

2021-11-23

## A Novel Parabolic Model of Instructional Efficiency Grounded on Ideal Mental Workload and Performance

Luca Longo

*Technological University Dublin, [luca.longo@tudublin.ie](mailto:luca.longo@tudublin.ie)*

Murali Rajendran

*Technological University Dublin*

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomart>



Part of the [Computer Sciences Commons](#), and the [Instructional Media Design Commons](#)

### Recommended Citation

Longo L., Rajendran M. (2021) A Novel Parabolic Model of Instructional Efficiency Grounded on Ideal Mental Workload and Performance. In: Longo L., Leva M.C. (eds) Human Mental Workload: Models and Applications. H-WORKLOAD 2021. Communications in Computer and Information Science, vol 1493. Springer, Cham. DOI: 10.1007/978-3-030-91408-0\_2

This Conference Paper is brought to you for free and open access by the School of Computer Sciences at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [gerard.connolly@tudublin.ie](mailto:gerard.connolly@tudublin.ie).



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](#)



# A Novel Parabolic Model of Instructional Efficiency Grounded on Ideal Mental Workload and Performance

Luca Longo<sup>(✉)</sup> and Murali Rajendran

Applied Intelligence Research Centre, School of Computer Science, City Campus,  
Technological University Dublin, Dublin, Ireland  
[luca.longo@tudublin.ie](mailto:luca.longo@tudublin.ie)

**Abstract.** Instructional efficiency within education is a measurable concept and models have been proposed to assess it. The main assumption behind these models is that efficiency is the capacity to achieve established goals at the minimal expense of resources. This article challenges this assumption by contributing to the body of Knowledge with a novel model that is grounded on ideal mental workload and performance, namely the *parabolic model* of instructional efficiency. A comparative empirical investigation has been constructed to demonstrate the potential of this model for instructional design evaluation. Evidence demonstrated that this model achieved a good concurrent validity with the well-known likelihood model of instructional efficiency, treated as baseline, but a better discriminant validity for the evaluation of the training and learning phases. Additionally, the inferences produced by this novel model have led to a superior information gain when compared to the baseline.

## 1 Introduction

The construct of efficiency, within the field of learning and instruction, is assumed to be the capacity to achieve established goals with minimal expenditure of effort or resources [12]. Models exist for assessing instructional efficiency, and they are based upon a measure of mental effort or workload exerted during a learning task, and a measure of test performance [36, 46, 47]. Ideally, any instructional activity conducted should be as efficient as possible and it is important to understand how particular approaches to learning influence the performance of learners. The above assumption that underpins efficiency is that low mental effort, with high performance, provides the best efficiency, whereas, high mental effort, with low performance, provides the worst efficiency. However, this article challenges this assumption by arguing that there are cases where it does not hold. Through the illustration of counter-examples, and by using the terminology brought forward in the years by Cognitive Load Theory, and its cognitive load types, this article proposes a novel model of instructional efficiency, named *the parabolic model*. This is inspired by the well-known assumption, within mental workload research, focused on the parabolic relationship between experienced mental workload and

performance, inspired by the seminal Yerkes-Dodson law, an inverted-U shaped model that links stress and performance [49].

The aim of this research is to propose and describe this novel model, by stating its main assumptions, contextualising it in the wider framework of instructional efficiency and cognitive load theory, as well as evaluating its effectiveness with a comparative empirical experiment. For these purposes, the reminded of this article focuses on presenting related work on cognitive load theory and on the construct of instructional efficiency, respectively in Sects. 2.1 and 2.2. The terminology, extracted from related work, is coupled with an introduction of the concepts of mental effort and mental workload in Sect. 2.3, serving as necessary information for discussing, in Sect. 2.4, the issues behind the main assumption of current instructional efficiency models. The article continues with the presentation of the novel *parabolic model* of instructional efficiency, in Sect. 3, along with its main assumption and formalities. Subsequently, a comparative quantitative empirical research experiment is designed and illustrated in Sect. 6, along with a research hypothesis aimed at testing this one of its kind model of instructional efficiency, with a set of evaluation criteria described in Sect. 3.2. Empirical resulting evidence is presented in Sect. 4 followed by a summary of this research, the implications for teaching and learning, and a delineation, in Sect. 5, of future work.

## 2 Related Work

### 2.1 Cognitive Load and Its Theory

Cognitive Load Theory (CLT) is a widely known theory in educational psychology which is used to enhance the learning phase by developing instructional material and applying instructional teaching techniques based on the limitations of the human cognitive architecture [33]. CLT provides an effective framework for designing and delivering work to learners of any standard. It is backed by empirical research supporting different amounts and types of instruction according to the level of learners and it enables instructors to provide well-crafted guidance in their topics. It states that effective learning can only take place where the cognitive capacity of an individual in a particular domain is not exceeded. The human cognitive architecture (HCA) provides a generic framework of the information-processing stages that learners use to encode, store, and modify information for the purposes of reasoning and decision making [1, 38]. It describes the necessary and sufficient conditions for a human to input, process and store information which in turn becomes knowledge. Cognitive load includes units of knowledge and elements of relationship. The cognitive load associated to a task is created when the units of knowledge interact with the relationship elements [44].

Sensory memory, short-term memory, also known as working memory, and long-term memory are three essential dimensions of HCA. Atkinson and Shiffrin (1968) proposed that the input of the information entered via sensory memory is processed in the working memory and then proceeds to be stored in the long-term memory [1]. Working memory is limited and it processes incoming information

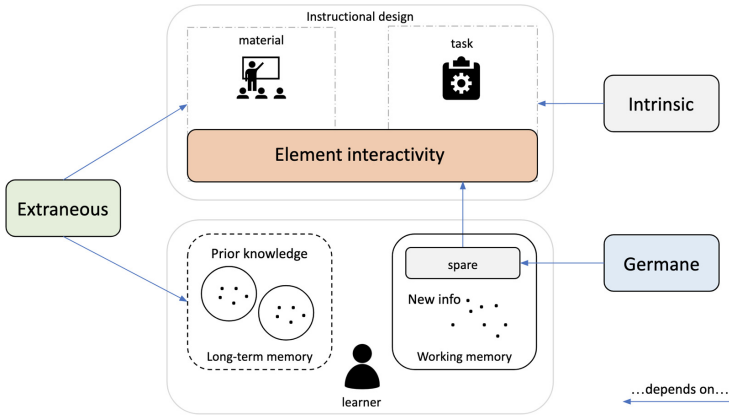
from sensory memory, long term memory instead is unlimited, highly structured and it stores relevant information as acquired knowledge [2, 28]. Short-term memory, as described by Miller (1956), has the capacity to hold seven plus or minus two chunks of information at any given time [28]. It is not specified whether the chunks of information were novel or familiar, interrelated or discrete, but rather that a chunk is a unit of knowledge. Long-term memory is a permanent store of experience, knowledge and process, all of which is held outside the conscious awareness until recalled in the working memory. It does not have an executive function [2]. The information stored in the long-term memory in knowledge structures of varying complexity is called ‘schemata’ [44]. A schemata makes the construction and transfer of knowledge possible, which equates to learning. The more schemata an individual holds for a particular topic, the more advanced they become in learning. Schema construction is believed to reduce the load in working memory. Leaving sufficient cognitive resources in the working memory to process new information is one of the core objectives of educational instructional design [33]. Explicit instructions are required to process information and build schemata of knowledge in working memory. Cognitive Load Theory distinguish three types of load, whose definitions have evolved over time: intrinsic, extraneous and germane loads [33]. Traditionally, CLT has focused on instructional methods aimed at decreasing the cognitive load experienced by learners so that their available resources can be fully devoted to learning. However, these have been redefined over time, and the latest belief, as shown in Fig. 1, is that:

- *intrinsic load* depends on the number of elements to be processed in working memory and on the characteristics of the learning task, which are believed to be fixed and immutable;
- *extraneous load* depends on the characteristic of the instructional material, and the characteristics on the instructional design, and on the prior knowledge of learners;
- *germane load* depends on the characteristics of the learner and the resources allocated to deal with the intrinsic load.

These definitions and terminology will be useful for a subsequent discussion on the limitations of current instructional efficiency models, as described in the following section, as well as the design of a novel model.

