Organization for Economic Co-operation and Development. (2016). *PISA 2015 assessment and analytical framework: Science, reading, mathematics and financial literacy*. Paris, France: OECD Publishing.
- Peirce, C. S. (1974). *Collected papers of Charles Sanders Peirce* (Vol. 2). Harvard University Press.
- Pimm, D. (1995). *Symbols and meanings in school mathematics*. London, England: Routledge.
- Poggio, J., Glasnapp, D. R., Yang, X., & Poggio, A. (2005). A comparative evaluation of score results from computerized and paper & pencil mathematics testing in a large scale state assessment program. *The Journal of Technology, Learning and Assessment*, 3(6).
- Ripley, M. (2009). Transformational computer-based testing. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment* (pp. 92–98). Luxembourg: Office for Official Publications of the European Communities.
- Russel, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7(20), 1–47.
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education policy analysis archives*, 5(3), 1–18.
- Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *The Teachers College Record*.
- Sandene, B., Horkay, N., Bennett, R. E., Allen, N., Braswell, J., Kaplan, B., & Oranje, A. (2005). *Online assessment in mathematics and writing*. Reports from the NAEP technology-based assessment project, research and development Series. National Center for Education Statistics.
- Sbaragli, S. (2008). La divisione, Aspetti concettuali e didattici [The division, conceptual and didactic aspects]. In B. D'Amore & S. Sbaragli (Eds.), *Didattica della matematica e azioni d'aula. Atti del convegno Incontri con la matematica n.* (22nd ed., pp. 151–154). Pitagora: Bologna.
- Sierpiska, A. (2004). Research in mathematics education through a keyhole: Task problematization. *For the Learning of Mathematics*, 24(2), 7–15.
- Thevenot, C., Barrouillet, P., & Fayol, M. (2004). Mental representation and procedures in arithmetic word problems: The effect of the position of the question. *L'Année Psychologique*, 104(4), 683–699.
- Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.
- Watson, A., & Ohtani, M. (2015). Themes and issues in mathematics education concerning task design: Editorial introduction. In A. Watson & M. Ohtani (Eds.), *Task design in mathematics education: An ICMI study* (22nd ed., pp. 3–15). New York, NY: Springer.
- Way, W. D., Davis, L. L., & Fitzpatrick, S. (2006). *Practical questions in introducing computerized adaptive testing for K-12 assessments*. San Antonio, TX: Pearson.
- Zan, R. (2012). La dimensione narrativa di un problema: Il modello C&D per l'analisi e la (ri)formulazione del testo. Parte II [The narrative dimension of a problem: The C&D model for the analysis and (re)formulation of the text. Part II]. *L'insegnamento della matematica e delle scienze integrate*, 35(2), 437–467.