

1Running head: INSTRUCTIONAL EFFICIENCY

**This is a pre-print of: Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43, 16-26.**

Copyright Taylor & Francis; *Educational Psychologist* is also available at

<http://www.informaworld.com/smpp/title~content=t775653642~db=all>

Instructional Efficiency: Revisiting the Original Construct in Educational Research

Tamara van Gog<sup>1,\*</sup> and Fred Paas<sup>1,2</sup>

<sup>1</sup> Educational Technology Expertise Center, Open University of The Netherlands

<sup>2</sup> Psychology Department, Erasmus University Rotterdam

\* Correspondence concerning this article should be addressed to Tamara van Gog, Educational Technology Expertise Center, Open University of The Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands. T: +31 45 5762276, F: +31 45 5762907, E: [tamara.vangog@ou.nl](mailto:tamara.vangog@ou.nl)

## 1 Abstract

This article revisits Paas and Van Merriënboer's (1993) measure of instructional efficiency, which can be applied by educational researchers to compare the effects of different instructional conditions on learning. This measure relied on performance and mental effort on the *test*, and as such gave an indication of the quality of learning outcomes. The acquisition of more (less) efficient cognitive schemata is indicated by combinations of high (low) performance and low (high) mental effort. This instructional efficiency measure has become widely adopted, but in an adapted form that incorporates mental effort invested in the learning phase instead of the test phase. This article demonstrates that the adaptation has important consequences for the construct of instructional efficiency, and for the type of conclusions that can be drawn. Examples are given to illustrate the various implications of different combinations of mental effort and performance measures in the light of more contemporary developments in educational research.

## Instructional Efficiency: Revisiting the Original Construct in Educational Research

In 1993, Paas and Van Merriënboer introduced a measure of efficiency of instructional conditions, based on test performance and mental effort invested to attain this test performance, that could aid researchers and instructional designers in comparing the effects of different instructional approaches on learning. What became known as the measure of “instructional efficiency”, has become widely adopted, but in an adapted form, based on mental effort invested in learning and test performance. This way, the construct that was measured changed from efficiency of instructional conditions in terms of learning *outcomes*, into efficiency in terms of the learning *process*. This article revisits Paas and Van Merriënboer’s measure of instructional efficiency in the perspective of its original intention and recent developments in educational research. We will argue here that: a) the use of this adapted measure may not have posed a problem and may have provided interesting information in earlier studies that focused on instructional formats that reduced learners’ mental effort investment in processes that are not effective for learning (i.e., extraneous cognitive load), and b) the use of the adapted measure may be problematic in other cases, for example in studies on instructional formats that seek to stimulate learners’ mental effort investment in processes that foster learning (i.e., germane cognitive load). We will first discuss cognitive load theory and the importance of measuring cognitive load, after which we will discuss the original and adapted instructional efficiency measures and provide guidelines regarding the conditions under which the different measures can be used.

### *Cognitive Load Theory and Educational Research*

Because there is a large body of literature describing cognitive load theory in detail (see e.g., Sweller, 1988, 2005; Sweller, Van Merriënboer, & Paas, 1998; Van Merriënboer & Sweller, 2005) we will only shortly summarize the most important elements of the theory. The central tenet of cognitive load theory is that in order to be effective, instruction should be

designed in alignment with learners' cognitive architecture. Human cognitive architecture is held to consist of a limited-capacity working memory and a virtually unlimited long-term memory. Working memory capacity is considered limited to holding seven plus or minus two elements or chunks of information (Miller, 1956), and even less when processing information (Cowan, 2000). It has also been suggested that it is not so much the number of information elements as it is the time that an information element could remain active without rehearsing it that defines working memory capacity (Baddeley, 1986). However, according to Cowan this is a controversial and unsettled issue, which is "nearly intractable because any putative effect of the passage of time on memory for a particular stimulus could instead be explained by a combination of various types of proactive and retroactive interference from other stimuli." (Cowan, 2000, p. 88). Either way, because information has to pass through working memory before it can be consolidated in long-term memory, the limited capacity of working memory can be considered the bottleneck for learning. Learning is considered to take place via schema construction, elaboration, and automation. Schemata are constructed by combining new, single information elements, into one larger element (schema), and are elaborated by adding new information elements to existing schemata. Through intensive and consistent practice, schemata can become automated, that is, executed without controlled processing.

Cognitive load theory distinguishes three types of cognitive load imposed by learning materials: intrinsic, extraneous, and germane cognitive load. Intrinsic cognitive load refers to the number of interacting information elements the material contains. Interactive elements have to be processed simultaneously in working memory for learning to commence (Chandler & Sweller, 1991), and consequently, learning new material that contains a high number of interacting elements will impose a high working memory load. Intrinsic load is influenced by expertise or prior knowledge: High prior knowledge learners have incorporated several or all

of the information elements presented in the learning materials in a cognitive schema.

Because this schema can be treated as a single element in working memory, intrinsic load decreases with increasing prior knowledge. Extraneous and germane load, in contrast, are not considered inherent to the learning material, but are imposed by the design of that material.

When the cognitive load imposed by the design is ineffective or detrimental for learning, it is called *extraneous* cognitive load; when it is effective for learning it is referred to as *germane* cognitive load (Sweller et al., 1998).