## 2.2 Instructional Efficiency

Efficiency of instructional designs in education is a measurable concept. Efficiency in the context of problem-solving, learning and instruction is the capacity to achieve established goals at the minimal expense of resources [12]. Pass and colleagues suggest that combining performance and mental effort measures allows the calculation of an index of mental efficiencies [35, 36]. Studies that investigated processing instructional efficiency made use of uni-variate scores to compare the impact of an experimental condition in respect to a control condition. Sweller (2010) argues that instructional effectiveness will be compromised



**Fig. 1.** Latest definition of the cognitive load types of Cognitive Load Theory [31], including intrinsic, extraneous and germane loads.

by the extent that instructional choices require learners to devote working memory resources to dealing with elements imposed by extraneous cognitive load [45]. It is also believed that, at a basic level, understanding efficiency is an essential precursor to assessing educational effectiveness and improvement. Various studies that propose measures of efficiency have been conducted in the past. The most common and widely used models of efficiency are discussed below.

**Deviational Model** - In search of a single measure to determine the relative efficiency of instructional conditions in terms of learning outcomes, Paas and Van Merriënboer developed a computational approach for combining a measure of performance with a measure of mental effort to attain efficiency [36]. This was characterised as the instructional condition efficiency. This is referred to as the Deviation model of efficiency by Hoffman and Schraw because this model computes the difference between a standardised score of performance and a standardised score of effort [12]. The reasoning behind this formula is based on the assumption that the resulting efficiency is high when an individual experiences high performance and low effort. Conversely, the resulting efficiency is low when an individual experiences low performance and high effort [36]. The deviation model of efficiency computes a measure of efficiency based on how the participant performs relative to the group [12]. It measures the distance from the observed score to the ideal efficiency slope. The deviation model provides a group-referenced score representing an individual efficiency that requires scores to be converted to a common scale. Formally,  $Efficiency = \frac{ZP - ZR}{\sqrt{2}}$ , where  $ZP$  is the standardised performance score and  $ZR$  the standardised effort score. If  $ZP - ZR > 0$ , then efficiency is positive. If  $ZP - ZR < 0$ , then efficiency is negative. According to the authors, the highest efficiency condition occurs when performance is maximal and effort is minimal. The lowest efficiency corresponds to the lowest performance and highest effort [36]. There are concerns expressed

by Hoffman and Schraw that the efficiency score computed by the deviational model is problematic because the standardised scores are affected by variability and performance of others within the group [12]. They also suggested that resulting scores should be interpreted cautiously as they may be conceptually incommensurate even if they can be mathematically identical in magnitude and direction. Additionally, the original deviational model was believed to assess only learning efficiency. However, over the years, a differentiation between various types of efficiency was made [47]. In particular, for the reminded or this article, in line with [46], training efficiency is referred to that measure computed with effort/mental workload obtained prior test performance, while learning efficiency with effort/mental workload obtained after it.

**Likelihood Model** - Another measure of efficiency widely employed within education is based upon the likelihood model put forward by Hoffman and Schraw [12]. Efficiency in this model is computed as a ratio of work output to input. In other words, a ratio of performance to perceived mental effort. Output is identified with learning and input is identified with time, work or effort [43]. Formally  $Efficiency = \frac{P}{R}$ , where  $P$  is the raw score of performance and  $R$  is the raw score of perceived effort. An estimation of the rate of change of performance is calculated by dividing  $P$  by  $R$  and the resulting ratio represents the individual efficiency based on individual scores of performance and effort [12]. The ratio ranges from zero to extensive positive values; it goes towards zero when performance is low and effort is high (low efficiency) and conversely, it goes towards the extensive positive value when performance is high and effort is low (high efficiency). The authors argue that, compared to the deviational model of efficiency, the likelihood model provides an unambiguous measure because the inputs are not standardised scores, and there is no restrictions in the range of efficiency scores. However, the resulting efficiency here is always going to be positive. It must be interpreted with caution because the formula assumes that the work input is not zero [11]. It is also acknowledged that efficiency scores based on this model is supposedly more reliable and sensitive to minor effect size changes compared to the deviational model. An extension of this model has been proposed by Kalyuga and Sweller where an extra reference to a *critical value* is used, under or above which the efficiency can be considered negative or positive [14]. The authors suggest to obtain the critical value by dividing the maximum performance score by the maximum effort exertable by a learner in order to establish whether that learner is competent or not. The ratio of the critical value is based on the underlying assumption that an instructional design is inefficient if a learner invests maximal effort in a task without reaching maximal performance and vice-versa [14]. Through this extended formula, the model evolves from one being able to define only positive efficiency scores to one capable of defining a positive or negative efficiencies.

### 2.3 Mental Effort and Mental Workload

Despite all the years of empirical research, no proven measure of the three cognitive load types of the Cognitive Load Theory have emerged. Similarly, no single measure of mental effort or workload exist, and many models have been proposed, each employing different techniques [19–21, 24, 40, 41]. The concept of cognitive load is mainly employed in the educational field, whereas the concept of mental workload, a psychological construct strictly connected to cognitive load, is employed mainly within ergonomics and human factors [7, 25]. Although very similar, some difference exists. Among them, the former relates mainly to working memory resources, whereas the latter takes into account other factors such as the level of motivation, stress and the physical demand experienced by humans as a consequence of a learning task. Despite of their different fields of research, they both assume that working memory limits, or a limited pool of resources and their capacity must be considered to predict performance while accomplishing an underlying task. Although the field of educational psychology is struggling to find ways of measuring cognitive load of learning tasks [27], there is an entire field within Ergonomics devoted to the design, development and validation of reliable measures of mental workload. Intuitively, mental workload (MWL) can be defined as the volume of cognitive work necessary for an individual to accomplish a task over time [22, 39]. However, many other definitions exists, making its formalisation a non-trivial goal. The measurement of cognitive load is of crucial importance for instructional research. The few efforts to measure it are almost exclusively concerned with performance measures [35]. Different techniques, with different advantages and disadvantages, have been proposed in education to measure cognitive load and they can be clustered in two main groups: subjective and objective measures [37]. Subjective measures are more suitable to be applied in an educational context and in general are easy to administer and analyse, in contrast to objective measures. Subjective measures, also referred to as self-reported measures, rely on the individual's perceived experience of the interaction with a learning task. They are based on the assumption that only the individual involved in the task can provide an accurate and precise judgement about the experienced load, as employed in a number of studies [13, 30]. The perception of the individual can be gathered through means of a survey or questionnaire. Subjective measures include both uni-dimensional and multidimensional approaches, which have been conceptualised, applied and validated. On one hand, the most commonly used subjective measures are uni-dimensional, for example the Rating Scale of Mental Effort (RSME) and its modifications [51]. They provide an index of overall workload/effort, but provide no information about its temporal variation. On the other hand, multidimensional measures, such as the NASA Task Load Index [9] and its raw version [3, 8, 10, 18], can determine the source of mental workload, as based upon different factors. They seem to be the most appropriate types of measures for assessing mental workload because they have demonstrated high levels of sensitivity and diagnosticity [42]. However, it was argued that their use is exceptional within education [4], with only a few studies investigating its validity and sensitivity [5, 6, 17].

