Conference papers

School of Computer Sciences

2020

Direct and Constructivist Instructional Design: A Comparison of Efficiency Using Mental Workload and Task Performance

Giuliana Orru

Luca Longo

Direct and Constructivist Instructional Design: A Comparison of Efficiency Using Mental Workload and Task Performance

2 authors:

Giuliano Orru
Technological University Dublin - City Campus
6 PUBLICATIONS   **73** CITATIONS

SEE PROFILE

Luca Longo
Technological University Dublin - City Campus
88 PUBLICATIONS   **1,058** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Community of inquiry and Cognitive load Theory View project

Direct and Constructivist Instructional Design: A Comparison of Efficiency Using Mental Workload and Task Performance

Giuliano Orru and Luca Longo[(✉)]

School of Computer Science, College of Health and Sciences, Technological University Dublin,
Dublin, Republic of Ireland
Luca.Longo@tudublin.ie

Abstract. Cognitive Load Theory is based upon the assumption that working memory can process only explicit and direct instructions. Therefore, it is believed that inquiries techniques, not employing explicit instructional methods for teaching, are set to fail. This paper aims to fill this gap by extending the traditional direct instruction teaching method, with a highly guided inquiry activity and comparing their efficiency. In detail, the efficiency of the former is expected to be lower than the efficiency of the latter hybrid method. The likelihood model of efficiency, originally based upon a unidimensional subjective measure of effort and an objective measure of performance, was originally applied and extended with a multidimensional measure of cognitive load. Empirical evidence partially supports the above hypothesis but reveals some limitations of the traditional statistical tools applied to group comparisons of small sample size, often the case in higher educational settings. This suggests that future scholars should dedicate effort on the identification and application of statistical methods for the analysis of the efficiency of instructional conditions in small sample-size groups.

Keywords: Cognitive load theory · Mental workload · Efficiency · Direct instruction methods · Inquiry methods

1 Introduction

Cognitive Load Theory (CLT), a well-known theory in educational psychology, is structured on the premise that working memory can process explicit and direct instructions only [1]. Kirschner and colleagues (2006) criticized experiments based on unguided collaborative methodologies because they tend to ignore the role of working memory in the process of learning. Consequently, teaching approaches that do not explicitly consider the importance of direct instructions are set to fail [1]. In CLT, the acquisition of knowledge is supposed to take place within human cognitive architecture whose core system is working memory. If explicit instructions are not considered, working memory cannot process the information related to any underlying learning task. This is the gap emerged from the literature review connected to CLT and the constructivists techniques usually used in collaborative learning. Jonassen (2009), replying to Kirchner et al.

(2006) affirms that, in the field of educational psychology, a comparison between the effectiveness of inquiry methods and direct instruction is not possible [2] because the two approaches come from different theory assumptions and they use different research methods. Constructivists methods usually employ qualitative analysis. On the contrary, direct instructions methods, as per those used in CLT, usually employ quantitative analysis [2]. Quantity and quality cannot be compared as they are different in their nature, consequently Jonassen (2009) states that the assumption of Kirchner and colleagues (2006) is not supported by any empirical comparison or evidence. Besides that, another problem is the lack of a shared learning outcome. The two methodologies must share the same dependent variable to be compared. In the case of the current set of experiments, the achievement of factual, conceptual and procedural knowledge is the shared dependent variable, evaluated by multiple choices questionnaire (MCQ) as performance test. The current research is motivated by the aforementioned gap between explicit instructional designs characteristic of CLT, and the features of the community of inquiry approach that focus on the learning connection between cognitive abilities and knowledge construction. The proposed solution is to extend the approach of direct instructions connected to a learning task with highly guided inquiries activities, the purpose is to establish whether this extension improves the efficiency of learners compared with learners who receive direct instructions only. Empirically evaluating instructional approaches is not a trivial task in Pedagogy. The likelihood efficiency measure, proposed by Hoffman and Schraw (2010), has been employed to tackle this problem on. The related formula is based upon two other measures: the cognitive load, namely the cognitive cost of learning task on working memory, and performance score [3]. The research question being proposed in this study is:

to what extent can an inquiry activity based upon cognitive trigger questions, when added to a direct instruction conventional teaching method, improve its efficiency, impact the effort and mental workload experienced by learners and improve their learning performance?

The current research proposes a requalified teaching methodology that is aimed at combining the cognitive approach of CLT focused on explicit instruction and the community of inquiry approach based on cognitive questions by extending the first with the second and comparing its efficiency with that associated to explicit instructions only. Explicit instructions are direct, precise, specific and clear teaching explanations aimed at facilitating how to deal with a learning task or problem solving. The proposed inquiry activity is aimed at engaging learners in the learning process by the use of dialogue focused on cognitive trigger questions.

The reminder of this paper is structured as it follows. Section 2 informs the reader on the related background and the literature review that originated the current research. Section 3 describes the design of the empirical comparative experiment and the methodology employed while Sect. 4 outlines the analysis of the results of such comparison, the related effect sizes and proposes an interpretation of data. Section 5 summarizes the paper highlighting its contribution and delineating future work.

2   Literature Review and Background

The aim of this section is to inform the reader with the basic notions and background on cognitive load theory, mental workload and its measurement, collaborative learning and instructional efficiency, theoretical contents that are critical to give an account of the layout of the proposed experiment described in Sect. 3.

2.1   Cognitive Load Theory

Cognitive Load Theory is widely known in educational psychology as a reference to improve the learning process by developing new instructional techniques aligned with the limitations of the human cognitive architecture [4]. This happens by optimizing the cognitive load imposed on working memories while carrying on a task. Cognitive load is, in fact, the cognitive cost, in terms of memory resources, experienced by learners when performing a learning task [5]. Working memory and long term-memory are the two dimensions of human cognitive architecture that stores information, retrieving and processing it for reasoning and decision making [6–8]. The premise for acquiring knowledge in CLT is that learners have to be instructed by means of direct instructional designs [4, 9]. Studies in CLT are based on the comparison between a control and experimental group of learners, the former is taught according to a conventional instructional procedure, the latter according to a new instructional procedure [10]. A test phase follows to see if there are any differences in learning outcomes. If statistical analyses on the test results demonstrate that learning improves by the new instructional procedure, then a new cognitive load effect is demonstrated and a superior instructional procedure generated as summarized by Sweller in his recent review of CLT [4]. Among others, the Collective Working Memory effect is particularly relevant to give an account of collaborative techniques that enable learners to share working memories while attending the same task. The assumption is that the use of working memory of many people can reduce the overall cognitive cost of that task. Collaborative techniques are supposed to engage students in higher order skills [11] and in activities valuable to the enhancement of learning such as self-directed learning, negotiating, meaning, verbalizing, explaining, justifying and reflecting, as well as giving each other mutual support [12]. Along with these positive findings, however, there is also a body of research showing mixed and negative benefits regarding both the learning process itself [13, 14] and the dynamics of group formation [15]. The main negative effect is the cognitive cost of information transfer: the transactive interaction, could generate too much cognitive load hampering the learning phase instead of facilitating it [16]. This depends on the complexity of the task. In tasks with high level of complexity, in fact, the cognitive cost of transfer is compensated by the advantage of using working memories of multiple people. In contrast, in tasks with low level of complexity, the working memory of one person is supposed to be sufficient and the transfer costs of communication might hamper the learning phase.

2.2   Mental Workload

According to Wickens (1979) '… the concept of operator workload is defined in terms of the human's limited processing resources'. His Multiple Resource Theory (MRT) states

that humans have a limited set of resources available for mental processes [17]. Mental Workload can be defined as 'the volume of cognitive work necessary for an individual to accomplish a task over time' [18, 19]. It is not 'an elementary property, rather it emerges from the interaction between the requirements of a task, the circumstances under which it is performed and the skills, behaviors and perceptions of the operator' [20–22]. However, these are only practical definitions, as many other factors influence mental workload [23–26]. The concept of cognitive load is mainly employed within Education whereas the concept of Mental Workload mainly in Ergonomics [27–30]. The former relates to working memory resources only, whereas latter takes into account other factors as the level of motivation, stress and the physical demand experienced by participants as a consequence of the task. Despite of their different fields of research, CLT and MRT share a common assumption: the limits of working memory must be considered to predict humans' performance while accomplishing an underlying task. A well-known multi-dimensional self-reporting measures of mental workload is the NASA-task Load Index (NASA-TLX) [31–33]. In contrast to unidimensional scales of overall cognitive load, such as effort and task difficulty proposed by Paas and Van Merriënboer (1993, 1994) [34, 35], Paas et al. (2003) [36] and Zijistra [37], NASA-TLX focuses on different components of load as per mental, physical, and temporal demands, frustration, effort, and performance. In education it is not widely employed, however, a number of studies have confirmed its validity and sensitivity [38–40]. A lighter version of the NASA-TLX exists. This is the RAW-NASA-TLX, in which the weighting process employed in the original questionnaire is eliminated.