Not surprisingly given the central tenet of cognitive load theory, the aim of researchers in the field of cognitive load theory has been to engineer the instructional control of cognitive load to provide the means to optimize cognitive load in learning arrangements. Please note the use of the term ‘optimize’; a common misunderstanding is that the aim is to reduce the load imposed on the learners’ working memory. This is not the case, rather, the aim is to avoid both overload and underload, because learning deteriorates under these conditions (Teigen, 1994; Young & Stanton, 2002), as well as to reduce the load imposed by processes that do not contribute to or hamper learning (i.e., extraneous load), and increase the load imposed by processes that foster learning (i.e., germane load; see also Paas, Renkl, & Sweller, 2003, 2004; Paas & Van Gog, 2006). For a long time intrinsic load was considered unalterable by instruction, but recently some research effort has been devoted to finding techniques to manage this load (e.g., Pollock, Chandler, & Sweller, 2002), which may be unavoidable in situations where tasks are extremely complex in order for learning to commence. Many ‘early’ research efforts have been devoted to finding instructional formats that reduce extraneous load, because it is imposed by processes that do not contribute to learning. Although extraneous load does not have a significant negative impact on learning when tasks are low in intrinsic load, it does so when tasks are high in intrinsic load. Therefore, reducing extraneous load is considered imperative for such tasks (Van Merriënboer & Sweller, 2005).

Examples of extraneous load reducing measures are: Using integrated text and diagram formats instead of split-source formats (split-attention effect; e.g., Chandler & Sweller, 1991), avoiding presentation of redundant information (redundancy effect; e.g., Chandler & Sweller, 1991), and making use of multiple modalities to present mutually referring textual and pictorial information (modality effect; e.g., Mousavi, Low, & Sweller, 1995). A more recent development is the search for instructional strategies that reduce extraneous and increase germane cognitive load (Paas, Renkl, et al., 2003; Sweller et al., 1998; Van Merriënboer & Sweller, 2005). When extraneous load is lowered, learners may have cognitive capacity left that can be invested in processes that impose a germane load, that is, which do contribute to learning. However, they are unlikely to engage in such activities spontaneously. Hence, research efforts have been directed towards identifying instructional techniques that stimulate learners to invest cognitive resources in activities relevant for learning. Examples of germane load inducing measures are imagination or self-explanation assignments to process the solution steps in worked examples more deeply (e.g., Atkinson, Renkl, & Merrill, 2003; Cooper, Tindall-Ford, Chandler, & Sweller, 2001), or study assignments of high-variability sequences of worked examples (e.g., Paas & Van Merriënboer, 1994). Another more recent development is that studies increasingly take into account the learners' levels of prior knowledge (see Kalyuga, Ayres, Chandler, & Sweller, 2003) or changes in the learners' level of knowledge during instruction (Atkinson et al., 2003; Kalyuga, 2006; Van Gog, Paas, & Van Merriënboer, 2006, in press).

#### *Measuring Cognitive Load in Educational Research*

Different subjective and objective techniques are available for measuring cognitive load during (online) or after (offline) task performance. We will only shortly describe those techniques; for an in-depth discussion of cognitive load measurement the reader is referred to Paas, Tuovinen, Tabbers, and Van Gerven (2003). Objective on-line measures include

physiological measures such as heart-rate variability (Paas & Van Merriënboer, 1994; note though, that they found this measure to be quite intrusive as well as insensitive to subtle fluctuations in cognitive load), psychophysiological measures such as eye-movement data (Van Gerven, Paas, Van Merriënboer, & Schmidt, 2004), secondary-task procedures (Brünken, Plass, & Leutner, 2003), as well as the more common time-on-task or response time measures.

Subjective measures usually consist of rating scales. The NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988) measures task load on several dimensions by subjective ratings participants provide of perceived: performance, effort, frustration, and mental, physical and temporal demands. Note that this measure of task load encompasses more than cognitive load, although it has been used in cognitive load research (see e.g., Gerjets, Scheiter, & Catrambone, 2004, 2006; Kester, Lehnen, Van Gerven, & Kirschner, 2006). The unidimensional 9-point symmetrical category rating scale developed by Paas (1992) presents participants with one item on which they have to translate their perceived amount of invested mental effort into a numerical value. Mental effort is defined as “the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task; thus, it can be considered to reflect the actual cognitive load” (Paas, Tuovinen, et al., 2003, p. 64). This 9-point mental effort rating scale has shown good internal consistency. For example, Paas, Van Merriënboer, and Adam (1994) report high reliability coefficients (Cronbach’s alpha) of 0.90 and 0.82 for this rating scale in two studies (i.e., Paas, 1992, and Paas & Van Merriënboer, 1994), Kester, Kirschner, & Van Merriënboer (2004) report a reliability of 0.91, and analyses of data of Van Gog et al. (2006, in press) reveal a reliability of .83 and .93. Moreover, the concurrent validity of the 9-point mental effort rating scale is shown by its sensitivity to detect variations in task complexity (see Paas et al., 1994), and to variations in intrinsic load during task performance (Ayres,

2006).

It should be noted though, that many authors have adapted the 9-point mental effort rating scale (see Table 1), using 5-point or 7-point scales, or asking participants not to rate mental effort, but to rate how difficult they perceived the task to be. Although the concepts of invested mental effort and perceived task difficulty are to some extent related, asking students to rate how much mental effort they invested in completing a task versus how difficult they perceived a task to be, are two different questions that can lead to different interpretations. The first question (invested mental effort) pertains to a process, and the perception thereof will likely involve more aspects than only the task itself, whereas the second question (perceived task difficulty) pertains mainly to the task. This difference will become more pronounced at the extreme end of the scale: Research has shown that when a learner perceives a problem to be extremely difficult, s/he may not be motivated to invest much effort in this problem (see e.g., Bandura, 1989; Cennamo, 1993; Paas, Tuovinen, Van Merriënboer, & Darabi, 2005; Pintrich & Schrauben, 1992). As such, questions of perceived task difficulty or invested mental effort will presumably lead to non-equivalent ratings.