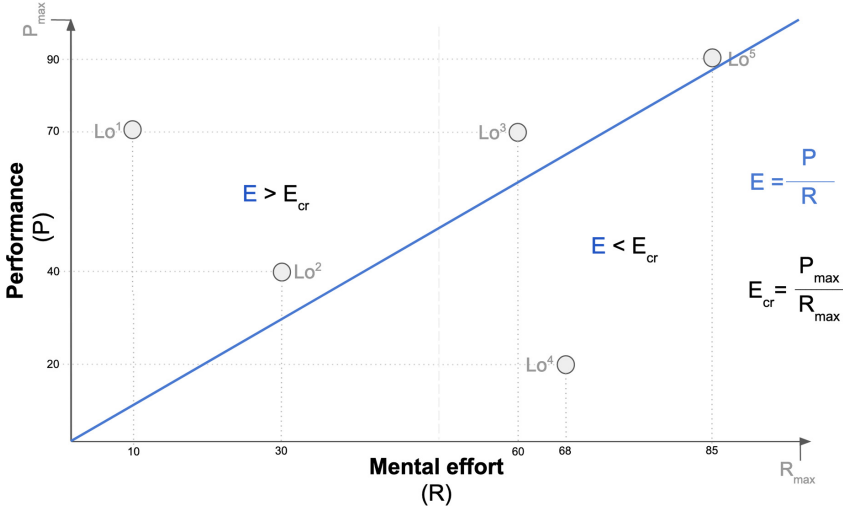
## 2.4 Discussion on Instructional Efficiency Models and Assumptions

The deviational and likelihood models of instructional efficiency, as described in Sect. 2.2, have their own advantages and limitations. The deviational model is group-referenced, and although is good to consistently investigate the variations of efficiency scores of learners within the same instructional design, it does not allow comparison across different instructional designs. In fact, the procedure that converts raw scores to standardised scores affects the variability and performance of learners within a group. Thus scores may be mathematically identical in magnitude and direction across different instructional designs, but they may be conceptually incommensurate. For example, a very difficult instructional design can generate the same distribution of standardised scores when compared to a very easy instructional design. The likelihood model aims at solving some of the problems associated to the deviational model, as it provides an unambiguous measure since the inputs are not standardised scores. Unfortunately, with null effort, the formula cannot work, as dividing performance by zero is mathematically not possible. This technical issue is rather minor since it is reasonable to assume that the plausibility of performance with no reported effort is improbable, and in practice, the scale for effort assessment always starts with values higher than zero. However, a major issue is that, since there are no restrictions in the range of efficiency scores - from greater than zero to infinite - it still makes comparison across different instructional designs difficult. In fact, the efficiency scores obtainable by a group of learners associated to a particular instructional design, can be totally the same when compared to a totally different instructional design, even if the magnitude of performance and effort are different. In other words, an efficiency score of 2 can be achieved both with a performance of 10 and an effort of 5, or with a performance of 20 and an effort of 10 ( $10/5 = 20/10$ ), assuming performance and effort are assessed using the same scale across the two instructional conditions. Eventually, given the definition of efficiency as the capacity to achieve established goals at the minimal expense of resources [12], it is here argued that negative efficiency scores do not make sense. In fact, the achievement of established goals, regardless if with a minimal or maximal expenses of resources, can be null, or equates to some positive extent, but can never be negative. In short, the opposite of achieving, which is not achieving, cannot be measured. Additionally, the fact that a model of instructional efficiency has infinite values can make comparisons across instructional designs more cumbersome. Therefore, it is argued that a measure of efficiency with a positive bounded range is ideal to tackled the above consideration. The deviational and likelihood models of efficiency share the same assumption of other models often employed in education, as synthesised below:

ASSUMPTION 1 (likelihood model)

*Instructional efficiency is higher if similar degrees of performance are achieved with less effort, or similarly, if higher degree of performance are achieved with the same expenditure of resources.*





**Fig. 2.** The likelihood model of efficiency with a critical value, and five learners who achieved different levels of performance, exerting different degrees of effort.

This assumption is the starting point of a critical discussion and a number of considerations aimed at challenging its validity. In order to elaborate on this, consider Fig. 2 where 5 learners ( $Lo^1 - Lo^5$ ) are depicted at the intersection of performance and experienced effort, with efficiency scores computed with the likelihood model (Table 1).

**Table 1.** An illustrative example with 5 learners who exerted different degrees of effort achieving different levels of performance, thus leading to different instructional efficiency scores as computed by the likelihood model [12] with a critical value  $P_{max}/R_{max} = 100/100 = 1$ .

Learner	Performance (P)	Effort (R)	Efficiency (E)
$Lo^1$	70	10	+7.00
$Lo^2$	40	30	+1.33
$Lo^3$	70	60	+1.16
$Lo^4$	20	68	- 0.29
$Lo^5$	90	85	+1.06

According to the likelihood model, by employing the critical value, four learners obtained positive instructional efficiency ( $Lo^1, Lo^2, Lo^3, Lo^5$ ) and one negative efficiency ( $Lo^4$ ), as per Table 1. Through the use of the terminology associated to the latest definition of the cognitive load types of the Cognitive Load

Theory [31] (Sect. 2.1), these learners had various experiences when exposed to a certain instructional design, and these are analysed below.

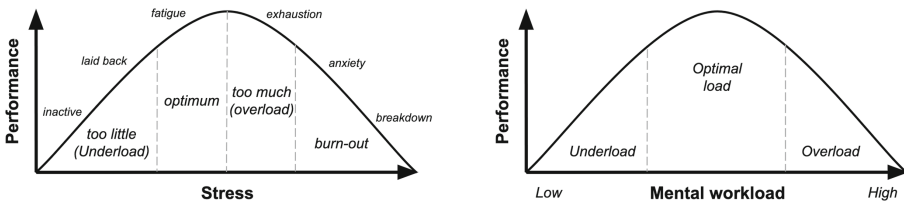
1.  $Lo^1$  - a learner who exerted very low germane load but achieved good performance. Intuitively, this could be a skilled learner with good prior knowledge. As a consequence, the very high efficiency score computed by the likelihood model (+7.0) is not fully justifiable and correctly reflecting the fact that the instructional design might have been probably of low utility for this learner, because of the low intrinsic load. Additionally, the extraneous load was probably intermediate, not influential nor redundant, as the learner achieved good performance anyway.
2.  $Lo^2$  - a learner who exerted minimal germane load and achieved medium performance. This is probably due to the medium extraneous load that means the instructional material might have not been fully engaging and stimulating and the intrinsic load was probably not optimal and not fully influential. Therefore, the efficiency score of the likelihood model (+1.33) is positive and close to the critical line (1). However, it is hard to establish to what extent the instructional design was efficient.
3.  $Lo^3$  - a learner who exerted medium germane load but achieved good performance. The extraneous load was probably optimal, either because of the good prior knowledge of the learner or because of the moderately efficient instructional material. Additionally, the intrinsic load might have been good as the learning task might have moderately and positively engaged the learner. Therefore, the efficiency score of the likelihood model (+1.16) is positive and close to the critical line (1). However, as for learner  $Lo^2$ , it is hard to establish to what extent the instructional design was efficient.
4.  $Lo^4$  - a learner who exerted a good portion of germane load but achieved poor performance. The extraneous load might have been low as the instructional material failed to promote learning and was probably not engaging, or the learner's prior knowledge was almost absent, not allowing to follow the material itself for the formation of knowledge. Also, the intrinsic load might have been probably high and the learner could not cope fully with the learning task. The likelihood model led to an efficiency score of  $-0.29$  which is very close to the critical line (1). Additionally, for the same level of effort 68, but with a lower level of performance, let's say 5, the likelihood model would have led to an efficiency score of  $5/68 = 0.073$ , which would have been deemed negative as below the critical value of 1, thus becoming  $-0.073$ . However, the latter case was clearly worse in term of efficiency than the former case, but its efficiency value is closer to 1 than the former, suggesting a slightly better efficiency (less negative) when in reality should have been deemed worse (more negative). Clearly, this is counter-intuitive, generating confusion during interpretations.
5.  $Lo^5$  - a learner who exerted high germane load and achieved good performance. The extraneous load might have been probably optimal as it might have motivated the learner to exert high germane load for coping with an high intrinsic load and thus correctly promoting the formation of knowledge. This

case might be seen as a situation in which a learner had high perseverance in exerting effort, despite the difficulty of the instructional design exposed to, but eventually achieved high performance. This can be regarded as a positive outcome, in terms of instructional efficiency. However, according to the likelihood model, this learner achieved a positive efficiency score of +1.06, very similar to those of learners  $Lo^2$  and  $Lo^3$ , which is clearly counter-intuitive since they either exerted significant lower effort or achieved significant lower performance.

According to the above considerations, examples, interpretations and discussion, a novel model of instructional efficiency is proposed in the next section. This model, grounded on measures of test performance and cognitive load, is aimed at tackling the above problems by providing scholars with instructional efficiency scores that should support higher interpretation of behavioural performance of learners and facilitate comparisons among instructional designs.