2.3 Genealogy of Community of Inquiry

On the relation between instructional technique and its capacity to have an impact on learning outcomes, John Dewey suggests to re-think the semantic distinction between 'Technique' as practice and 'Knowledge' as pure theory. Practice, in fact, is not foundationalist in its epistemology anymore, in other words it does not require a first principle as its theoretical foundation. In the philosophy of Dewey, 'Technique' means an active procedure aimed at developing new skills starting from the redefinition of the old as resumed in the 'Theory of Inquiry' of Dewey [41]. Therefore, the configuration of epistemic theoretical knowledge is a specific case of technical production and 'Knowledge' as a theory is the result of 'Technique' as practice. Both are deeply interconnected and they share the resolution of practical problems as starting point for expanding knowledge [41]. The aim of this pragmatist approaches to learn is improving the techniques of humans by a process of autocorrective feedback within inquiring environments [41]. Inquiry is proposed as teaching and learning technique that is deeply linked with a continuous autocorrective process of knowledge's development: this is the reason why Dewey identifies technology, the discipline whose focus is the study of techniques, with education. Through the process of inquiry an unsatisfactory situation can be converted into satisfactory by connecting all of its constituent into a coherent and unified whole [41]. This is the reason why, in educational contexts, inquiry techniques are proposed to improve the comprehension of complex learning tasks [42]. The research of Garrison (2007) conducted on the community of inquiry online was influenced by the philosophy

of education of Dewey and by its notion of inquiry. Garrison provided a clear exemplification of the social and cognitive structure a community of inquiry consists of [42]. The social context is based on democratic features as free risk expression, encouragement and collaboration. The cognitive dimensions, instead, consist in exploring a problem by exchanging information on its constituent parts, integrating that information by connecting related ideas and solving it suggesting alternatives and new ideas. The core ability in solving a problem consists in connecting the right tool to reach a specific aim. The community of inquiry may be defined as 'a teaching and learning technique, an instructional technique of a group of learners who, through the use of dialogue, examine the conceptual boundary of a problematic concept proceeding all the parts this problem is composed of in order to solve it' [30]. Inquiring may be defined, in a nutshell, as critical and creative thinking. The former consists in connecting tools with aims consistently, in order to solve problems and expanding human knowledge, the latter consists in coherently connecting the constituents parts of a problem with its whole [43]. Dewey and after him Lipman who extended the community of inquiry with a philosophical model of reasoning, connected the meaning of inquiry with the meaning of community, the individual and the community can only exists in relation to each other along a continuous process of adaptation that ends up with their reciprocal, critical and creative improvement [41–43]. In line with the definition of inquiry, a pedagogical framework grounded in the 'Philosophy for Children' proposed by Mathew Lipman exists (the project NORIA) [44]. It proposes a set of cognitive questions aimed to exercise the cognitive abilities of a learner and to develop a higher level of thinking. The goal of the trigger questions is to support the development of the cognitive skills of learners [44]. Among others, they are aimed at developing cognitive skills of conceptualization by comparing and contrasting, defining, classifying, and reasoning by relating cause and effect, tools and aims, parts and whole and by establishing criteria [44].

2.4 Taxonomy of Anderson and Models of Instructional Efficiency

The taxonomy of educational objectives proposed by Anderson distinguishes between dimensions of knowledge (factual, conceptual, procedural and metacognitive knowledge) and dimension of cognitive process (remember, understand, apply, analyze evaluate and create) [45]. These dimensions of educational objectives assume a critical importance in the design of the experiment proposed in Sect. 3. If a performance test, as per Multiple Choice Questions, is aligned with the aforementioned dimensions of knowledge, then it can be employed to compare the performance of students who receive different teaching methods based on the same contents. In problem solving, learning and instruction, efficiency is generally defined in terms of ability to reach established goals by minimal expenditure of time, effort or cognitive resources [3]. The Likelihood model computes a measure of efficiency based on the ratio of work output to work input. Output can be identified with learning, input with work, time or effort [46]. The likelihood model has been widely used in educational psychology to analyze relative gain between two variables as a consequence of a methodological intervention [47, 48]. It is based on the ratio between performance and effort, a raw score for test performance or a learning outcome denoted as P divided by a raw score for time or effort denoted as R:

$$Efficiency \ = \ P \ / \ R$$

R can be any self-report or an objective measure based on time or cognitive resources employed [3]. An estimation of the rate of change of performance is calculated by dividing P per R. This ratio diverges from zero to extensive positive values, it goes towards zero when performance is low and effort is high (low efficiency), it goes towards extensive positive values when performance is high, and effort is low (high efficiency). The result represents the individual efficiency based on individual scores [3].

2.5 Summary

The gap emerged from the literature review points to the relationship between working memory and constructivists techniques: inquiry techniques are believed to fail in educational contexts because of their lack of direct and explicit instructions that are instead required by working memory to process information as proposed within Cognitive Load Theory [49]. However, this claim is not backed up by empirical tests and a lack of comparison of their effectiveness exists. This is justified by the fact that, on one hand, constructivists techniques usually employ qualitative research methods, while, on the other hand, those employed within cognitive load theory are quantitative. This, make such a comparison non-trivial. However, multiple choice questionnaires, as extensively used within education [50], is a potential tool for supporting such a comparison and allowing the evaluation of factual, conceptual and procedural knowledge as well as the assessment of learning produced with different instructional approaches. Formulas of efficiency for the evaluation of instructional conditions exist, such as the likelihood model, based upon the combination of test performance scores and a measure of effort or cognitive load. The likelihood model seems to be effective in detecting changes within experimental studies.

3 Design, Material and Methodology

Given the research question of Sect. 1, a primary research experiment has been designed and the following research hypothesis was set:

H1: if an explicit instructional design method is extended with an inquiry activity focused on cognitive trigger questions, then, its efficiency is improved, the experienced effort and mental workload of learners is impacted, and their learning performance is positively enhanced.

The approach employed in CLT and the collective working memories principles recently summarized by Kirschner and colleagues [51], have been taken in consideration in the design of the experiment. In the current research, in fact, two instructional design conditions were compared: one following the direct instruction approach to learning (instructional condition 1), and one that extends this with a collaborative activity inspired by the community of inquiry approach to learning (instructional condition 2). In detail, the former involved a theoretical explanation of an underlying topic, whereby an instructor presented information through direct instructions along a set of slides. The latter involved the extension of the former with a highly guided inquiry activity based upon cognitive trigger questions. These questions, aligned to the Anderson's taxonomy,

are supposed to develop cognitive skills in conceptualizing and reasoning that stimulate knowledge construction in working memory [52]. All lecturers at Technological University Dublin, School of Computer Science were contacted by email and invited to take part in the experiment. Only a number of them accepted and among these, only those using the traditional direct instruction approach to learning were selected. Each of these lecturers identified one or two suitable topics, already part of their modules, for experimental purposes. The material required by each selected lecturer was: 1) a set of slides on the selected topic. 2) A computer connected to a projector to display these slides in the classroom. 3) MCQ as originally proposed by Haladyna (2002) [50] designed on the contents of the slides to evaluate factual, conceptual and procedural knowledge. 4) a set of trigger questions designed for the selected topics as part of the inquiry activity of instructional condition 2. An example of guidelines for the inquiry activity (instructional condition 2) and examples of trigger cognitive questions on the topic 'Semantic Web' are depicted in Table 1.

Table 1. Community of Inquiry guidelines and examples of trigger questions employed during the inquiry activity in the topic 'Semantic Web'.

Section 1: Take part in a group dialog considering the following democratic habits: free-risk expression**,** encouragement, collaboration and gentle manners.