Because the NASA-TLX is a multidimensional scale, it is usually administered only once, at the end of a learning or test phase. The unidimensional 9-point mental effort rating scale can be used for multiple measurements during an experiment, that is, a single measurement after each task. Consequently, this scale allows for a more fine-grained analysis of the invested mental effort, and allows for estimations of internal consistency (see Paas et al., 1994). Because the subjective cognitive load measures are usually recorded after each task (9-point mental effort rating scale) or after a series of tasks has been completed (NASA-TLX), they are usually considered to be off-line measures. However, Ayres (2006) showed that the 9-point mental effort rating scale can be used during task performance as well. Whereas on-line measures are very suitable for studying fluctuations in cognitive load during



task performance, subjective off-line data are often easier to collect and analyze, and seem to provide a good indication of the overall cognitive load a task imposed (Paas, 1992; Paas & Van Merriënboer, 1993; Paas, Tuovinen, et al., 2003). Note that all measures, whether objective or subjective, online or offline, provide indications of cognitive load as a whole, rather than of its constituent aspects (i.e., intrinsic, extraneous, germane). Although some attempts have been made, at present no measure exists that can distinguish between intrinsic, extraneous and germane load imposed by a task. However, information on variations in the different constituent aspects of overall cognitive load can be obtained by varying only one aspect of load (e.g., intrinsic load) and holding the others constant (cf. Ayres, 2006), or by holding intrinsic load constant and regarding mental effort scores in relation to performance scores, because extraneous and germane load are defined as being detrimental or beneficial for learning, respectively. This issue is addressed in more detail in the following section.

*Instructional Efficiency: How the Relationship between Mental Effort and Performance Measures Can Be Used to Evaluate Instruction*

As indicated in the previous section, mental effort measures reflect the cognitive capacity allocated to accommodate to the demands imposed by the task (i.e., the actual cognitive load). The task can be either an instructional task (i.e., during the learning phase) or a test task designed to measure learning outcomes. In this section we will discuss what kind of information mental effort measurements during and after the learning phase can provide, in relation to measures of performance during and after the learning phase. Our use of the term ‘performance’ here, is aligned with the most common use of the term in the field, that is, as an evaluation of the learning outcomes in terms of a scoring of correct answers, or errors made<sup>1</sup>.

*Mental Effort Measurements During the Learning Phase*

As mentioned before, next to the intrinsic load imposed by the number of interacting information elements of a task, there is extraneous and germane load imposed by the design

of the task. Extraneous cognitive load is defined as being ineffective for learning, that is, it does not contribute to, or may even harm learning outcomes. Germane cognitive load, in contrast is defined as being effective for learning (see Sweller et al., 1998). In other words, when a task is high in extraneous load, it requires high levels of mental effort investment, but this effort is invested in processes that do not contribute to or even hamper learning. On the other hand, when a task is high in germane load, it also requires high levels of mental effort investment, but this effort is invested in processes that foster learning. The main aim of cognitive load theory research, is to find instructional formats that reduce extraneous and increase germane load, which challenge the learner to invest effort, but in processes that are relevant for learning.

However, when we know the mean rating of invested mental effort during the learning phase of a group of learners, we do not know whether they invested this effort in processes that were relevant (germane) or irrelevant (extraneous) for learning, if we do not regard their learning outcomes. For example, when we know that two groups of learners (A and B), of equal levels of expertise (i.e., materials will impose a comparable intrinsic load), both indicated that they on average invested 'high mental effort' (i.e., a mean rating of 7 on a 9-point scale) during learning with two different instructional formats A (group A) and B (group B), we do not know very much. However, if we would also know that the learning outcomes of group B were higher than those of group A, we can conclude that the high level of effort invested by group B in studying format B was apparently more effective for learning than the same level of effort invested by group A studying format A.

Nonetheless, measuring mental effort during the learning phase may provide some interesting information to researchers and instructional designers. For example, if, given equal intrinsic load, two instructional formats (A and B) lead to equal learning outcomes, but format A requires significantly less investment of mental effort during learning than format B, then it

might be more desirable to implement format A in practice. However, the phrase “equal learning outcomes” is of pivotal importance here, and should encompass more than a performance score, as we will show in the next paragraph.

### *Mental Effort Measurements During the Test Phase*

We already mentioned the relationship between cognitive load and expertise: With increasing levels of expertise or prior knowledge, individuals increasingly acquire, elaborate, and/or automate task-relevant cognitive schemata, which decreases the intrinsic load imposed by completing such a task for learners with more prior knowledge compared to learners with less prior knowledge. Therefore, individuals with more expertise are able to attain equal or higher levels of performance with less investment of mental effort (cf. Anderson & Fincham, 1994; Yeo & Neal, 2004). This also implies that learners who managed to gain more knowledge during the learning phase as a result of a more effective type of instruction, should experience less cognitive load when completing test tasks than learners who received a less effective instructional format (who, as a result, learned less). For example, consider two groups of learners (X and Y) learning with two instructional formats (X and Y, respectively), that both managed to reach a high mean test performance score of 80% correctly answered. Group X reported a mean mental effort invested in the test tasks of 4 (out of 9) and group Y of 7 (out of 9). If we would rely only on performance scores, we would (erroneously) conclude that both instructional formats were equally effective. However, if we would take a closer look and take into account the mental effort scores, it becomes clear that the quality of learning of group X, was higher than that of group Y; in other words, learners in group X have gained more expertise. This example illustrates the importance of relying not only on measures of test performance to draw conclusions about the effects of instructional conditions, but to consider mental effort invested in solving the test problems as an important element of the learning outcomes. Mental effort in combination with performance measures,

will provide us with a better, more subtle, indicator of the quality of learning outcomes, that is, in terms of the efficiency of cognitive schemata acquired, elaborated, or automated as a result of instruction, and hence, with a better indicator of the quality of different instructional conditions. Nevertheless, as can be seen in Table 1, many studies do not even measure mental effort during the test phase.

*Efficiency: Combining Mental Effort and Performance Measures*

*Original instructional efficiency measure.* In searching for a single measure to determine the relative efficiency of instructional conditions in terms of learning *outcomes*, Paas and Van Merriënboer (1993) developed a computational approach for combining measures of test performance with measures of mental effort invested to attain this test performance. Performance and mental effort scores on the test are first standardized. Then, mean standardized test performance (P) and test mental effort (E) scores attained by learners

in a certain condition are entered into the following formula:  $Efficiency = \frac{zP_{test} - zE_{test}}{\sqrt{2}}$ .