### 3 Design and Methodology

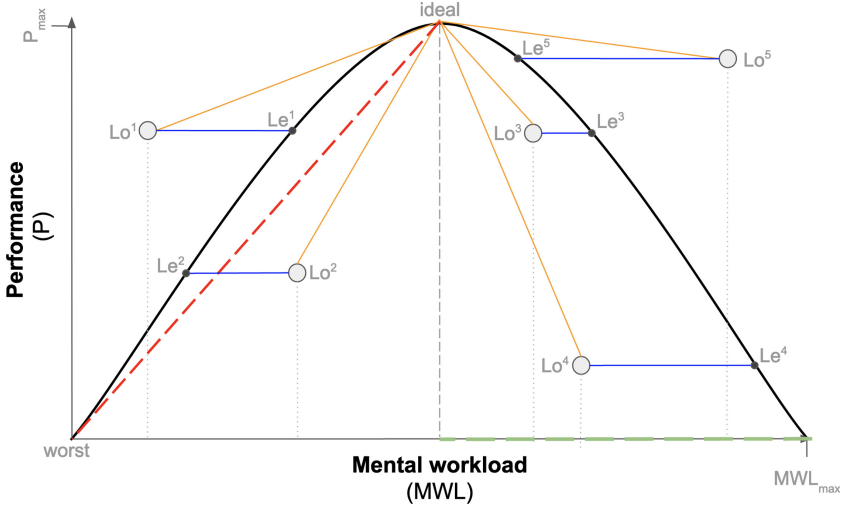
The discussion of Sect. 2.4 serves as the main motivation for the design of a novel model of instructional efficiency that is referred to as the *parabolic model of instructional efficiency*. The word ‘parabolic’ originates from a well-known assumption within mental workload research. This assumption focuses on the parabolic relationship between experienced mental workload and performance, as depicted in Fig. 3 (right). This relationship has been theorised by many scholars within mental workload research over the last 50 years [15, 16, 29, 48, 50], and originally motivated by the Yerkes-Dodson law, as depicted in Fig. 3 (left), whereby the relationship between pressure and performance was formulated from empirical research [49], the well-known inverted U-model. Starting from this reasonable assumption, a novel model of instructional efficiency is proposed, as depicted in Fig. 4.



**Fig. 3.** The Yerkes-Dodson inverted U-model of stress and performance (left) and the assumed relationship between mental workload and performance (right)

This new model expects that performance, intended as test performance of learners, achieves maximality when experienced mental workload and expenditure of resources or effort is moderate, and it achieves minimality when mental

workload is either too little (underload) or too much (overload). The optimal point for performance is depicted with the label *ideal* in Fig. 4, which equates to the top of the inverted parabola (black) that models the expectation itself. This is the first difference with state-of-the-art models of instructional efficiency, as discussed in Sect. 2.4, whereby efficiency is the linear capacity to achieve established goals with minimum expenditure of effort or resources [12], while in the parabolic model, it is the non-linear capacity to achieve established goals with moderate expenditure of effort or mental workload.



**Fig. 4.** A novel parabolic model of instructional efficiency based on ideal mental workload and performance, with 5 illustrative learners. (Color figure online)

Secondly, it is argued that the efficiency associated to a learner depends on the distance between a point at the intersection of exerted mental workload (*MWL*) and observed test performance (*P*) to the *ideal* point. Formally:

$$D(Lo^x, ideal) \tag{1}$$

with *D* is the function of the euclidean distance between two points, which is the length of the line segment between the point  $Lo^x$  (the observed intersection of mental workload and performance for a learner *L*) and the *ideal* point ( $MWL_{max}/2, P_{max}$ ). The yellow segments of Fig. 4 depict these distances for five illustrative learners. The goal of this measure of distance is to penalise learners far from the *ideal* point, in terms of efficiency, and award those close to it. Consequently, the *worst* performing learner is one who exerts minimal mental workload, the closer available point to zero on the *x* axis, with null performance. Therefore, the longest distance from the *ideal* point equates to the red dashed line in Fig. 4. Formally:

$$D(worst, ideal) \tag{2}$$

with  $D$  is the function of the euclidean distance between two points, which is the length of the line segment between the point *worst*  $(0, 0)$  and the *ideal* point  $(MWL_{max}/2, P_{max})$ . Thus, the proportional distance associable to a learner, is the ratio between the formula 1, the observed distance, and the formula 2, the maximum distance to the *ideal* point. Such ratio is inversely proportional to instructional efficiency, which, in this model is set in the range  $[0, 1] \in \mathfrak{R}$ . In other words, the higher the ratio, the lower the efficiency, and vice-versa. Thus, the instructional efficiency associable to a learner is the ratio between the distance to the *ideal* point over the maximum distance to it, subtracted from 1, the maximum value of efficiency. Formally:

$$1 - \frac{D(Lo, ideal)}{D(worst, ideal)} \quad (3)$$

Formula 3 is aimed at penalising those learners far from the *ideal* point, in terms of instructional efficiency. Additionally, it penalises even more learners who have achieved minimal performance, given minimal or maximal mental workload, than those who achieved maximum performance with minimal or maximal exertion of mental workload. This is because the distance between the point  $(0, P_{max})$  and the *ideal* point, is lower than the distance between the point  $(0, 0)$  and the *ideal* point. In other terms,  $D((0, P_{max}), ideal) < D((0, 0), ideal)$ .

Beside the distance to the *ideal* point, and the penalisation of learners far from it, another factor to consider is the expected behaviour of learners, given experienced mental workload and achieved performance. This is modeled with the parabola (black) of Fig. 4, following the assumption stated above, and depicted in Fig. 3 (right). In detail, given an observed level of performance, the parabola defines a corresponding level of mental workload that a learner is expected to exert for reaching such performance. Departures from such an expected point on the parabola equates to penalisations, in terms of instructional efficiency. This can be modelled with another measure of distance between the *observed* behaviour ( $o$ ), the actual point at the intersection of actual mental workload and actual performance, and the *expected* behaviour ( $e$ ), the point at the intersection between expected mental workload and actual performance on the parabola, which defines a straight horizontal segment (blue segments in Fig. 4). Formally,

$$D(Lo, Le) \quad (4)$$

with  $D$  is the function of the euclidean distance between two points, which is the length of the line segment between the observed point at the intersection between observed mental workload and performance of learner  $L$  ( $Lo$ ), and the point at intersection of expected mental workload given a certain level of experienced performance ( $Le$ ). The longest straight segment that can be drawn from the parabola equates to the half of the mental workload range  $MWL_{max}/2$  (green segment). This is the maximum distance from the parabola that models a learner who achieved null or maximal performance, with either minimal or maximal mental workload (the four corners of the diagram of Fig. 4), the farthest points from the expected points on the parabola. Thus, the proportional

distance associable to a learner is the ratio between the formula 4, the distance to the expected point on the parabola, and  $MWL_{max}/2$ , the maximal distance achievable from the parabola. Such ratio is inversely proportional to instructional efficiency, which, as mentioned above, is set in the range  $[0, 1] \in \mathfrak{R}$ . Thus, the instructional efficiency associable to a learner, in this case, is the ratio between the distance to the expected point on the parabola, over the maximal distance to it, subtracted from 1, the maximal value of efficiency. Formally:

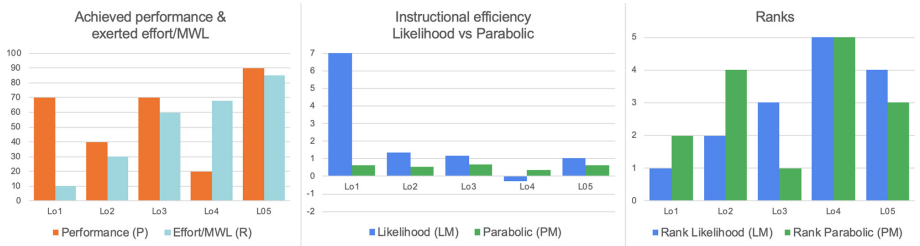
$$1 - \frac{D(Lo, Le)}{MWL_{max}/2} \tag{5}$$

Formulas 3 and 5 are two terms of instructional efficiency, and since they are percent proportions, they are bounded in the same range  $[0, 1] \in \mathfrak{R}$ . They are treated with equal priority, thus have the same importance. Eventually, averaging them give a combined measure of instructional efficiency, also bounded in  $[0, 1] \in \mathfrak{R}$ . Formally:

$$E : [0, 1] \in \mathfrak{R}$$

$$E = \frac{1}{2} \left( 1 - \frac{D(Lo, ideal)}{D(worst, ideal)} \right) + \frac{1}{2} \left( 1 - \frac{D(Lo, Le)}{MWL_{max}/2} \right) \tag{6}$$

This novel measure of instructional efficiency is aimed at solving the theoretical issues emerged in the discussion of Sect. 2.4. In fact, let us consider the illustrative learners, their performance and effort scores, presented in Table 1, and depicted in Figs. 2 and 4. The application of the parabolic model to these learners generates a new set of efficiency scores, as depicted in Fig. 5.