Section 2: Answer the questions below and follow these instructions:
- Exchange information related to the underlying topic
- Connect ideas in relation to this information
- FIRST find an agreement about each answer collaboratively, THEN write the answer by each group member individually

Trigger questions (examples below with meta-cognitive function elicited):
- What does a Triple define? (Conceptualizing)
- How a Triple is composed of? (Reasoning)
- What is Linked Data? (Conceptualizing)
- What does a RDF File contain? (Reasoning)
- What does RDF identify by using XML namespace? (Reasoning)

The first section explains the social nature of the inquiry activity while the second outlines the cognitive process involved in answering the trigger questions. Trigger questions are adapted from the work of Satiro (2006) [44]. As shown in Table 2, they are aimed at developing cognitive skills of conceptualization by comparing and contrasting, defining, classifying, and reasoning by relating cause and effect, tools and aims, parts and whole and by establishing criteria [44].

MCQs and trigger questions were ratified by lecturers to guarantee the preservation of the semantic behind each question and answer. All selected topics were supposed to be of high difficulty to justify the need of the collaborative activity. The students associated to the classes of selected lecturers were informed on the criteria of voluntary acceptance of the experiment and anonymity of any published data. In detail, to guarantee good standards of ethical research and scholarly practice, study information and consent

Table 2. Examples of trigger cognitive questions employed during the inquiry activity for the topics "Research Hypothesis", 'Problem Solving' and 'Operating System'.

Topic	Trigger question	Goal
Research hypothesis	What is a 'research hypothesis?	Conceptualizing
Research hypothesis	Which are the criteria you must consider in order to test a research hypothesis?	Reasoning
Problem solving	What is the 'lateral thinking'?	Conceptualizing
Problem solving	Which is the right set of actions to analyze 'facts and logic'?	Reasoning
Operating systems	Can you define the name of each section of the following full filename?	Conceptualizing
Operating systems	What is the function of the forc() command in Linux?	Reasoning

form were distributed at the beginning of each class and these were previously approved by the Ethical Committee of Technological University Dublin under the criteria of the European Code of Conduct for Research Integrity and the IUA Policy Statement on Ensuring Research Integrity in Ireland. Fidelity to this ethical protocol was ensured by the first author of this paper who gathered data. After signing study information and consent forms, the students who accepted to participate in the study were randomly divided in two groups: control and experimental groups. Both received direct instructions (instructional condition 1) while only the experimental group subsequently participated in the inquiry activity (instructional condition 2). The control group received the Rating Scale of Mental Effort and the NASA Task Load aimed at quantifying their self-reported effort and overall cognitive load respectively, as well as a multiple-choice questionnaire (MCQ) associated to the topic they were taught. The experimental group was in turn split in teams of three or four students for performing the inquiry activity. Subsequently, guidelines instructed each member of each team to write down the shared answer individually. This is a strategy designed to elicit metacognition and facilitate the process and transfer of information in working memory. After that, each student in the experimental group received the questionnaire aimed at quantifying their self- reported effort and mental workload, similarly to the students in the control group. Moreover, in order to make the relation between the outcomes of the inquiring process (the written answers to the trigger questions) and the achievement of knowledge as explicit as possible, the students in each team were allowed to use the answers, agreed during the inquiry activity, while answering the MCQ. This was assumed to be an advantage if the inquiring activity would have produced the right answer. However, if the shared answers were wrong, the extra-support provided by the inquiry activity was assumed to be a disadvantage. The contents under evaluation were exactly the same for both groups, consequently, the inquiring trigger questions were identical to those provided in the MCQ. The aim was to evaluate the impact of the inquiry technique on the achievement of knowledge and comparing it with the impact of direct instructions only. However, the way answers

were required was different. The MCQ showed the right answer that was selectable out of four options. On the contrary, during the inquiry activity, the answers to the trigger questions were supposed to be elaborated, constructed among the member of each team by reaching an agreement. Eventually, two slightly different designs of the described experimental study were planned, and these differ in the length of each questionnaire and in the type of measure of efficiency: training efficiency in the first was extended with learning efficiency in the second. These two experimental designs are detailed in the following sections as 'tuning' and 'experimental' phase respectively.

3.1   Tuning Phase: Participants and Procedures

Four lecturers and four different topics were selected in the first semester of the academic year 2018/19 involving a total of 122 students aged between 20 and 25 for bachelor and between 25 to 60 for masters. Table 3 provides details on these topics, academic stage, participants, number of slides and delivery length for the instructional condition 1 and 2.

Table 3. Description of each taught topic and associated information for the tuning phase.

Topic	Level	# participants	Control group	Experimental group	# slides	Length (mins)
Advanced database	B.Sc. (4th year)	25	13	12	28	50
Research Methods	M.Sc.	26	12	14	20	35
Amazon Cloud Watch	B.Sc. (3th year)	29	15	14	25	25
Semantic Web	B.Sc. (1th year)	42	26	16	55	75

The first design involved the use of the Rating Scale Mental Effort (RSME) and the original Nasa Task Load Index (NASA-TLX) respectively as measures of self-reported effort and cognitive load. The likelihood model of efficiency, as described in Sect. 2, was used with these two measures and the performance scores from the MCQs.

As shown in Fig. 1, students in the experimental group took part in a collaborative and inquiry activity. With a measure of performance (the multiple-choice score in percentage) and the overall cognitive load scores computed with the RSME and NASA-TLX (answered by students before the performance test) instruments, a measure of *training efficiency,* where the overall cognitive load was measured before the MCQ [53], was calculated using the likelihood model described in Sect. 2.

3.2   Experimental Phase: Participants and Procedures

Six lecturers and six different topics were delivered during the second semester of the academic year 2018/2019 involving a total of 160 students aged between 20 and 25 for
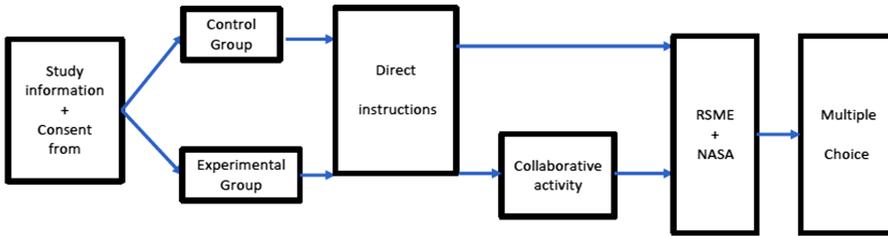
Fig. 1. The layout of the first set of experiment aimed at comparing the efficiency of two instructional conditions, one employing a direct instructional approach, and one extending this with a collaborative activity.

bachelor and between 25 to 60 for masters. Table 4 provides details on these topics: academic stage, participants, number of slides and delivery length for the instructional conditions 1 and 2. As depicted in Fig. 2, the Experimental phase was slightly different from the Tuning phase.

Table 4. Description of each taught topic and associated information for the experimental phase (C = control group, E = experimental group).

Topic	Level	# students	C	E	# slides	Length (mins)
Research Methods	M.Sc.	29	14	15	30	27
Research Hypothesis	M.Sc.	36	20	16	21	25
Geo Spatial Data	M.Sc.	12	5	7	60	40
Operating System	B.Sc. (4th Year)	39	20	19	142	60
Problem Solving	M.S.C	25	14	11	70	90
Data Mining	B.Sc. (4th Year)	19	10	9	62	60
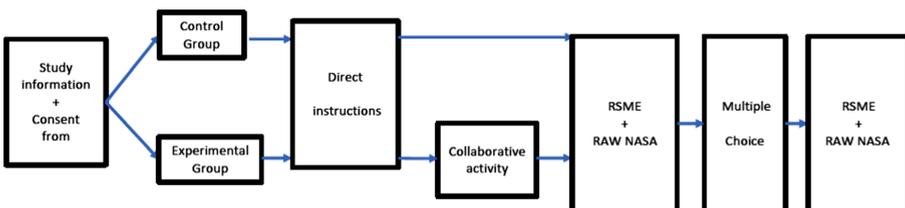
Fig. 2. A refined layout of the second set of experiments aimed at comparing the efficiency of two instructional conditions, one employing a direct instructional approach, and one extending this with a collaborative activity.

In detail, the pair comparison of the NASA-TLX instrument (15 questions) was eliminated in favor of the RAW-NASA-TLX. This was possible because it has been

empirically demonstrated that even by removing this pairwise comparison, the computation of the subjective mental workload experienced by students can be still considered valid [54]. A further difference showed in Fig. 2, is that students were asked to fill the questionnaire associated to the self-reported overall effort and cognitive load not only before the MCQ but also after. This change was introduced to verify the impact of the compared instructional conditions along different stages of the learning process. As suggested by Van Gogh and Paas (2008), in fact, filling questionnaires on cognitive load before and after the performance test corresponds to two different measures: training efficiency the former and learning efficiency the latter [53].