Although developed for measuring the efficiency of instructional conditions, the combination of performance and mental effort is also indicative of expertise (the quality of acquired cognitive structures), and hence, this measure can also be used outside instructional situations to evaluate learners' relative levels of expertise/acquired cognitive structures (cf. Van Gog, Paas, & Van Merriënboer, 2005).

Although Paas and Van Merriënboer (1993) proposed the use of performance and mental effort in the efficiency measure, other measures of cognitive load might also be used. For example, since time-on-task can be regarded as an objective measure of cognitive load, one could also define an efficiency measure based on test performance and time-on-task invested in the test. At least, this is the case when time-on-task is not restricted by the experimenter or instructor. Under conditions where time-on-task is restricted, it will no longer

have a very strong relation with cognitive load, and it may be more useful to rely on mental effort invested in the test, or to incorporate both performance, time-on-task, and mental effort invested in the test into the efficiency measure (cf. Tuovinen & Paas, 2004).

*Adapted instructional efficiency measure.* As is shown in Table 1, many studies have adopted, but simultaneously adapted the original instructional efficiency measure, by combining test performance with mental effort invested in the learning phase instead of the test phase:  $Efficiency = \frac{zP_{test} - zE_{learning}}{\sqrt{2}}$ . In fact, of the 37 studies in this Table after 1993, only 4 used the original measure! Whereas the original measure was developed for examining instructional efficiency in terms of learning *outcomes*, that is, to obtain a more subtle measure of the quality of the cognitive structures acquired *as a consequence of* instruction than performance scores only, the adapted measure examines instructional efficiency in terms of the learning *process*, that is, combining the effort investment during instruction with the test performance attained as a result of instruction. The latter, adapted, measure reflects a somewhat more general economic cost-benefit view of efficiency.

This difference between the original and the adapted measure has important implications, because mental effort invested in the learning phase and mental effort invested in the test phase are very different (at least from a research perspective; for the learner they may not be, s/he just rates the effort s/he invested in a task, whether it is a learning task or a test task). As mentioned before, it is very hard to interpret mental effort scores in the learning phase, because the exact processes in which a learner invests effort in the learning phase are often unknown, and the relative contributions of those processes to effort investment are unknown. What we do know, however, is that the design of the instruction in the different conditions will have a major influence, and hence, that these processes differ between conditions. The same argument holds for effects on time-on-task during the learning phase

when one would want to use this variable to measure efficiency rather than mental effort. In contrast, in the test phase, problems are identical for all students and mental effort is influenced mainly by their knowledge, which allows for a more unequivocal interpretation of the mental effort scores on the test phase across conditions.

As a result, there are significant differences in what is actually being measured by the original and adapted measures, that is, the first defines the construct of instructional efficiency in terms of learning *outcomes* (original), whereas the second defines the construct of instructional efficiency in terms of the learning *process* (adapted). This also implies that there are differences regarding the type of conclusions that can be drawn from these measures. We will return to this issue in more detail in the next section.

*Efficiency measure for adaptive task selection.* Efficiency measures have also found use in algorithms for adaptive, personalized task selection, as is shown in Table 1 (measure descriptions preceded by <sup>1</sup>). Because the amount of mental effort invested in a task combined with the performance on that task can be used to assess learners' acquired cognitive structures, this information can be used to dynamically adapt task sequences. That is, depending on efficiency scores on a certain task, the complexity or amount of support provided by a next task can be tailored to the learner's current knowledge state. This strategy ensures that each next task is in optimal alignment with the individual learner's cognitive architecture.

#### *Original Versus Adapted Instructional Efficiency Measures*

In this section, it is argued that although it measures a very different kind of efficiency than the original efficiency measure, the adapted efficiency measure might provide useful information in some circumstances, such as when the aim is to decrease extraneous load, but not in others, such as when the aim is to increase germane load.

*Efficiency Measures in Studies Aiming to Decrease Extraneous vs. Increase Germane Load*

In comparing instructional materials, the adapted measure seems most useful in situations where the aim is to reduce cognitive load during learning (cf. the ‘early’ research on cognitive load theory that aimed at the development of instructional techniques that reduce extraneous load). Consider for example a comparison of learning from conventional problem solving, which imposes a high extraneous load, with learning from studying worked examples, which reduces extraneous load for novice learners. However, when instruction aims to stimulate learners to invest high levels of effort in processes relevant for learning (cf. the ‘later’ research on cognitive load theory focusing on both reducing extraneous load and increasing germane load), the adapted measure does not seem very suitable. Consider for example a comparison of learning from studying regular (product-oriented) worked examples, which reduce extraneous load compared to conventional problem solving, with learning from studying process-oriented worked examples, which reduce extraneous load compared to conventional problem solving and increase germane load compared to regular, product-oriented worked examples for novice learners. Instructional formats that successfully substitute germane cognitive load for extraneous cognitive load, would require more investment of effort from learners during the learning phase than formats that only decrease extraneous load. Because the adapted measure uses only one aspect of learning outcomes, that is, a performance score, and it is highly unlikely that this increase in effort would be proportional to the increase in performance, efficiency according to the adapted measure would be lower in the germane load inducing condition. This might lead to the erroneous conclusion that this condition is less desirable than the other.

We will illustrate these points with empirical data. We will first provide an example of a condition under which the adapted measure can be useful, and then provide an example of a condition in which it is not, which will illustrate how the original and adapted efficiency measures can lead to very different conclusions when applied under the wrong conditions.

These examples are based on parts of the data sets of two empirical studies by Van Gog et al. (2006, in press)<sup>2</sup>.