**Fig. 5.** A comparison between the likelihood model and the parabolic model of instructional efficiency, with five illustrative types of learner.

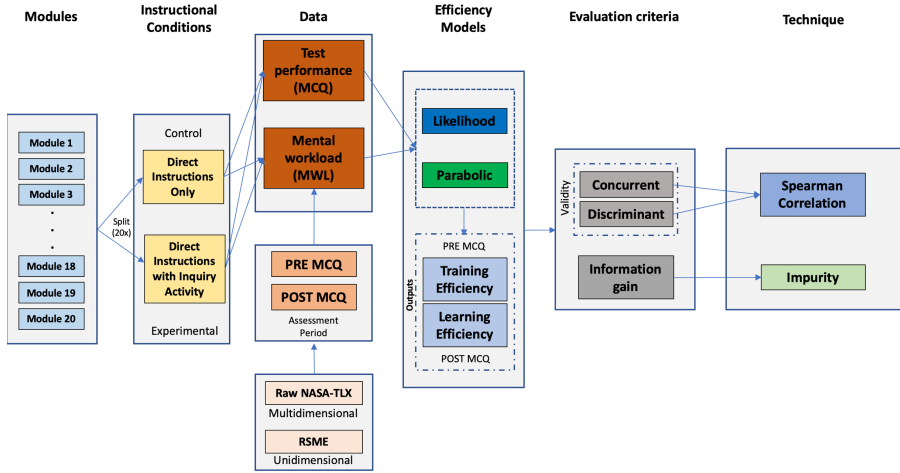
As it is possible to see, the efficiency score computed by the parabolic model (PM) for the first learner  $Lo^1$  has been drastically reduced from the score computed by the likelihood model (LM). This is because of the bounded range within

0 and 1 of the parabolic model as well as its capacity to penalise learners whose intersection between performance and mental workload is not close to the *ideal* point on the parabola, and far from the expected exertable mental workload level given the achieved performance. However, the rank of  $Lo^1$ , out of five learners, is 1 (the best) for the likelihood model, and 2 for the parabolic model, thus not drastically different in the group. On the one hand,  $Lo^2$  has instead earned a rank of 4 with the parabolic model, against 2 reached with the likelihood model, loosing position in the group. This is reasonable, given the fact that this learner has exerted very low effort, with a moderate performance, thus the underlying instructional design should not be considered as very efficient, because probably it did not engage the learner. On the other hand,  $Lo^3$  has achieved rank 1 with the parabolic model, against rank 3 with the likelihood model. In this case, the efficiency of the underlying instructional design could be considered good, as it might have engaged the learner because of the good level of performance reached.  $Lo^4$  was the worst both for the likelihood and the parabolic models, demonstrating a consistent outcome. This learner exerted a good amount of mental workload, but achieved very low performance, thus reasonably deriving that the underlying instructional design was not efficient. For this learner, it is also possible to see the effect of the parabolic model not to produce negative efficiencies, as instead possible with the likelihood model. Eventually,  $Lo^5$  was a learner who exerted a very high mental workload, but still positively achieving a very high performance. Although this is good, because the performance scores actually points to a good learning experience, the exertable effort could have been reduced, thus the underlying instructional condition cannot be considered optimal in terms of efficiency.

### 3.1 Research Hypothesis and Comparative Design

In order to gauge the empirical value of the parabolic model for tackling the challenging technical problem of instructional efficiency assessment within education, a comparative experiment has been designed to assess its validity and informativeness. This experiment is replicable and indeed should be replicated multiple times across various studies in order to validate such a novel model, and this article is the very first attempt towards this goal. A research hypothesis is set below and the detailed research design is depicted in Fig. 6.

*'H<sub>1</sub>: IF the parabolic model (PM) is employed to compute the instructional efficiencies of a set of instructional designs,  
THEN it is expected that it exhibits a good concurrent validity with the likelihood model (LM) but a superior discriminant validity and higher information gain'*



**Fig. 6.** The comparative experimental design diagram with details on instructional conditions, collected data, models of instructional efficiency employed, the evaluation criteria and methods employed to assess them.

Data for this experiment was collected in various taught modules at the Technological University Dublin, in the context of a larger research project in education [31,32,34]. All procedures performed involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Ethical approval was obtained by the ethics committee of the above institution. Informed consent was obtained from all individual participants involved in the experiment, who were also provided with a study information sheet and could withdraw the study at any point in time. This experiment deals with two instructional design conditions. The first condition followed the traditional direct instruction approach to learning, while the second design extended it with a collaborative inquiry activity. In detail, the former approach involved a theoretical explanation of a chosen topic, whereby the instructor presented the information through direct instructions, projected to a white board, and verbally narrated. The latter approach extended the former approach with a guided inquiry activity amongst participants based on the use of cognitive trigger questions. Both groups received direct instructions, while only the experimental group subsequently participated in the collaborative inquiry activity. In the wider research project that originated such an experiment [31–34], the purpose of the second condition was to establish whether the additional inquiry activity could improve the efficiency of learners when compared to those learners who receive direct instructions only. Further details on the rationale behind the two conditions is not necessary here, since the goal is the comparison of two instructional efficiency models. For further detail, the reader is referred to our other published work [31,32,34]. In the experiment set here, any two different



instructional conditions, assuming their difference, could have also been suitable. In fact, the more replications of this experiment, the better for its evaluation and thus validation.

In the first instructional condition, after the topic was presented to the class by an instructor, on the one hand the control group participants received questionnaires aimed at quantifying the effort and mental workload they experienced, using the Rating Scale Mental Effort (RSME) [51] and the raw version of NASA task load index (R-NASA) [9], models described in Sect. 2.3. Subsequently to these self-reporting questionnaires, a multiple-choice questionnaire (MCQ) associated to the taught topic was administered to the control group. On the other hand, the experimental group, the remaining half of the class, was split into teams of three or four participants for the inquiry activity. The participants discussed and exchanged information related to the topic leading to informed agreements collaboratively. The participants then wrote the shared answers individually to the cognitive trigger questions, part of the inquiry activity. Subsequently, the participants in the experimental group, similarly to those in the control group, received the same self-reporting questionnaires (RSME, R-NASA, and MCQ). Once the participants in both groups completed the MCQ, they were provided again with another RSME and Raw-NASA questionnaires. Filling the questionnaire on both occasions (pre and post MCQ) allows for the computation of both the training and the learning efficiencies, as they are related to different stages of the learning process [47]. Efficiency scores were then calculated using both the RSME and the R-NASA for each model of efficiency, namely the likelihood model (LM), used as baseline, and the parabolic model (PM), resulting in a total of eight different efficiency models, as shown in Table 2.

**Table 2.** Experimental configuration of the likelihood model (LM) and the parabolic model (PM) of instructional efficiency, with both the use of the Rating Scale Mental Effort (RSME) and the raw NASA Task Load index (RNASA) for the assessment of mental workload (MWL), and a multiple choice questionnaire (MCQ) for the assessment of test performance.

Model label	Efficiency type	Instructional model	MWL	Performance
LrEff-LM-RNasa	Learning	Likelihood (LM)	R-NASA	MCQ
LrEff-LM-Rsme	Learning	Likelihood (LM)	RSME	MCQ
LrEff-PM-RNasa	Learning	Parabolic (PM)	R-NASA	MCQ
LrEff-PM-Rsme	Learning	Parabolic (PM)	RSME	MCQ
TrEff-LM-RNasa	Training	Likelihood (LM)	R-NASA	MCQ
TrEff-LM-Rsme	Training	Likelihood (LM)	RSME	MCQ
TrEff-PM-RNasa	Training	Parabolic (PM)	R-NASA	MCQ
TrEff-PM-Rsme	Training	parabolic (PM)	RSME	MCQ

### 3.2 Evaluation Criteria

To assess the instructional efficiency models, namely the likelihood model (LM) and the parabolic model (PM), with their 8 variations, a set of evaluation criteria are selected, as described below, namely concurrent validity, discriminant validity and information gain.