4   Results and Discussion

The scatterplots depicted in Fig. 3 show the overall relations between test performance (MCQ), the overall cognitive load (Nasa Task Load Index) and mental effort (RSME) respectively pre and post MCQ. The line in blue represents the linear regression of these two measures. As noticeable, in most of the cases, the overall increment of mental workload and effort does not affect the performance test as measured by multiple choice questionnaires. The latter seems to be independent from the amount of mental workload and effort experienced during the deliveries of instructional materials. This suggests these measures of load and performance are independent and their combination might deliver more insight. Consequently, this fully justifies the use of the selected model of efficiency. Table 5 shows means and standard deviations of RSME, NASA-TLX and MCQ associated to each topic and related group. All the experimental groups experienced, on average, more overall effort (RSME) and more overall cognitive load (NASA) than the control group. Intuitively, this can be attributed to the extra mental cost required by collaboration. As noticeable in Table 5, the collaborative activity increased also the overall level of performance of learners belonging to the experimental group across all experiments. To verify the normality of the distribution of the data a Shapiro Wilk test was computed followed by a T-Test for normal distributions (p-values > 0.05) and a Mann Whitney U Test (M.W.T) for not normal distributions (p-values < 0.05) to compare the means of control and experimental groups. Outliers were computed and eliminated. Table 6 shows the results of the related T-Test and Mann Whitney U Test.

Higher values of T-Value, in connection to the number of participants in each group, indicate that a large difference exists between data related to control and experimental groups. Higher values of M-Value, instead, indicate that the difference between groups is due to the experimental intervention rather than to chance. However, any statistical difference between groups is given by the P-Value.

Despite all experimental groups performed higher than the control groups (Table 5), the P-Value of T-Test shown in Table 6 is statistically relevant in the MCQ of Semantic Web only. This finding confirms the research hypothesis proposed: the inquiry activity statistically increases the performance of related group in Semantic Web. Unfortunately, in relation to the others MCQ, the research hypothesis is not supported by the same evidence, the related T-Test and Mann Whitney U Test, in fact, are all above the significance level. Given the dynamics of third-level classes and the heterogeneity of students having different characteristics such as prior knowledge and learning strategy, this was
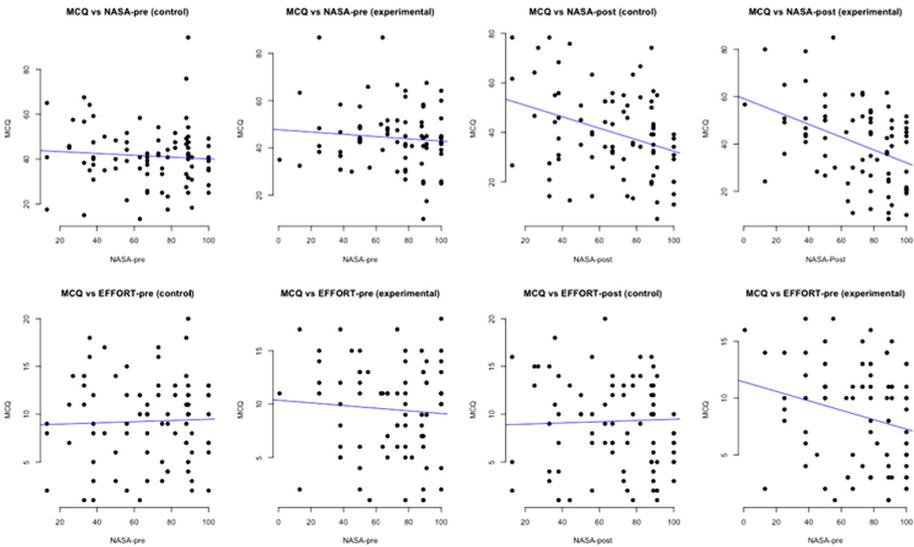
Fig. 3. Overall relations in control and experimental groups between mental workload, mental effort (measured before and after the MCQ) and test performance. The linear regression is represented by the line in blue (Color figure online).

Table 5. Means and standard deviations of the RSME, NASA-TLX and MCQ grouped by control and experimental groups (C = control group, E = experimental group)

Topic	RSME mean (SD)		NASA mean (SD)		MCQ mean (SD)	
	C	E	C	E	C	E
Advanced Database	36.00 (12.83)	47.91 (13.72)	43.61 (15.39)	47.80 (9.91)	42.92 (21.26)	54.91 (14.27)
Amazon Cloud Watch	56.00 (25.64)	68.57 (32.07)	50.00 (8.19)	54.45 (16.12)	61.33 (15.52)	66.42 (13.36)
Research Methods	47.08 (8.38)	67.85 (23.67)	49.38 (9.37)	49.85 (8.96)	68.41 (15.72)	69.57 (18.88)
Semantic Web	61.92 (29.19)	66.31 (32.08)	47.74 (10.98)	50.62 (9.43)	34.42 (18.10)	47.12 (18.77)

not a surprising outcome. Moreover, all experiments were conducted in real educational environments, consequently the collection of related data might have been affected by the 'noise' that characterizes the composition of a group of learners. In other words, external factors as the 'background noise', might have partially influenced the results. As per Table 6, the perceived effort of the experimental group in Research Methods statistically increases. Probably, the experimental group experienced more effort than the control group as a consequence of the inquiry activity whose impact presumably

Table 6. Values (V) of T-Test (T) or Mann Whitney U Test (M) and related P-values (P-V) (at significance level < 0.05) of the multiple choice percentage scores (MCQ), perceived effort scores (RSME) and the overall cognitive load scores (NASA-TLX).

Topic	MCQ		RSME		NASA	
	V	P-V	V	P-V	V	P-V
Advanced Dataset	53.5 (M)	0.18 (M)	42.5 (M)	0.52 (M)	−4.38 (T)	0.39 (T)
Amaz. Cloud Watch	85 (M)	0.4 (M)	−12.5 (T)	0.25 (T)	−4.58 (T)	0.15 (T)
Research Methods	−1.15 (T)	0.86 (T)	38 (M)	**0.01** (M)	−2.25 (T)	0.5 (T)
Semantic Web	−12.7 (T)	**0.03** (T)	203 (M)	0.9 (M)	−2.88 (T)	0.39 (T)

augmented the transactive cost of communication [51]. Unfortunately, in regard to the related measure of efficiency depicted in Table 7, this extra cost of communication (measured as higher effort) did not have a positive effect: the experimental group in Research Methods class, in fact, performed lower than the control group. Table 7 lists the efficiency scores across groups and topics computed with RSME and with NASA TLX employing the likelihood formula. The efficiency computed with the RSME is higher in the experimental group of Semantic Web only, whereas the efficiencies of the experimental groups in Advanced Database, Amazon Cloud Watch and Research Methods are lower than the control groups. A more coherent picture emerges when, as represented in Table 7, the efficiency is computed with the NASA-TLX. In fact, the efficiency scores are, on average, always slightly higher in the experimental group.

Table 7. Means and standard deviations of the efficiencies scores computed with the likelihood formula by using the Rating Scale Mental Effort (RSME) and the Nasa Task Load Index (NASA TLX) (C = control group, E = experimental group).

Topic	RSME mean (SD)		NASA-TLX mean (SD)	
	C	E	C	E
Advanced Database	1.3 (0.77)	1.23 (0.51)	1.15 (0.79)	1.24 (0.65)
Amazon Cloud Watch	1.4 (0.98)	1.33 (1.01)	1.24 (0.29)	1.34 (0.98)
Research Methods	1.50 (0.45)	1.15 (0.52)	1.42 (0.38)	1.44 (0.48)
Semantic Web	0.65 (0.41)	0.92 (0.6)	0.7 (0.38)	0.96 (0.44)

Again, a Shapiro Wilk test was computed to verify the normality of the distribution of the data followed by a T-Test for normal distributions and a Mann Whitney U Test (M.W.T) for not normal distributions. Despite of the average increment of the efficiency scores summarized in Table 7, these, as noticeable in Table 8, are not statistically different across design conditions with the NASA-TLX. All p-values, in fact, are greater than the significance level (0.05).