Compared to learning from conventional problem solving, studying carefully designed worked examples would be hypothesized to reduce extraneous load, which would lead to less investment of mental effort in the training, higher performance on the test, and less/equal effort investment in the test. A summary of such data from an empirical study (Van Gog et al., 2006) is presented in Table 2. If we analyze efficiency using the adapted and original measures, we find that both lead to the conclusion that studying worked examples is more efficient: adapted,  $t(28) = 3.642, p = .001$  (two-tailed); original,  $t(28) = 2.197, p = .036$  (two-tailed). So when the aim is to study formats that reduce extraneous load, and these formats are successful in doing so, the adapted measure can be used to come to the same conclusion, but one needs to realize that it provides information about efficiency in terms of the process of learning rather than learning outcomes.

A germane load inducing measure combined with worked examples, could be the use of process-oriented worked examples, which aim to increase students' understanding of the solution procedures presented in the examples (see Van Gog, Paas, & Van Merriënboer, 2004). As mentioned above, studying regular, product-oriented worked examples, would be hypothesized to reduce extraneous load (compared to conventional problem solving). Studying process-oriented worked examples, however, would be hypothesized to not only reduce extraneous load (compared to conventional problem solving) but also increase germane load (compared to product-oriented worked examples) for novice learners. This would be expected to lead to more investment of mental effort in the learning phase when studying process-oriented worked examples, but because this effort is invested in processes relevant for learning, one would also expect an equal/higher performance score on the test, and less investment of mental effort in the test compared to product-oriented worked



examples. A summary of such data from an empirical study (Van Gog et al., in press) is presented in Table 3. Here we find that if we use the adapted measure, we have to conclude that both formats are equally efficient:  $t(79) = .038, p = .970$  (two-tailed). If we look at the learning outcomes, however, using the original measure, we find that studying process-oriented worked examples is more efficient for learning than studying product-oriented worked examples:  $t(79) = 2.073, p = .041$  (two-tailed). So, in studies that do not aim to *lower* investment of effort during learning, but rather *increase* it to increase learning outcomes, and are successful in doing so, the adapted measure that relies on mental effort during learning is not very useful.

#### *Efficiency Measures in Studies on Adaptive Task Selection*

In addition, in studies on adaptive task selection, there is another argument for not using the adapted efficiency measure. As Table 1 shows, some studies on adaptive task selection have not only used an efficiency measure for adaptive task selection, but also applied the adapted instructional efficiency measure to draw conclusions about the differential effects of conditions, that is, they used the standardized mean mental effort invested in the learning tasks combined with test performance. This is problematic, because adaptive task selection based on efficiency implies that one is directly influencing the mental effort that needs to be invested in each next learning task (this is kept at an optimal level), whereas this is not the case in the other conditions. Using the mental effort invested in the learning tasks in the efficiency formula to compare conditions, then, would be inappropriate in these studies, because this measure has been influenced by the instructional manipulation. Therefore, using the *original* instructional efficiency measure would also be more appropriate in studies such as these.

#### *Efficiency Measures: More General Arguments*

A final, more general argument for using the original efficiency measure instead of the

adapted measure is related to the differences between instructional conditions that are characteristic for instructional research. As mentioned previously, instructional conditions normally differ not only with regard to task format but also with regard to the type of cognitive processes required (cf. germane load inducing learning activities), completion time needed, and so on. As it is difficult or impossible to disentangle these factors it is not clear which of them, and to what extent, has contributed to the invested mental effort in the learning phase. This also goes for efficiency measures based on performance and time-on-task, because the same problems for interpreting mental effort invested in the learning phase also hold for interpreting time-on-task invested in the learning phase. In contrast, in the test phase, problems are identical for all students, which allows for a more unequivocal interpretation of the mental effort (and time-on-task) scores. Research has also shown that data acquired during the learning phase might easily lead to incorrect conclusions about learning outcomes. Among others, Bjork (1999) shows that processes that are helpful to learning may lead to slow performance gains during the learning phase (and/or high cognitive load; cf. germane load), whereas processes that lead to high performance gains during the learning phase may not actually have effects on learning as measured by learning outcomes.

Regarding the use of a perceived difficulty rating in the adapted efficiency measure, which many authors do (see Table 1), researchers should be aware that they are not only diverting from the original measure, but may also be measuring yet another construct than the adapted measure based on mental effort would. As mentioned previously, the question to rate perceived task difficulty or invested mental effort, will lead to different interpretations and outcomes, especially at the extreme end of the rating scale. When a learner perceives a problem to be extremely difficult, s/he may not be motivated to invest much effort in this problem. Hence, under such conditions, the outcomes of the effort and difficulty questions that are inserted in the efficiency formula are completely opposite. Using a difficulty rating in

the original definition of the efficiency measure seems even less sensible, because the efficiency of an instructional condition would then be considered high in case of high performance on a perceived easy test task, which does not seem indicative of quality of learning outcomes –high performance on a difficult task would be, but this would not be efficient according to the formula.

In sum, when analyzing efficiency of instructional conditions, we would strongly recommend researchers to always use the original measure based on mental effort (rather than difficulty) ratings and performance scores in the test phase. In certain situations, the adapted measure may provide some useful additional information, but it: a) should not replace the original measure, b) should not be used when instructional conditions that decrease extraneous and increase germane load are compared to conditions that only decrease extraneous load, and c) should not be used when one of the instructional conditions incorporated an efficiency measure for adaptive task selection.

### Discussion

This article revisited the ideas behind the original measure of instructional efficiency, in an attempt to clarify that what has become known as *the* instructional efficiency measure of Paas and Van Merriënboer (1993), is in fact an adaptation. This should not be taken to imply that the studies summarized in Table 1 are suspect, for two reasons illustrated by the examples based on empirical data: 1) the use of the adapted measure may be useful when the goal is to reduce extraneous load, and 2) in studies that seek to enhance germane load, the use of the adapted measure might lead to the conclusion of no significant differences (and in theory it might even lead to an opposite effect). In this case, researchers may have calculated the adapted efficiency measure, but are not likely to report it because it does not add interesting results, and therefore the studies for which our argument would have consequence, are not likely to be in this Table.