*Concurrent validity* is the capacity of a measure to actually represent an underlying construct, in this case instructional efficiency, and it can be demonstrated by correlating the efficiency scores generated by a new model, here the parabolic model (PM), with those from another well-known model, in this case the likelihood model (LM). Concurrent validity is in practice measured with a correlation test.

*Discriminant validity* is aimed at measuring whether two constructs that theoretically should not be highly related to each other are, in fact, not related to each other. In this experiment the capability of the likelihood model, and the parabolic model is assessed, to better separate training and learning efficiencies, as two independent constructs and, in practice, this is achieved by performing a correlation test between the training and learning efficiency scores of each model.

*Information gain* is the level of impurity in a group of observations. It measures how much ‘information’ a variable provides about a target variable. In this research, it is used to understand which instructional efficiency model leads to efficiency scores that provide a higher amount of information about whether an observation belongs to a specific group (control or experimental). It is a measure of reduction in entropy (H) by transforming a set of data in some way. In practice, it is calculated by comparing the entropy of the data before and after a transformation.  $H = \sum_i -p_i \log_2 p_i$ , with  $p$  the probability of class  $i$  that means the proportion of class  $i$  in a set. The higher the entropy the less the information content. The idea is to look at how much it is possible to reduce the entropy of the parent node (group) by segmenting on a given child (efficiency score). Formally,  $IG = H_p - \sum_{i=1}^n p_{ci} H_{ci}$ , where  $H_p$  is the entropy of the parent node,  $n$  is the number of child segments of the target variable,  $p_{ci}$  is the probability that an observation is in child  $i$  (the weighting), and  $H_{ci}$  is the entropy of child  $i$ . To compute the information gain the following steps are necessary:

1. split the efficiency score variable of each efficiency model into 5 bins;
2. calculate the entropy for each bin;
3. calculate the proportion of all records in each bin;
4. compute the entropy of the parent node (group);
5. subtract the sum of the entropy scores of the bin weighted by the proportion of data they represent from the root node entropy, obtaining information gain.
6. repeat the steps above for all 20 taught modules and for each instructional efficiency model and their variations.

The information gain associated to each version of an instructional efficiency model will be assigned a rank value from 1 to 8, 1 to the model with the most information gain units (I.G.), 8 to the model with the least I.G. units, per taught module, and the rank values will be aggregated across all the taught 20 modules.

This aggregated value is referred to as the sum of rank values. Subsequently, the efficiency models will be ranked again from 1 to 8 based on this sum of rank values, with 1 indicating the least sum of rank values achieved, 8 the highest. This will be referred to as the ranking position across models. In summary, the efficiency models that achieve lower ranking position will be deemed superior, in terms of average information gain across taught modules.

## 4 Results and Evaluation

The data collected contained 455 observations across 20 university modules in computer science as per Table 3. Most of the taught topics were different and data was collected over 3 years, each over two academic semesters, and participants in each class were different than those in all the other classes, thus independent to each other.

**Table 3.** Distribution of participants across taught modules in computer science and how they were split across control and experimental groups.

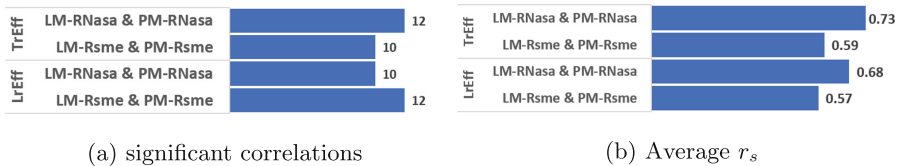
Module ID	Module name	Control	Experimental	Total learners
1	Research methods	14	15	29
2	Research hypothesis	20	16	36
3	Visualising geo spatial data	5	7	12
4	Operating systems	20	18	38
5	Problem solving	14	11	25
6	Data mining	10	9	19
7	Literature review	7	8	15
8	Research hypothesis	8	8	16
9	Strings	10	12	22
10	Program design	15	15	30
11	Machine learning	5	8	13
12	Image processing	7	9	16
13	Research methods	8	9	17
14	Statistics	6	7	13
15	IT forensics	19	14	33
16	Literature comprehension	7	9	16
17	Virtual memory	8	7	15
18	Research hypothesis	18	14	32
19	Literature review	16	15	31
20	Operating systems	14	13	27

The following scores were computed using collected data:

1. RSME (Rating Scale Mental Effort) scores (R) for each participant: pre and post a multiple choice questionnaire (training and learning);
2. a Raw NASA-TLX (Nasa Task Load Index) scores (R) for each participant: pre and post a multiple choice questionnaire (training and learning);
3. a performance score (P), in percentage, gathered from the application of a multiple choice questionnaire tailored to each taught module, and designed with the expert lecturer for that taught topic;
4. an efficiency score for each participant, by employing the two mental workload assessment techniques of 1 and 2, both for the two phases (training and learning), and for both the two models of instructional efficiency (likelihood and parabolic). In total 8 efficiency scores are produced for each participant, in line with the models defined in Table 2.

### 4.1 Concurrent Validity

All the efficiency scores were paired up with each other for each learner, and for each instructional efficiency model (LM and PM) and their 8 variations (Table 2), grouped by taught module (Table 3). The non-parametric Spearman correlation test was performed for each of these paired lists. The Pearson correlation instrument was not used since its assumptions were not met. Figure 7(a) illustrates the number of statistically significant correlations of the models, grouped by learning and training efficiency (LrEff, TrEff), and by the mental workload self-reporting assessment instrument employed (RSME, R-NASA). Figure 7(b) shows the average correlation coefficient  $r_s$  of these significant correlations.



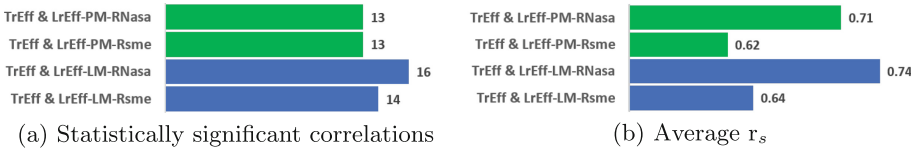
**Fig. 7.** Concurrent validity of the parabolic model (PM) of instructional efficiency with the likelihood model (LM) with a) depicting the number of significant correlation tests across the 20 taught modules, and b) their average Spearman correlation coefficients.

Out of 20 possible modules, the parabolic model has shown, for more than half of them, a high concurrent validity with the likelihood model, with average correlation coefficients higher than 0.57, these being statistically significant ( $\alpha < 0.05$ ). The lack of a perfect correlation was expected, since the parabolic model measures instructional efficiency differently than the likelihood model. However, the fact that for more than half of the modules the two models of instructional

efficiency produce scores moderately and highly correlated, it shows that the underlying construct they are actually measuring is similar, and not drastically different.

## 4.2 Discriminant Validity

The correlation between the training and learning efficiency scores of each model (PM and LM) were examined using Spearman’s correlation test to determine their discriminant validity across the 8 models (Table 2) and the taught modules (Table 3). The Spearman test was used because the assumption of the Pearson test were not met. Figure 8a illustrates the number of statistically significant correlations between the efficiency score pairs, while Fig. 8b their average spearman’s correlation ( $r_s$ ).



**Fig. 8.** Discriminant validity of the parabolic model (PM) of instructional efficiency with the likelihood model (PM) with a) depicting the number of significant correlation tests across the 20 taught modules, and b) their average Spearman correlation coefficients. (Color figure online)

Results demonstrate that the parabolic model (PM) has led to both the lowest number of significant correlations (13 + 13, Fig. 8a), against those of the likelihood model (16 + 14), and for those significant, it exhibited the lowest correlation coefficients (0.71, 0.62, Fig. 8b) against those of the likelihood model (0.74, 0.64). This is promising as it shows that the parabolic model led to compute training and learning efficiency scores that are more distinctive, since they do not correlate, when compared to those generated by the likelihood model. In other words, evidence suggests that the parabolic model (green bars of Fig. 8) has a superior discriminant validity than the likelihood model (blue bars).