Table 8. Values (V) of T-Test (T) or Mann Whitney U Test (M) and related P-values (P-V) of the analysis of variance of the efficiency scores with the NASA Task Load Index (NASA-TLX) and the Rating Scale Mental Effort (RSME).

Topic	NASA-TLX		RSME	
	V	P-V	V	P-V
Advanced Database	68	0.61 (M)	0.06	0.8 (T)
Amazon Cloud Watch	96	0.71 (M)	86	0.42 (M)
Research Methods	−0.015	0.93 (T)	0.35	0.08 (T)
Semantic Web	−0.21	0.1 (T)	−0.31	**0.042** (T)

The evidence of the positive impact of the collaborative inquiry activity is limited to Semantic Web class. Here, higher efficiency of experimental group is statistically supported by the P-Value of T-Test. Table 7 shows that measuring the cognitive load by unidimensional or multidimensional instruments have an impact regardless of the model of efficiency employed. In fact, with the unidimensional RSME, the experimental group, on average, had a lower efficiency than the control group across topics (this is not true for Semantic Web). On the other hand, with the multidimensional NASA-TLX, the efficiency of the control group was always somewhat better than the experimental across topics. In contrast, the efficiency of the experimental group computed with the RSME in Semantic Web was higher than the control group and the difference was statistically significant as showed by T-Test in Table 8. Moreover, the layout of the inquiry activity boosted the related performance (MCQ) in all classes (Table 5). According to collaborative cognitive load theory, nine principles, can be used to predict the impact of collaborative activities on related performance and cognitive load of learners [51]. Among these, task complexity is particularly relevant to justify or not the implementation of collaborative activities. In relation to the current set of experiments, three factors were observed to infer task complexity: amount of content delivered, time employed for its delivery and level of prior knowledge of learners. Indeed, where the level of complexity overcomes working memories limits, collaboration is critical to share information and to free memory resources up [55]. According to this, Advanced Database was delivered in 50 min by 28 slides, Amazon Cloud Watch for 25 min and 25 slides were employed, Research Methods in 35 min by 20 slides, Semantic Web in 75 min by 55 slides. Prior knowledge could be inferred from the year the topic was delivered: Semantic Web first year BSc in Computer Science; Amazon Cloud Watching third year, Advanced Database fourth year, and Research Methods post-graduate level. Semantic Web was the learning task with the higher level of complexity in terms of number of slides (55), delivery time (75 min) and prior knowledge (first year). Results are in line with the assumption of collaborative cognitive load theory: collaborative learning is more effective when the level of the complexity of an instructional design is high [56]. In fact, on one hand Advanced Database, Amazon Cloud Watching, and Research Methods are of lower complexity to justify the utility of a collaborative activity that involves sharing working memory resources among different learners. On the other, the higher complexity of

Semantic Web justifies the utility of collaborative activities and the exploitation of extra memory resources among different learners in processing information and enhancing the learning outcomes [57]. In relation to the second set of the implemented experiments along the second semester 2018/19, Table 9 shows means and standards deviations of MCQ scores for each associated group and topic.

Table 9. Means and standard deviations of the multiple-choice percentage scores grouped by taught topic (C = control group, E = experimental group)

Topic	C mean (SD)	E mean (SD)
Research Methods	71.50 (22.12)	75.33 (14.36)
Research Hypothesis	82.35 (17.37)	89.00 (13.91)
Geo Spatial Data	45.40 (20.88)	44.94 (25.78)
Operating System	65.50 (22.04)	84.36 (11.77)
Problem Solving	76.21 (24.17)	54.81 (25.26)
Data Mining	37.80 (14.43)	32.22 (11.04)

As observable in Table 9, the performance in the MCQ of the experimental group in Research Methods, Research Hypothesis, and Operating System is, on average, higher than the control group but lower in Problem Solving, Visualizing Geospatial Data and Data Mining. Table 10 displays means and standard deviations of the perceived mental effort pre and post MCQ associated to each group and topic. The experimental group in Research Methods, Visualizing Geo Spatial Data and Problem Solving experienced, on average, more effort than the control group when the effort was measured before of the MCQ. Intuitively this can be attributed to the extra mental cost required by the communication developed within the collaborative activity. This is not true for Research Hypothesis, Operating System and Data Mining where the effort of the experimental groups is lower. Moreover, in Research Methods, Research Hypothesis and Problem Solving the effort post MCQ of the control groups is higher than the experimental but lower in Visualizing Geo Spatial Data Operating system and Data Mining.

The assumption of sharing working memories is valid for the analysis of the mental workload also. As displayed in Table 11, the experimental groups experienced higher overall cognitive load (measured before the MCQ) than the control groups in all classes except of Research Hypothesis. The perceived overall cognitive load of the control groups (measured after the MCQ) is higher in Research Methods, Research Hypothesis and Operating System but lower in Visualizing Geo Spatial Data, Problem Solving and Data Mining. In Research Hypothesis only, the experimental groups experienced, over all variables, less effort and cognitive load than the control groups (Table 10 and Table 11). Results in Research Hypothesis may be interpreted as a consequence of the inquiry activity that freed working memories up and optimized cognitive load by sharing mental resources among multiple working memories. To test the statistical relevance of results depicted in Tables 9, 10 and 11, a Shapiro Wilk test was performed to verify the

Table 10. Means and standard deviations of the Rating Scale Mental Effort (RSME) scores PRE and POST execution of the multiple-choice questionnaire (C = control group, E = experimental group)

Topic	Pre mean (SD)		Post mean (SD)	
	C	E	C	E
Research Methods	49.35 (29.13)	56.53 (35.31)	51.21 (26.38)	50.13 (26.61)
Research Hypothesis	55.00 (26.73)	37.20 (22.30)	60.60 (28.91)	41.69 (25.99)
Geo Spatial Data	39.80 (35.95)	47.85 (22.51)	53.80 (37.69)	79.57 (24.69)
Operating System	45.90 (32.38)	36.16 (26.27)	31.50 (20.65)	37.42 (34.33)
Problem Solving	39.35 (22.75)	48.72 (25.90)	50.92 (27.39)	47.18 (22.55)
Data Mining	54.10 (30.07)	38.55 (23.59)	56.30 (34.39)	82.55 (26.81)

normality of the distribution of the data followed by a T-Test for normal distributions and a Mann Whitney U Test (M.W.T) for not normal distributions to compare the means of control and experimental groups. Outliers were spotted and eliminated

Table 12 shows P-Values statistically relevant in the RSME pre-MCQ and post-MCQ of Research Hypothesis and in the RSME post-MCQ of Data Mining. As shown in Table 10, the control group in Research Hypothesis experienced, on average, more effort than the experimental that, instead, perceived less effort, likely because of sharing working memory [58]. In Data Mining instead, the experimental group perceived more effort likely as a consequence of the collaboration activity that increased the communicative costs on the working memories of multiple learners [51]. Table 12 depicts also two statistically significant P-Values in the MCQ of Operating System and Problem Solving.

Table 11. Means and standard deviations of the mental workload scores computed with the RAW Nasa Task Load Index (C = control group, E = experimental group)

Topic	PRE mean (SD)		POST mean(SD)	
	C	E	C	E
Research Methods	43.86 (11.66)	47.50 (17.73)	41.36 (20.68)	41.00 (19.07)
Research Hypothesis	44.66 (13.24)	41.45 (08.9)	40.25 (15.39)	37.05 (16.28)
Geo Spatial Data	35.5 (19.50)	46.19 (09.29)	42.00 (11.17)	47.85 (05.94)
Operating System	36.7 (13.62)	42.98 (15.24)	32.29 (16.52)	29.29 (14.88)
Problem Solving	42.2 (15.07)	43.86 (06.54)	40.00 (17.65)	42.65 (12.38)
Data Mining	42.66 (7.56)	46.38 (18.00)	50.58 (16.46)	58.14 (15.60)

Table 12. Values (V) of T-Test (T) or Mann Whitney U Test (M) and related P-values (P-V) of the multiple choice percentage scores (MCQ), the Rating Scale Mental Effort scores (RSME) and the mental workload scores computed with the Raw Nasa Task Load Index (RAW NASA-TLX) using a two tails distribution and two-sample equal variance