The studies in Table 1 served the purpose of illustrating that many authors only seem to know the adapted measure, and not the original one, which has hardly been used in educational research. Moreover, because many authors in this Table refer to the original measure but use an adapted measure, they do not seem to be aware of the fact that they use an adapted measure. Although this adapted measure can be argued to measure some kind of efficiency of instructional conditions (i.e., in terms of the learning process), it does not measure the efficiency of instructional conditions in the original definition (i.e., in terms of learning outcomes), which has consequences both for application of this adapted measure and conclusions that can be drawn from it. This article intended to make educational researchers more aware of the differences between the original and adapted measures and the conditions under which they can and cannot be applied.

Here, however, we should ask a logical question that we did not address yet: Why did the measure become adapted and then accepted so quickly? The answer may lie in a rhetoric question: *What's in a name?* The term “efficiency of instructional conditions” (or “instructional efficiency”) itself may well be one of the reasons why the measure became adapted in the first place, and why the adaptation was not really noticed. It is not so difficult to imagine that this term is interpreted as referring to the efficiency of the *process* of instruction, because, as mentioned before, this reflects a more general, economic notion of efficiency in terms of process costs and attained benefit. Somehow, regarding instructional efficiency as the amount of effort one has to invest when one is confronted with the instruction (i.e., in the learning phase) in relation to performance attained as a result of the instruction, instead of as referring to the *results* of instruction (i.e., *learning outcomes*), does seem more intuitive. That the term instructional efficiency indeed seems to evoke some ambiguity in interpretation is also reflected in the alternative terms that the authors who did realize they were using adapted versions started to use. For example, even though most

authors simply use the term instructional efficiency, the adapted measure has also been called “training efficiency” (e.g., Salden, Paas, Broers, & Van Merriënboer, 2004; Van Gerven, Paas, Van Merriënboer, & Schmidt, 2002), which seems indeed more accurate for the adapted measure and helps to distinguish the original measure from the adapted one, and “learning efficiency” (e.g., Kester et al., 2006), which is still somewhat ambiguous in whether it refers to a process or outcome. The original measure has also been called “performance efficiency” (e.g., Kester, Kirschner, & Van Merriënboer, 2006; Kester, Lehnen, et al., 2006), and this term has also been used in situations where the efficiency measure was used as a measure of relative expertise (Van Gog, et al., 2005). The term “mental efficiency” has been used in adaptive task selection settings (e.g., Camp, Paas, Rikers, & Van Merriënboer, 2001; Salden et al., 2004), where efficiency scores per task served as input for selection of a next task. Both “performance efficiency” and “mental efficiency” seem to more clearly convey the meaning of the original measure, that is, stressing the efficiency of acquired cognitive structures as a consequence of instruction, than “instructional efficiency” does.

Interesting directions for future research would be to study the viability of alternative efficiency measures. For example, other subjective ratings such as NASA-TLX (cf. Kester, Lehnen, et al., 2006) might be used. As mentioned before, it can be very useful to incorporate time-on-task on the test as an objective measure of cognitive load (when time-on-task is not restricted), but it might also be possible to use objective measures such as secondary task or eye movement data. Moreover, efficiency measures incorporating more variables than performance and cognitive load can be conceived of (cf. Tuovinen & Paas, 2004), for example, measures of motivation have been included by some (Hummel, Paas, & Koper, 2004). However, in exploring these alternative measures, care should be taken, to define clearly what these efficiency measures intend to measure. Moreover, attention for the nature of the relationships between the different measures included in the efficiency measure is

necessary, for example, motivation is likely to influence effort investment (cf. Paas et al., 2005), so it is questionable whether adding it as a separate variable to the efficiency measure has added value.

In addition, it might also be interesting to investigate the use of the efficiency measure in longitudinal studies of learning or expertise development (cf. Van Gog et al., 2005). The efficiency measure is relative, that is, it allows for comparing the efficiency of one group/individual with another group/individual, and does not position an individual on an exact point of the continuum from novice to expert. However, given a set of standard test tasks, it might be possible and useful to take ones own past efficiency on these tasks as a reference point to measure learning (i.e., increase in efficiency) over time. Again, because not only performance would be regarded, but also invested effort (or some other measure such as time-on-task), more subtle improvements in learning outcomes could be detected. This would not only give learners and their teachers, instructors, mentors, or coaches a more accurate picture of learners' individual progress, but might also be more motivating for learners.

## References

- Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1322-1340.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*, 774-783.
- Ayres, P. (2006). Using subjective measures to detect variations of intrinsic load within problems. *Learning and Instruction*, *16*, 389-400.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, *25*, 725-739.
- Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher and A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.
- Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*, 53-61.
- Camp, G., Paas, F., Rikers, R., & Van Merriënboer J. J. G (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior*, *17*, 575-595.
- Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology*, *95*, 629-640.
- Cennamo, K. S. (1993). Learning from video: Factors influencing learners' preconceptions and invested mental effort. *Educational Technology Research and Development*, *41*(3), 33-45.

- Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research, 15*, 345-367.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.
- Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied, 7*, 68-82.
- Corbalan, G., Kester, L., & Van Merriënboer, J. J. G. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science, 34*, 399-422.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences, 24*, 87-114.
- Cuevas, H. M., Fiore, S. M., & Oser, R. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science, 30*, 433-464.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33-58.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction, 16*, 104-121.
- Halabi, A. K., Tuovinen, J. E., & Farley, A. A. (2005). Empirical evidence on the relative efficiency of worked examples versus problem-solving exercises in accounting principles instruction. *Issues in Accounting Education, 20*, 21-32.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.),



- Human mental workload* (pp. 139-178). Amsterdam: Elsevier Science.
- Hasler, B. A., Kersten, B., & Sweller, J. (in press). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology*.
- Hummel, H. G. K., Paas, F., & Koper, E. J. R. (2004). Cueing for transfer in multimedia programmes: Process worksheets vs. worked-out examples. *Journal of Computer Assisted Learning, 20*, 387-397.
- Kalyuga, S. (2006). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J., Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*, 1-17.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126-136.
- Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5-23.
- Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors, 46*, 567-581.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development, 53*(3), 83-93.
- Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2004). Timing of information

- presentation in learning statistics. *Instructional Science*, 32, 233-252.
- Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31, 167-185.
- Kester, L., Lehnen, C., Van Gerven, P. W. M., & Kirschner, P. A. (2006). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior*, 22, 93-112.
- Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, 88, 49-63.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psychological Review*, 63, 81-97.
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99-113.
- Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development*, 53(3), 35-45.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319-334.
- Ngu, B. H., & Rethinasamy, S. (2006). Evaluating a CALL software on the learning of English prepositions. *Computers & Education*, 47, 41-55.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology*, 84, 429-434.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design:

- Recent developments. *Educational Psychologist*, 38, 1-4.
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32, 1-8.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63-71.
- Paas, F., Tuovinen, J. E., Van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, 53(3), 25-34.
- Paas, F., Van Gerven, P. W. M., & Wouters, P. (in press). Instructional efficiency of animation: Effects of interactivity through mental reconstruction of static key frames. *Applied Cognitive Psychology*.
- Paas, F., & Van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, 16, 87-91.
- Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, 35, 737-743.
- Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122-133.
- Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive-load in instructional research. *Perceptual and Motor Skills*, 79, 419-430.
- Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive

- engagement in classroom academic tasks. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 149–183). Hillsdale, NJ: Erlbaum.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61-86.
- Salden, R. J. C. M., Paas, F., Broers, N. J., & Van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in Air Traffic Control training. *Instructional Science, 32*, 153-172.
- Salden, R. J. C. M., Paas, F., Van der Pal, J., & Van Merriënboer, J. J. G. (2006). Dynamic task selection in flight management system training. *International Journal of Aviation Psychology, 16*, 157-174.
- Salden, R. J. C. M., Paas, F., & Van Merriënboer, J. J. G. (2006). Personalised adaptive task selection in air traffic control: Effects on training efficiency and transfer. *Learning and Instruction, 16*, 350-362.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-295.
- Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory and Psychology, 4*, 525–547.
- Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257-287.
- Tuovinen, J., & Paas, F. (2004). Exploring multidimensional approaches to the efficiency of

- instructional conditions. *Instructional Science*, 32, 133-152.
- Tuovinen, J. E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, 91, 334-341.
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., Hendriks, M., & Schmidt, H. G. (2003). The efficiency of multimedia learning into old age. *British Journal of Educational Psychology*, 73, 489-505.
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2002). Cognitive load theory and aging: Effects of worked examples on training efficiency. *Learning and Instruction*, 12, 87-105.
- Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology*, 41, 167-174.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science*, 32, 83-98.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, 19, 205-221.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, 16, 154-164.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (in press). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*.
- Van Merriënboer, J. J. G., Schuurman, J. G., De Croock, M. B. M., & Paas, F. (2002).

- Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training. *Learning and Instruction, 12*, 11-39.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147-177.
- Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice, and performance: Effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology, 89*, 231-247.
- Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *The Journal of Experimental Education, 67*, 197-217.
- Yeung, A. S., Jin, P., & Sweller, J. (1997). Cognitive load and learner expertise: split-attention and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology, 23*, 1-21.
- Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A new explanation for the effects of mental underload on performance. *Human Factors, 44*, 365-375.

## Footnotes

<sup>1</sup> Though we would ideally prefer performance measures to incorporate mental effort (or other cognitive load) measures.

<sup>2</sup> The Van Gog et al. (2006) study consisted of 4 conditions, data of 2 conditions are presented here. The Van Gog et al. (in press) study consisted of two alternating learning and test phases: Learning1 – Test 1 – Learning 2 – Test 2. The data presented here are only those of the first half of the experiment (i.e., Learning 1 and Test 1), in which students studied either process-oriented or product-oriented worked examples in the learning phase.

Table 1

*Type of Efficiency Measures Applied in Studies from 1993-2006*

<b>Studies</b>	<b>Learning Phase</b>	<b>Test Phase</b>	<b>Efficiency Measure</b>
Paas & Van Merriënboer (1993)		ME (9-pt effort); P	* ME Test; P Test
Paas & Van Merriënboer (1994)	ME (9-pt effort)	ME (9-pt effort); P	* ME Test; P Test
Cerpa, Chandler, & Sweller (1996)	ME (9-pt difficulty)	P	ME Learn; P Test
Marcus, Cooper, & Sweller (1996)	ME (7-pt difficulty)	P	ME Learn; P Test
Tindall-Ford, Chandler, & Sweller (1997)	ME (7-pt effort)	P	ME Learn; P Test
Yeung, Jin, & Sweller (1997)	ME (9-pt difficulty)	P	ME Learn; P Test
Kalyuga, Chandler, & Sweller (1998)	ME (7-pt difficulty)	P	ME Learn; P Test
Kalyuga, Chandler, & Sweller (1999)	ME (7-pt difficulty)	P	ME Learn; P Test
Tuovinen & Sweller (1999)	ME (9-pt effort)	ME (9-pt effort); P	ME Learn; P Test
Yeung (1999)	ME (9-pt difficulty)	P	ME Learn; P Test
Kalyuga, Chandler, & Sweller (2000)	ME (7-pt difficulty)	P	ME Learn; P Test
Camp, Paas, Rikers, & Van Merriënboer (2001)	ME (5-pt effort); P		<sup>1</sup> ME Learn; P Learn
			ME Learn; P Learn
Kalyuga, Chandler, & Sweller (2001)	ME (7-pt difficulty)	P	ME Learn; P Test
Kalyuga, Chandler, Tuovinen, & Sweller (2001)	ME (9-pt difficulty)	P	ME Learn; P Test
Cuevas, Fiore, & Oser (2002)	ME (7-pt difficulty)	P	ME Learn; P Test
Pollock, Chandler, & Sweller (2002)	ME (7-pt difficulty; 9-pt difficulty)	P	ME Learn; P Test
Van Gerven, Paas, Van Merriënboer, & Schmidt (2002)	ME (9-pt difficulty)	P	ME Learn; P Test
Van Merriënboer, Schuurman, De Croock, & Paas (2002)	ME (9-pt effort)	P	ME Learn; P Test
Carlson, Chandler, & Sweller (2003)	ME (7-pt effort); P	P	ME Learn; P Learn
			ME Learn; P Test
Van Gerven, Paas, Van Merriënboer, Hendriks, & Schmidt	ME (9-pt difficulty)	P	ME Learn; P Test