## 4.3 Information Gain

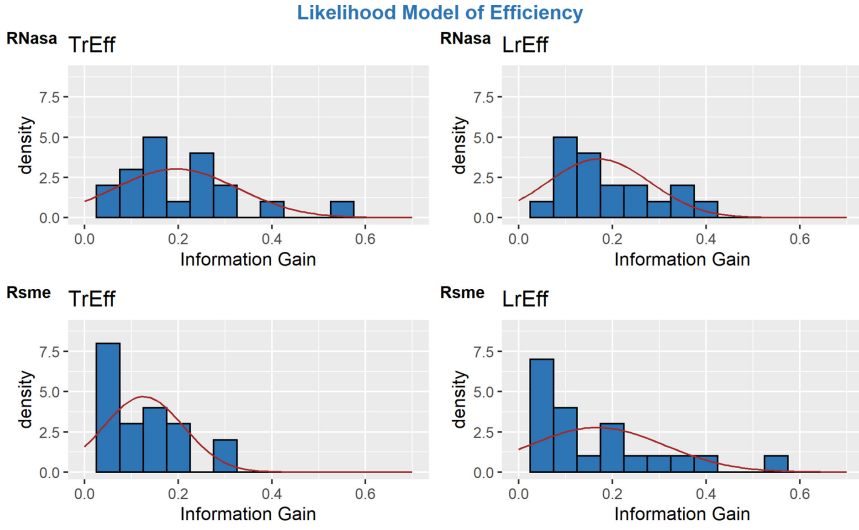
Information gain was calculated for each of the efficiency models (Table 2) in order to explore which one provides the most information about whether or not an observation belongs to the control or experimental group, for each of the 20 modules (Table 3). In other words, through a measure of information gain, it is possible to determine which efficiency model provides the ‘purest’ segmentation with respect to these two groups. Figure 9 shows the distribution of the information gain values for these models.

Table 4 provides a summary of all the information gain units (I.G. units) calculated for all of these efficiency models across the 20 modules, and ranked by lower to higher, that means from the model that exhibits on average lower entropy to that of higher entropy.

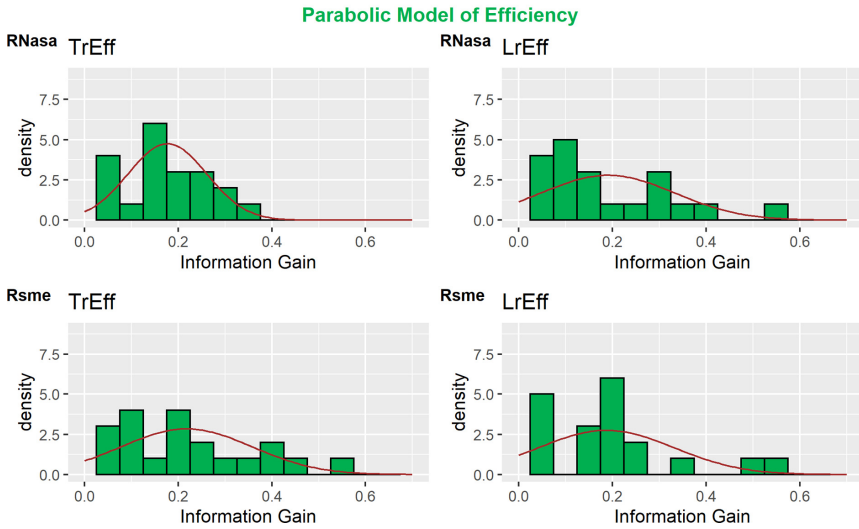
**Table 4.** Ranking of efficiency score models based on the Information gain (I.G.) units, with Tr indicating the training phase, Lr the learning phase, Rsme the Rating Scale Mental Effort, RNasa the Raw Nasa Task load Index, PM the parabolic model, and LM the likelihood model of instructional efficiency.

Efficiency model	Rank Val. Sum	I.G. Avg	I.G. StDev.	Ranking Pos.
<b>TrEff-PM-Rsme</b>	74	0.218	0.141	1
TrEff-LM-RNasa	80	0.195	0.132	2
<b>TrEff-PM-RNasa</b>	83	0.176	0.084	3
<b>LrEff-PM-Rsme</b>	88	0.187	0.145	4
LrEff-LM-RNasa	90	0.171	0.109	5
<b>LrEff-PM-RNasa</b>	91	0.190	0.143	6
LrEff-LM-Rsme	97	0.167	0.144	7
TrEff-LM-Rsme	109	0.126	0.085	8

The parabolic model of instructional efficiency, as it is possible to observe from the ranks of Table 4 (highlighted rows), provided higher information gain (less entropy) across model variations when compared to the variations of the likelihood model, thus had lower rank values on average. In fact, for the training phase, the parabolic model had a sum of rank values lower than the likelihood model (74 vs 80). Similarly, for the learning phase, it has a lower rank values than the likelihood model (88 vs 90). The application of the unidimensional measure of effort, or that of the multi-dimensional measure of mental workload, did not provide clear evidence of the superiority of one measure over the other for producing lower ranks. Thus they seem equivalent, and as a consequence, the multidimensional measure might be preferable, as it can give more information on the learner’s behaviour for post-hoc analysis. In summary, the parabolic model tends to generate efficiency scores with higher information gain than those produced by the likelihood model, as confirmed by the general lower rank values, and higher averages of information gain units.



(a) likelihood model (LM)



(b) parabolic model (PM)

**Fig. 9.** Density plots of information gain values per efficiency model, with Tr indicating the training phase, Lr the learning phase, Rsme the Rating Scale Mental Effort, RNasa the raw Nasa Task load Index, PM the parabolic model, and LM the likelihood model of instructional efficiency.

## 5 Summary and Future Work

Instructional efficiency is a measurable concept within education and it has gained importance for the quantitative assessment of the efficiency of instructional designs. After a description of the deviational and the likelihood model of instructional efficiency, two well-known models employed within education, and after highlighting their advantages and disadvantages, the constructs of effort and mental workload have been briefly reviewed. This is because the above models of instructional efficiency are based upon a measure of effort or workload, which is combined with a measure of test performance. Additionally, on one hand, the main assumption behind these models is that efficiency is the capacity to achieve established goals, assessed with test performance measures, at the minimal expense of resources [12], assessed with perceived effort or mental workload. However, on the other hand, this article, through the proposal of a novel model of instructional efficiency, the *parabolic model*, challenged this assumption. Through illustrative examples, and by employing the various cognitive load types, fundamental components of the Cognitive Load Theory, it has been shown how this assumption is not always reasonable. This novel and one of its kind model is based upon the assumption that the relationship between mental workload and performance has an inverted-U shape, similarly to the well-known Yerkes–Dodson law of stress and performance [49]. Performance should peak with moderate mental workload, an *ideal* point, and be at the minimal level when experienced mental workload is either at a minimal or maximal level, that means underload or overload.

An empirical experiment has been designed, part of a larger study within education, aimed at investigating the inferences produced by this novel model when compared to the likelihood model of instructional efficiency. Empirical evidence is promising as it shows how the parabolic model has a good concurrent validity with the likelihood model, reasonably demonstrating its capacity to actually measure the construct of instructional efficiency. However, it has a better discriminant validity as demonstrated by a lower correlation between the efficiency scores of the training and the learning phases. Eventually, the parabolic model has led to efficiency scores with a higher information gain, as assessed by employing a measure of information entropy, highlighting its potential for the evaluation of instructional designs with practical implications for the disciplines of teaching and learning and generally for education. Future work are needed to further validate the parabolic model of instructional efficiency. This includes the replication of the experiment conducted in this study over additional instructional designs, and a further assessment of the efficiency scores generated by the parabolic model across a larger set of evaluation criteria [23, 26], such as for example sensitivity, reliability, predicting validity.