Topic	RSME				MCQ		RAW NASA-TLX			
	Pre		Post				Pre		Post	
	V	P-V	V	P-V	V	P-V	V	P-V	V	P-V
Research Methods	−0.5 (T)	0.961 (T)	1 (M)	0.91 (T)	104 (M)	1 (M)	−3.6 (T)	0.52 (T)	98 (M)	0.78 (M)
Research Hypothesis	21.2 (T)	**0.017** (T)	18.9 (T)	0.04 (T)	120 (M)	0.21 (M)	154 (M)	0.86 (M)	158 (M)	0.96 (M)
G. Spatial Data	9.5 (M)	0.2 (M)	-25 (T)	0.18 (T)	0.45 (T)	0.97 (T)	−10 (T)	0.23 (T)	−5 (T)	0.26 (T)
Operating System	155 (M)	0.47 (M)	175 (M)	0.89(M)	87 (M)	**0.005** (M)	−6.2 (T)	0.18 (T)	161 (M)	0.59 (M)
Problem Solving	−9.3 (T)	0.346 (T)	3.7 (T)	0.71 (T)	40 (M)	**0.04** (M)	−4.2 (T)	0.3 (T)	−2 (T)	0.67 (T)
Data Mining	26.5 (M)	0.133 (M)	−33 (T)	**0.02** (T)	5.5 (T)	0.362 (T)	42 (M)	0.84 (M)	−7 (T)	0.32 (T)

In Operating System the experimental group, as per Table 9, performed higher than the control group. In contrast, in Problem Solving the experimental group performed lower. To give a more precise account of these opposite results, a qualitative interpretation of the inquiry activity is proposed as following: the design of the trigger questions related to Operating System might have implied a technical discussion on how to use the commands functions in Linux. In contrast, the design of the questions related to Problem Solving might have generated a dialogue more complex whose abstract and theoretical nature, based on the six different way of thinking proposed by De Bono 2017 [59], might have hampered the comprehension of the delivery. Table 13 and Table 14 show that the efficiencies scores of the experimental groups, computed pre and post MCQ with RSME and RAW NASA, are always higher in Research Hypothesis and Operating System. On the contrary, the control groups in Visualizing Geo Spatial Data, Problem Solving and Data Mining performed always better. This is valid for Research Methods too, a part of the efficiency pre-MCQ computed with the RAW NASA that is equal for both groups.

Again, a Shapiro Wilk test was computed to test the normality of the distribution of the data followed by a T-Test for normal distributions and a Mann Whitney U Test for not normal distributions to compare the means of control and experimental groups. Outliers were spotted and eliminated. Table 15 points out P-Values statistically significant in the efficiency of Research Hypothesis class considering the RSME before and after the MCQ. Here, the distribution of the efficiency in the experimental groups, shown in Table 13, is higher than the control groups. Unfortunately, despite all members of

Table 13. Means and standard deviations of the efficiency scores computed with the likelihood model using the Rating Scale Mental Effort (RSME) grouped by taught topics, control and experimental groups, pre and post multiple choice (C = control group, E = experimental group)

Topic	PRE mean (SD)		POST mean (SD)	
	C	E	C	E
Research Methods	6.84 (19.0)	2.24 (2.19)	7.96 (23.91	5.91 (16.1)
Research Hypothesis	2.20 (1.96)	14.5 (33.3)	2.34 (2.56)	12.44 (28.4)
Geo Spatial Data	1.55 (1.02)	1.13 (0.82)	1.09 (0.82)	0.68 (0.53)
Operating System	2.54 (2.24)	10.7 (19.5)	7.87 (16.8)	10.8 (21.3)
Problem Solving	7.57 (0.92)	1.54 (1.05)	2.10 (1.60)	1.36 (0.78)
Data Mining	1.09 (1.05)	1.04 (0.64)	1.22 (1.11)	0.44 (0.23)

Table 14. Means and standard deviations of the efficiency scores computed with the likelihood model using the Raw Nasa Task Load Index (NASA-TLX), grouped by taught topics, control and experimental groups, pre/post multiple choice (C = control group, E = experimental group)

Topic	PRE mean (SD)		POST mean (SD)	
	C	E	C	E
Research Methods	1.86 (1.11)	1.86 (1.05)	3.35 (4.66)	2.31 (1.39)
Research Hypothesis	1.96 (0.67)	2.28 (0.77)	2.58 (1.87)	3.13 (2.26)
Geo Spatial Data	1.70 (1.69)	0.95 (0.54)	1.06 (0.50)	0.98 (0.60)
Operating System	2.08 (1.04)	2.40 (1.71)	2.53 (1.52)	3.92 (2.51)
Problem Solving	2.04 (0.91)	1.25 (0.64)	2.51 (1.66)	1.33 (0.67)
Data Mining	0.92 (0.41)	0.80 (0.39)	0.86 (0.44)	0.59 (0.24)

experimental groups knew what they had to do (take part in a dialogue), how (democratically social setting structured by cognitive phases), with whom they had to work with and what they had to communicate about (answering cognitive trigger questions), these findings are the only evidences emerged within this set of studies in favor of the proposed research hypothesis.

Table 15 represents statistically significant P-Values in the efficiency of Problem Solving either computed before and after the MCQ with RAW-NASA-TLX. Here, the average mean in the efficiency of the control group, is higher than the experimental (Table 14). Consequently, it can be deducted that the impact of the experimental design on the efficiencies of learners was negative. The direct instructions whereby the topic was delivered, indeed, were supposed to be enough to inform the students. In contrast, the design of the inquiry activity might have created a redundant effect on the comprehension of the delivery, decreasing the efficiency in experimental group instead of

Table 15. Values (V) of T-Test (T) or Mann Whitney U Test (M) and related P-values (P-V) of the efficiency scores computed with the likelihood model grouped by pre and post multiple choice questionnaire, the mental workload instruments (Rating Scale Mental Effort or Raw Nasa Task Load Index) with 2 tailed distribution and two sample equal variance.

Topic	RSME				RAW-NASA-TLX			
	PRE	PRE	POST	POST	PRE	PRE	POST	POST
	V	P-V	V	P-V	V	P-V	V	P-V
Research Methods	97.5 (M)	0.74 (M)	90.5 (M)	0.53 (M)	92.5 (M)	0.591 (M)	98 (M)	0.78 (M)
Research Hypothesis	97 (M)	**0.04** (M)	89.5 (M)	**0.02** (M)	138.5 (M)	0.49 (M)	142 (M)	0.58 (M)
G. Spatial Data	0.41 (T)	0.45 (T)	0.48 (T)	0.24 (T)	14 (M)	0.63 (M)	−0.11 (T)	0.77 (T)
Operating System	138 (M)	0.22 (M)	143 (M)	0.29 (M)	177 (M)	0.94 (M)	122 (M)	0.09 (M)
Problem Solving	52 (M)	0.18 (M)	54 (M)	0.22 (M)	0.93 (T)	**0.02** (T)	1.17 (T)	**0.03** (T)
Data Mining	40 (M)	0.72 (M)	22 (M)	0.06 (M)	0.11 (T)	0.53 (T)	0.26 (T)	0.13 (T)

improving it. Considering the cost of communication experienced answering the trigger questions, the inquiry activity does have an influence on the efficiency of learners. In order to better define the relation between 'communicative cost' of inquiring and complexity of the delivery, it is worthy accounting for the design of the inquiring activity in Research Hypothesis and Problem Solving, whose respective trigger questions were more technical in the former and more theoretical in the latter. To sum up, the cost of communication in the experimental design of Problem Solving probably increased the element interactivity of the learning task (its difficulty), generating extraneous instead of germane load. In other words, the extra cost of communication experienced during the inquiring activity of Problem Solving might have confused the experimental group instead of benefiting it as for the Research Hypothesis topic. Table 16 shows that the effect sizes computed on T-Test of the efficiency and MCQ scores in Semantic Web topic are medium under the criteria proposed by Cohen (2013) but in the zone of desired effects as per the barometer of effect sizes proposed in Fig. 4 [60].

Based on the criteria proposed by Cohen (2013), Table 16 shows large and medium effect sizes computed for Mann Whitney U Tests [61]. As observable per Table 16, the evidence against the hypothesis are confirmed by the large effect size found in the MCQ of Problem Solving and in its efficiency pre and post MCQ with RAW-NASA-TLX. Nonetheless, the evidences in favor of the hypothesis are strongly confirmed as well because the effect size for efficiency and MCQ in Semantic Web is medium under the Cohen's criteria but within the zone of desired effects as per Hattie's barometer. Moreover, Table 16 shows that the effect size of efficiency pre and post MCQ in Research

Table 16. Effect sizes grouped by taught topic, instrument and statistical T-test (T) or Mann Whitney U Test (M) and their sizes (medium ≥ 0.5 and <0.8; large ≥ 0.8 for T-test) (large ≥ 0.5 for Mahan Whitney U Test) [61].