(2003)			
Hummel, Paas, & Koper (2004)	ME (9-pt effort); TT; MOT; P	P	<sup>2</sup> ME Learn; MOT Learn; TT Learn; P Learn  <sup>2</sup> ME Learn; MOT Learn; TT Learn; P Test ME Learn; P Test
Kalyuga, Chandler, & Sweller (2004)	ME (7-pt difficulty; 9-pt difficulty)	P	ME Learn; P Test
Kester, Kirschner, & Van Merriënboer (2004)	ME (9-pt effort)	P	ME Learn; P Test
Moreno (2004)	ME (10-pt difficulty)	P	ME Learn; P Test
Salden, Paas, Broers, & Van Merriënboer (2004)	ME (5-pt effort); TT; P	P	<sup>1</sup> ME Learn; P Learn  <sup>2</sup> ME Learn; TT Learn; P Test ME Learn; P Test
Halabi, Tuovinen, & Farley (2005).	ME (5-pt effort)	P	ME Learn; P Test
Kalyuga & Sweller (2005)	ME (9-pt difficulty); P		<sup>1</sup> ME Learn; P Learn
Moreno & Valdez (2005)	ME (7-pt difficulty)	P	ME Learn; P Test
Corbalan, Kester, & Van Merriënboer (2006)	ME (7-pt effort); P		ME Learn; P Learn
Kalyuga (2006)	ME (9-pt difficulty); P	P	<sup>1</sup> ME Learn; P Learn
Kester, Kirschner, Van Merriënboer (2006)	ME (9-pt effort)	ME (9-pt effort); P	ME Learn; P Test ME Learn; P Test
Kester, Lehnen, Van Gerven, & Kirschner (2006)	ME (NASA-TLX)	ME (NASA-TLX); P	* ME Test; P Test ME Learn; P Test
Ngu & Rethinasamy (2006)	ME (7-pt effort)	P	ME Test; P Test ME Learn; P Test
Salden, Paas, Van der Pal, & Van Merriënboer (2006)	ME (5-pt effort); P	P	<sup>1</sup> ME Learn; P Learn

Salden, Paas, & Van Merriënboer (2006)	ME (5-pt effort); P	ME (5-pt effort); P	ME Learn; P Test <sup>1</sup> ME Learn; P Learn
Hasler, Kersten, & Sweller (in press)	ME (9-pt effort)	P	<sup>2</sup> ME Test; P Test ME Learn; P Test
Paas, Van Gerven, & Wouters (in press)	ME (9-pt effort)	ME (9-pt effort);P	* ME Test; P Test
Van Gog, Paas, & Van Merriënboer (in press)	ME (9-pt effort)	ME (9-pt effort);P	* ME Test; P Test

*Note 1.* Studies are listed in chronological, then alphabetical order. *Note 2.* ME = Mental Effort; P = Performance; TT = Time on Task; MOT =

motivation; Learn = learning phase; Test = test phase; \* = Original measure; 1 = efficiency computed for adaptive/dynamic task selection during

training (i.e., not based on mean score of entire learning phase but per task); <sup>2</sup> = multidimensional measure (see Tuovinen & Paas, 2004). *Note 3.*

This Table provides an overview of scientific articles in which use of an efficiency measure was reported, and shows what kind of efficiency measure was used. This overview was generated as follows. We surveyed the list of publications referring to Paas and Van Merriënboer (1993) in Thomson Web of Science and Google Scholar on January 24, 2007, as well as in press articles we knew of, and included the empirical and published/in press articles that applied the efficiency measure in some way in this table (i.e., theoretical or review articles, dissertations, proceedings or conference papers, empirical articles referring to but not applying the efficiency measure, and unpublished not in press manuscripts presented on private websites were ignored).

Table 2

*Summary of Data per Condition (Van Gog et al., 2006)*

	Conventional problems ( $n = 16$ )		Worked examples ( $n = 14$ )	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Training mental effort (0-9)	5.22	1.56	4.01	1.82
Test performance (0-3)	1.36	.40	1.83	.48
Test mental effort (0-9)	4.97	1.59	4.71	1.54
Original efficiency	-.37	.85	.42	1.11
Adapted efficiency	-.54	.70	.61	1.02

Table 3

*Summary of Data per Condition (Van Gog et al., in press)*

	Product-oriented worked examples ( $n = 41$ )		Process-oriented worked examples ( $n = 41$ )	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Training mental effort (0-9)	3.18	1.36	3.50	1.57
Test performance (0-3)	1.70	.52	1.81	.45
Test mental effort (0-9)	5.43	1.12	4.83	1.47
Original efficiency	-.24	1.04	.24	1.05
Adapted efficiency	-.004	.97	.004	1.04