## References

1. Atkinson, R., Shiffrin, R.: Human memory: a proposed system and its control processes. In: Psychology of Learning and Motivation, vol. 2, pp. 89–195. Academic Press (1968)



2. Baddeley, A.: Is working memory still working? *Am. Psychol.* **56**, 851–864 (2001)
3. Bittner, A.C., Jr., Byers, J.C., Hill, S.G., Zaklad, A.L., Christ, R.E.: Generic workload ratings of a mobile air defense system (LOS-F-H). *Proc. Hum. Factors Soc. Annu. Meet.* **33**(20), 1476–1480 (1989)
4. De Jong, T.: Cognitive load theory, educational research, and instructional design: some food for thought. *Instr. Sci.* **38**(2), 105–134 (2010)
5. Fischer, S., Lowe, R.K., Schwan, S.: Effects of presentation speed of a dynamic visualization on the understanding of a mechanical system. *Appl. Cogn. Psychol.* **22**(8), 1126–1141 (2008)
6. 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)
7. Hancock, G., Longo, L., Hancock, P., Young, M.: Mental workload. In: Salvendy, G., Karwalski, W. (eds.) *Handbook of Human Factors & Ergonomics*, 5 edn., Chap. 7. Taylor & Francis (2021)
8. Hart, S.G.: NASA-task load index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **50**(9), 904–908 (2006)
9. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) *Human Mental Workload*, *Advances in Psychology*, vol. 52, pp. 139–183. North-Holland, Oxford (1988)
10. HENDY, K.C., Hamilton, K.M., Landry, L.N.: Measuring subjective workload: when is one scale better than many? *Hum. Factors* **35**(4), 579–601 (1993)
11. Hoffman, B.: Cognitive efficiency: a conceptual and methodological comparison. *Learn. Instr.* **22**(2), 133–144 (2012)
12. Hoffman, B., Schraw, G.: Conceptions of efficiency: applications in learning and problem solving. *Educ. Psychol.* **45**(1), 1–14 (2010)
13. Junior, A.C., Debruyne, C., Longo, L., O’Sullivan, D.: On the mental workload assessment of uplift mapping representations in linked data. In: Longo, L., Leva, M.C. (eds.) *H-WORKLOAD 2018*. CCIS, vol. 1012, pp. 160–179. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-14273-5\\_10](https://doi.org/10.1007/978-3-030-14273-5_10)
14. Kalyuga, S., Sweller, J.: Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Education Tech. Research Dev.* **53**, 83–93 (2005)
15. Kantowitz, B.H.: 3. mental workload. In: *Advances in Psychology*, vol. 47, pp. 81–121. Elsevier (1987)
16. Kantowitz, B.H.: Attention and mental workload. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, pp. 3–456. SAGE Publications, Los Angeles (2000)
17. 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)
18. Liu, Y., Wickens, C.D.: Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. *Ergonomics* **37**(11), 1843–1854 (1994)
19. 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](https://doi.org/10.1007/978-3-642-23768-3_43)

20. Longo, L.: Formalising human mental workload as non-monotonic concept for adaptive and personalised web-design. In: Masthoff, J., Mobasher, B., Desmarais, M.C., Nkambou, R. (eds.) UMAP 2012. LNCS, vol. 7379, pp. 369–373. Springer, Heidelberg (2012). [https://doi.org/10.1007/978-3-642-31454-4\\_38](https://doi.org/10.1007/978-3-642-31454-4_38)
21. Longo, L.: Formalising human mental workload as a defeasible computational concept. Ph.D. thesis, Trinity College (2014)
22. Longo, L.: A defeasible reasoning framework for human mental workload representation and assessment. *Behav. Inf. Technol.* **34**, 758–786 (2015)
23. 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: Proceedings of the 10th International Conference on Computer Supported Education, CSEDU 2018, Funchal, Madeira, Portugal, 15–17 March 2018, vol. 2, pp. 166–178 (2018)
24. 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](https://doi.org/10.1007/978-3-642-15314-3_6)
25. Longo, L., Leva, M.C. (eds.): H-WORKLOAD 2017. CCIS, vol. 726. Springer, Cham (2017). <https://doi.org/10.1007/978-3-319-61061-0>
26. 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., Uhomiohi, J. (eds.) CSEDU 2018. CCIS, vol. 1022, pp. 384–413. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-21151-6\\_19](https://doi.org/10.1007/978-3-030-21151-6_19)
27. Longo, L., Orru, G.: Evaluating instructional designs with mental workload assessments in university classrooms. *Behav. Inf. Technol.* **0**(0), 1–31 (2020)
28. Miller, G.A.: The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**(2), 81–97 (1956)
29. Moray, N.: *Mental Workload: Its Theory and Measurement*, vol. 8. Springer, Boston (2013). <https://doi.org/10.1007/978-1-4757-0884-4>
30. Moustafa, K., Longo, L.: Analysing the impact of machine learning to model subjective mental workload: a case study in third-level education. In: Longo, L., Leva, M.C. (eds.) H-WORKLOAD 2018. CCIS, vol. 1012, pp. 92–111. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-14273-5\\_6](https://doi.org/10.1007/978-3-030-14273-5_6)
31. 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: Proceedings of the 10th International Conference on Computer Supported Education (CSEDU 2018) vol. 2, pp. 292–302 (2018)
32. Orru, G., Longo, L.: Direct instruction and its extension with a community of inquiry: a comparison of mental workload, performance and efficiency. In: Proceedings of the 11th International Conference on Computer Supported Education, CSEDU 2019, Heraklion, Crete, Greece, 2–4 May 2019, vol. 1, pp. 436–444 (2019). <https://doi.org/10.5220/0007757204360444>
33. 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](https://doi.org/10.1007/978-3-030-14273-5_3)

34. Orru, G., Longo, L.: Direct and constructivist instructional design: a comparison of efficiency using mental workload and task performance. In: Longo, L., Leva, M.C. (eds.) *H-WORKLOAD 2020*. CCIS, vol. 1318, pp. 99–123. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-62302-9\\_7](https://doi.org/10.1007/978-3-030-62302-9_7)
35. Paas, F., Van Merriënboer, J.J.G., Adam, J.: Measurement of cognitive load in instructional research. *Percept. Mot. Skills* **79**, 419–430 (1994)
36. Paas, F., Van Merriënboer, J.J.: The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Hum. Fact. J. Hum. Fact. Ergon. Soc.* **35**(4), 737–743 (1993)
37. Plass, J.L., Moreno, R., Brünken, R.: *Evaluating Instructional Designs Cognitive Load Theory*, 1st edn. Cambridge University Press, New York (2010)
38. Reed, S.K.: *Human Cognitive Architecture*, pp. 1452–1455. Springer, Boston (2012). <https://doi.org/10.1007/978-1-4419-1428-6>
39. 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](https://doi.org/10.1007/978-3-319-44944-9_19)
40. Rizzo, L., Longo, L.: Inferential models of mental workload with defeasible argumentation and non-monotonic fuzzy reasoning: a comparative study. In: *Proceedings of the 2nd Workshop on Advances in Argumentation in Artificial Intelligence*, co-located with XVII International Conference of the Italian Association for Artificial Intelligence, AI<sup>3</sup>@AI\*IA 2018, Trento, Italy, 23 November 2018, pp. 11–26 (2018)
41. 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. In: *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence AI3AI-IA*. CEURS (2017)
42. Rubio Valdehita, S., Ramiro, E., García, J., Puente, J.: Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Appl. Psychol.* **53**, 61–86 (2004)
43. Smith, P.C., Street, A.: Measuring the efficiency of public services: the limits of analysis. *J. Roy. Stat. Soc. Ser. A (Stat. Soc.)* **168**(2), 401–417 (2005)
44. Sweller, J.: Evolution of human cognitive architecture. *Psychol. Learn. Motiv.* **43**(2003), 216–266 (2003)
45. Sweller, J.: Element interactivity and intrinsic, extraneous, and Germane cognitive load. *Educ. Psychol. Rev.* **22**, 123–138 (2010)
46. Tuovinen, J.E., Paas, F.: Exploring multidimensional approaches to the efficiency of instructional conditions. *Instr. Sci.* **32**, 133–152 (2004)
47. Van Gog, T., Paas, F.: Instructional efficiency: revisiting the original construct in educational research. *Educ. Psychol.* **43**(1), 16–26 (2008)
48. Wickens, C.D.: Multiple resources and mental workload. *Hum. Factors* **50**(3), 449–455 (2008)
49. Yerkes, R.M., Dodson, J.D., et al.: *The Relation of Strength of Stimulus to Rapidity of Habit-Formation. Punishment: Issues and Experiments*, pp. 27–41 (1908)
50. Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A.: State of science: mental workload in ergonomics. *Ergonomics* **58**(1), 1–17 (2014)
51. Zijlstra, F., Doorn, L.: The construction of a scale to measure perceived effort. Department of Philosophy and Social Sciences, p. 53, January 1985