Topic	Instrument (details)	Test	Effect	Size
Res. Methods	RSME	M	Large	0.718
Semantic Web	Efficiency Likelihood Model	T	Medium	0.542
Semantic Web	MCQ	T	Medium	0.691
Data Mining	RSME (post MCQ)	T	Large	0.845
Problem Solving	Efficiency Likelihood Model (RAW NASA pre-MCQ)	T	Large	0.982
Problem Solving	Efficiency Likelihood Model (RAW NASA post-MCQ)	T	Large	0.891
Research Hypothesis	Effort (pre MCQ)	T	Medium	0.712
Research Hypothesis	Effort (post MCQ)	T	Medium	0.618
Problem Solving	MCQ	M	Large	0.71
Operating System	MCQ	M	Large	0.73
Research Hypothesis	Efficiency Likelihood Model (RSME Pre-MCQ)	M	Large	0.72
Research Hypothesis	Efficiency Likelihood Model (RSME Post-MCQ)	M	Large	0.721
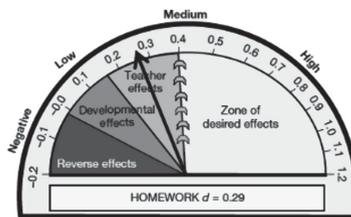
Fig. 4. A typical barometer of influence. This aims to judge the success of educational innovations relative to a hinge point, a zone of desired effects empirically found at d = 0.4 from which influences start to have the greatest impact on student achievement outcomes [60].

Hypothesis is large: here the statistical results confirmed that the experimental design increased the efficiencies of learners. This tendency is further supported by the large effect size found in the MCQ of Operating System and by the medium effect size found in the perceived effort pre and post MCQ in Research Hypothesis where the inquiry activity statistically decreased the perceived effort optimizing the efficiencies of learners.

Very often, the constructivist community of inquiry method was employed in contexts of learning where the process for forming knowledge is ill structured, as for instance in topics whose goal is to reach an agreement on ethical issues by learners and construct knowledge collaboratively. On the contrary, direct instruction methods were employed

in context of learning where the process for forming knowledge is well structured. Their respective research approaches to evaluate learning outcomes, in fact, are different: mainly quantitative the former and mainly qualitative the latter. Their respective learning outcomes are different too. Consequently, it was critical to establish shared learning outcomes that have been identified in factual, conceptual and procedural knowledge for both approaches. The main limitation of the studies under examination was that the design of the deliveries and the related element interactivity, were not under the direct control of the researchers. However, the conditions of the delivery were identical for both the control and experimental groups, with the only difference that the latter performed a subsequent inquiry activity. The comparison of efficiency, in fact, occurred within members of groups of the same class who received the same instructions in a number of controlled experiments. Results showed that changes in the independent variable (the instructional design) partially affected the effort and cognitive load experienced by students, their performance and the related efficiency of learners. The analysis of variance computed on the average of these variables per groups was not statistical different either. By a deeper analysis of data, it emerged that the trigger questions are the core structure of the experimental design and are supposed to generate a dialogue whose transactive cost of communication can increase or decrease the efficiency of learners.

5   Conclusion

A literature review revealed a lack of comparison of the efficiency of direct instruction and inquiry teaching methods. Motivated by the statement provided by Kirshner and colleagues (2006) whereby inquiries techniques are believed to be ineffective in the absence of explicit direct instructions, an empirical experimental study has been designed. In detail, a comparison of the efficiency of the traditional instructional design, based upon explicit direct instructions, and its extension with a highly guided inquiry activity, was proposed. The research hypothesis was that if the traditional explicit instructional method is extended with an inquiry technique, then, its efficiency is higher than by employing the former method alone. Efficiency was measured by employing the likelihood model of efficiency [3]. The original models are based upon a unidimensional measure of subjective perceived human effort and objective performance. In this study, these were extended with a multidimensional measure of cognitive load. In detail, the Rating Scale Mental Effort [37] has been selected as the unidimensional measure while the Nasa Task Load Index and its shorter version, the RAW Nasa [20], have been chosen as multi-dimensional measures of cognitive load. In relation to objective performance, multiple choice questionnaires were employed, tailored to each experimental study and the taught content to learners. A number of lecturers were involved in this study and their classes were split in control and experimental groups. The former received direct instructions, while the latter also participated in an inquiry activity based on trigger questions to support the development of cognitive skills of conceptualization and reasoning. Results showed partial evidence in favour of the hypothesis, supported by good effect sizes. However, a clear separation and a statistical significant difference of the efficiency between the groups of students who have received direct instructional method alone, and those who have received its extended version with the inquiry technique, is hard to be

achieved because of the small sample sizes of these groups in typical higher-education classes, usually in the order of ten/twenty students per group. Therefore, further statistical tests for small sample-size groups comparison should be explored such as cluster analysis and shift functions. Future work will focus on the design of a lighter collaborative inquiry activity to mitigate the extra-cost of communication among students and thus supporting a measurement of their cognitive load that better reflect their experience within this activity itself. The aim is to promote a dialogical environment, increase the germane load of students and minimize their extraneous load with expected positive consequences on instructional efficiency. This research contributes to the existent body of knowledge by demonstrating how the impact of the community of inquiry method can be empirically tested using existing measures of efficiency, cognitive load and performance. In doing so, it requalifies the constructivist approach of the community of inquiry technique by creating a replicable experiment for enhancing learning that extends the traditional method for teaching, based upon direct instructional designs only.

References

1. Kirschner, P.A., Sweller, J., Clark, R.E.: Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educ. Psychol. **41**(2), 75–86 (2006)
2. Jonassen, D.: Reconciling a human cognitive architecture. In: Constructivist Instruction, pp. 25–45. Routledge (2009)
3. Hoffman, B., Schraw, G.: Conceptions of efficiency: applications in learning and problem solving. Educ. Psychol. **45**(1), 1–14 (2010)
4. Sweller, J., van Merriënboer, J.J., Paas, F.: Cognitive architecture and instructional design: 20 years later. Educ. Psychol. Rev. **31**, 1–32 (2019)
5. Orru, G., Longo, L.: The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: a review. In: Longo, L., Leva, M.C. (eds.) H-WORKLOAD 2018. CCIS, vol. 1012, pp. 23–48. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14273-5_3
6. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**(2), 81 (1956)
7. Atkinson, R.C., Shiffrin, R.M.: The Control Processes of Short-Term Memory. Stanford University, Stanford (1971)
8. Sweller, J.: Working memory, long-term memory, and instructional design. J. Appl. Res. Mem. Cogn. **5**(4), 360–367 (2016)
9. Sweller, J.: Cognitive Load Theory, Evolutionary Educational Psychology, and Instructional Design. In: Geary, D.C.C., Berch, D.B.B. (eds.) Evolutionary Perspectives on Child Development and Education. EP, pp. 291–306. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29986-0_12
10. Orru, G., Longo, L.: Direct instruction and its extension with a community of inquiry: a comparison of mental workload, performance and efficiency. In: CSEDU (1), pp. 436–444 (2019)
11. Sloffer, S., Dueber, B., Duffy, T.M.: Using asynchronous conferencing to promote critical thinking: two implementations in higher education. In: HICSS, p. 1083. IEEE (1999)
12. Van Boxtel, C., Van der Linden, J., Kanselaar, G.: Collaborative learning tasks and the elaboration of conceptual knowledge. Learn. Instr. **10**(4), 311–330 (2000)

13. Gregor, S.D., Cuskelly, E.: Computer mediated communication in distance education. J. Comput. Assist. Learn. **10**(3), 168–181 (1994)
14. Heath, E.: Two cheers and a pint of worry: an on-line course in political and social philosophy. Teach. Philos. **20**(3), 277–300 (1997)
15. Paas, F., Sweller, J.: An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. Educ. Psychol. Rev. **24**(1), 27–45 (2012)
16. Kirschner, F., Paas, F., Kirschner, P.A.: Individual and group-based learning from complex cognitive tasks: effects on retention and transfer efficiency. Comput. Hum. Behav. **25**(2), 306–314 (2009)
17. Wickens, C.D.: Measures of workload, stress and secondary tasks. In: Moray, N. (ed.) Mental Workload. NATO Conference Series, vol. 8. Springer, Boston (1979). https://doi.org/10.1007/978-1-4757-0884-4_6
18. Longo, L.: A defeasible reasoning framework for human mental workload representation and assessment. Behav. Inf. Technol. **34**(8), 758–786 (2015)
19. Rizzo, L., Dondio, P., Delany, S.J., Longo, L.: Modeling mental workload via rule-based expert system: a comparison with NASA-TLX and workload profile. In: Iliadis, L., Maglogiannis, I. (eds.) AIAI 2016. IAICT, vol. 475, pp. 215–229. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44944-9_19
20. Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 904–908. Sage Publications, Sage (2006)
21. Longo, L., Orru, G.: An evaluation of the reliability, validity and sensitivity of three human mental workload measures under different instructional conditions in third-level education. In: McLaren, B.M., Reilly, R., Zvacek, S., Uhomoibhi, J. (eds.) CSEDU 2018. CCIS, vol. 1022, pp. 384–413. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21151-6_19
22. Longo, L.: Designing medical interactive systems via assessment of human mental workload. In: International Symposium on Computer-Based Medical Systems, pp. 364–365 (2015)
23. Longo, L., Barrett, S.: Cognitive effort for multi-agent systems. In: Yao, Y., Sun, R., Poggio, T., Liu, J., Zhong, N., Huang, J. (eds.) BI 2010. LNCS (LNAI), vol. 6334, pp. 55–66. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15314-3_6
24. Longo, L.: Formalising human mental workload as a defeasible computational concept. Ph.D. thesis. Trinity College Dublin (2014)
25. Longo, L.: Experienced mental workload, perception of usability, their interaction and impact on task performance. PLoS ONE **13**(8), e0199661 (2018)
26. Longo, L., Leva, M.C. (eds.): H-WORKLOAD 2018. CCIS, vol. 1012. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14273-5
27. Longo, L.: Human-computer interaction and human mental workload: assessing cognitive engagement in the world wide web. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011. LNCS, vol. 6949, pp. 402–405. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23768-3_43
28. Longo, L.: Subjective usability, mental workload assessments and their impact on objective human performance. In: Bernhaupt, R., Dalvi, G., Joshi, A., Balkrishan, D.K., O'Neill, J., Winckler, M. (eds.) INTERACT 2017. LNCS, vol. 10514, pp. 202–223. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67684-5_13
29. Rizzo, L., Longo, L.: Inferential models of mental workload with defeasible argumentation and non-monotonic fuzzy reasoning: a comparative study. In: AI^3 @ AI* IA, pp. 11–26 (2018)
30. Orru, G., Gobbo, F., O'Sullivan, D., Longo, L.: An investigation of the impact of a social constructivist teaching approach, based on trigger questions, through measures of mental workload and efficiency. In: 10th International Conference on Computer Supported Education (CSEDU 2018), pp. 292–302 (2018)

31. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
32. Rizzo, L.M., Longo, L.: Representing and inferring mental workload via defeasible reasoning: a comparison with the NASA task load index and the workload profile (2017)
33. Longo, L.: On the reliability, validity and sensitivity of three mental workload assessment techniques for the evaluation of instructional designs: a case study in a third-level course. In: 10th International Conference on Computer Supported Education (CSEDU 2018), pp. 166–178 (2018)
34. Paas, F.G., Van Merriënboer, J.J.: The efficiency of instructional conditions: an approach to combine mental effort and performance measures. Hum. Factors J. Hum. Factors Ergon. Soc. **35**(4), 737–743 (1993)
35. Paas, F.G., Van Merriënboer, J., Adam, J.J.: Measurement of cognitive load in instructional research. Percept. Mot. Skills **79**(1), 419–430 (1994)
36. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.: Cognitive load measurement as a means to advance cognitive load theory. Educ. Psychol. **38**(1), 63–71 (2003)
37. Zijlstra, F.R.H.: Efficiency in work behavior: a design approach for modern tools. Doctoral thesis. Delft University Press (1993)
38. Gerjets, P., Scheiter, K., Catrambone, R.: Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? Learn. Instr. **16**(2), 104–121 (2006)
39. Kester, L., Lehnen, C., Van Gerven, P.W., Kirschner, P.A.: Just-in-time, schematic supportive information presentation during cognitive skill acquisition. Comput. Hum. Behav. **22**(1), 93–112 (2006)
40. Gerjets, P., Scheiter, K., Catrambone, R.: Designing instructional examples to reduce intrinsic cognitive load: molar versus modular presentation of solution procedures. Instr. Sci. **32**(1/2), 33–58 (2004)
41. Dewey, J.: Logic-The Theory of Inquiry. Read Books Ltd., Redditch (2018)
42. Garrison, D.R.: Online community of inquiry review: social, cognitive, and teaching presence issues. J. Asynchronous Learn. Netw. **11**(1), 61–72 (2007)
43. Lipman, M.: Thinking in Education. Cambridge University Press, Cambridge (2003)
44. Sátiro, A.: Jugar a pensar con mitos: este libro forma parte dle Proyecto Noria y acompña al libro para niños de 8-9 años: Juanita y los mitos. Octaedro (2006)
45. Krathwohl, D.R.: A revision of bloom's taxonomy: an overview. Theory Pract. **41**(4), 212–218 (2002)
46. Smith, P.C., Street, A.: Measuring the efficiency of public services: the limits of analysis. J. R. Stat. Soc. Ser. A Stat. Soc. **168**(2), 401–417 (2005)
47. Kalyuga, S., Sweller, J.: Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. Educ. Technol. Res. Dev. **53**(3), 83–93 (2005)
48. Warnick, E.M., Bracken, M.B., Kasl, S.: Screening efficiency of the child behavior checklist and strengths and difficulties questionnaire: a systematic review. Child Adolesc. Ment. Health **13**(3), 140–147 (2008)
49. Plass, J.L., Moreno, R., Brünken, R.: Cognitive Load Theory. Cambridge University Press, Cambridge (2010)
50. Haladyna, T.M., Downing, S.M., Rodriguez, M.C.: A review of multiple-choice item-writing guidelines for classroom assessment. Appl. Meas. Educ. **15**(3), 309–333 (2002)
51. Kirschner, P.A., Sweller, J., Kirschner, F., Zambrano R., J.: From cognitive load theory to collaborative cognitive load theory. Int. J. Comput.-Support. Collab. Learn. **13**(2), 213–233 (2018). https://doi.org/10.1007/s11412-018-9277-y
52. Popov, V., van Leeuwen, A., Buis, S.: Are you with me or not? Temporal synchronicity and transactivity during CSCL. J. Comput. Assist. Learn. **33**(5), 424–442 (2017)

53. Van Gog, T., Paas, F.: Instructional efficiency: revisiting the original construct in educational research. Educ. Psychol. **43**(1), 16–26 (2008)
54. Bittner Jr, A.C., Byers, J.C., Hill, S.G., Zaklad, A.L., Christ, R.E.: Generic workload ratings of a mobile air defense system (LOS-FH). In: Proceedings of the Human Factors Society Annual Meeting, pp. 1476–1480. SAGE Publications, Sage (1989)
55. Kirschner, F., Paas, F., Kirschner, P.A.: Task complexity as a driver for collaborative learning efficiency: the collective working-memory effect. Appl. Cogn. Psychol. **25**(4), 615–624 (2011)
56. Kirschner, P., Kirschner, F., Janssen, J.: The collaboration principle in multimedia learning. Camb. Handb. Multimed. Learn. **2**, 547–575 (2014)
57. Kirschner, F., Paas, F., Kirschner, P.A.: A cognitive load approach to collaborative learning: united brains for complex tasks. Educ. Psychol. Rev. **21**(1), 31–42 (2009)
58. Kirschner, F., Paas, F., Kirschner, P.A.: Superiority of collaborative learning with complex tasks: a research note on an alternative affective explanation. Comput. Hum. Behav. **27**(1), 53–57 (2011)
59. De Bono, E.: Six thinking hats. Penguin UK (2017)
60. Hattie, J.: Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement, p. 378. Routledge, London (2010)
61. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences. Routledge, Abingdon (2013)