Progress in Science Education
2022, Vol. 5, No. 1, page 6-32
ISSN 2405-6057
DOI.10.25321/prise.2022.1275

---

## Special Issue
Tasks in Science Education

# An evidence-based Approach to Tasks in Science Education: Meta Analytical and other Quantitative results

Andreas Müller[1], Alexander Brown[2]

Received: February 2021 / Accepted: April 2021

## Structured Abstract

**Background:** Tasks and "Task culture" play a large role in science education and related areas (STEM: science, technology, engineering, mathematics). Establishing a reasoned overview of empirical (mainly meta-analytical) evidence relevant for the field provides useful background for research and practice.
**Purpose:** The main purpose of the present work is to present and discuss meta analytical and other quantitative results concerning various settings (e.g., classroom questions, homework), forms (worked examples, concept maps, experimental tasks) and "philosophies" (e.g., problem-based learning), as well as learner and classroom characteristics related to tasks in STEM education. Classroom practice plays a large role in the discussion.
**Sample/setting:** Meta-analyses and other studies covering several decades, (science) subjects, and geographical locations are reviewed. Age groups range from primary to tertiary education, with the main body in lower and upper secondary education. Where influences of age, gender, or academic level have been found, they are reported and discussed.
**Design and Methods:** Meta-analytical results are discussed in terms of Cohen $d$ (and of $g$, a small-sample correction) as a measure of effect sizes. Interpretative background of the latter is provided.
**Results:** Effect sizes for more than 20 independent and moderating variables related to tasks are reported and discussed. Notably, the role of feedback appears crucial in several cases, especially for homework. Some results may appear unexpectedly high (e.g., self questioning, $d = 0.64$) or low (e.g., inquiry based learning, $d = 0.31$), deserving more attention in practice and teacher education in the field.
**Conclusions:** Tasks have earned their place in science classrooms, but they must be used judiciously. The evidence based perspective on tasks reveals a rich variety of forms, purposes, and outcomes, confirming a subject specific "task culture" as a key element of science education, helpful and stimulating for researchers and practitioners. It shows how a deliberative, research informed practice can (*i*) go far beyond an understanding of tasks as dull, repetitive "drill and practice", restricted to surface knowledge and (*ii*) counter certain illusions (e.g., regarding classroom questions) and ideologies (e.g., regarding homework) related to tasks. We see "task culture", combined with an evidence-based perspective, as a promising framework for classroom practice, teacher education and research in the STEM field.
**Keywords:** *exercises, problems, tasks, task culture, evidence-based science education*

---

[1]Section Physique, Faculté des Sciences, Université de Genéve, [2]St Genis Pouilly, France
✉ andreas.mueller@unige.ch

# 1 Introduction

## 1.1 Rationale and Objective: An Evidence-Based Approach to the Essential Role of Tasks for Learning[1]

> *[T]asks are recognizable and consequential units of analysis in the development and implementation of curriculum, instruction, and assessment.* Tekkumru-Kisa et al. (2020, p. 607)

Tasks play a fundamental role at several key stages in a learner's journey and can thus serve as basic building blocks for educational quality and efficacy:

(i) Learning tasks, including "worked-out examples", are a teaching tool whose impact has been established in several studies (Renkl et al., 2002).
(ii) Practice and application tasks feature increasing independence and autonomy, as recommended by classical education theory (from Pestalozzi to the reform movement) and (almost universally) by modern education psychology (Levin & Arnold, 2006).
(iii) Diagnostic and evaluation tasks have at least three important educational functions: for working on misconceptions (Viennot, 1996; Vosniadou, 2013; Neidorf et al., 2020); for scientifically based competence models (Konsortium HarmoS Naturwissenschaften+, 2009, chap.7.4); and they promote transparency in expectations related to evaluation (Crahay, 2006, p. 129; Hattie, 2012, p. 129)

The following account of the importance of tasks draws on work in science, technology, engineering, mathematics (STEM) and other branches of discipline-based education. In the notion of "task culture", emphasis is placed on the multifaceted nature of their forms and functions and the benefits they display when used in a well-connected and systematic way.

In mathematics education, Steinbring (1989) proposed the notion of "task systems" as structured series of tasks to successfully develop concepts in probability. Stein and colleagues (1996; 2000) have introduced the mathematical task framework (MTF) as a set of "orchestrated" learning opportunities (Silver & Herbst, 2007), essential for teachers' work (and students' learning). An interesting aspect, among others, is that while the feature of "challenge" in tasks is often advocated for by some authors (CTGV, 1992, Birenbaum, 1996; Eccles et al., 2015) and might indeed be conducive for learning, it can just as easily be an obstacle, if not matched to students' competencies and perceptions (cf. sect. 1.4.3). To establish a balance in this regard is one objective in the MTF (Silver & Herbst, 2007). In a large scale mathematics education project (COACTIV, Löwen et al., 2013), tasks were shown to have an essential role, offering "a wealth of pedagogical, cognitive, communicative, comprehension-oriented, and other opportunities" for learning (Neubrand et al., 2013). Their "orchestration", taking into account teacher, student, and instructional factors in a systematic way is understood as an "opportunity structure for the learning process" (Fig. 1).

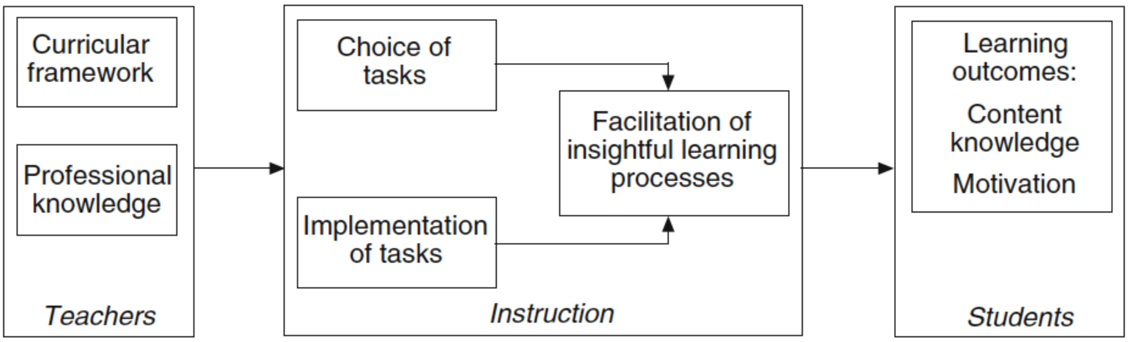

**Fig. 1.** Tasks as the basis of opportunity structures for insightful learning processes (adapted from Neubrand et al., 2013).

In science education, the field of educational activities and interventions is ascribed to the umbrella term "task culture" (in German: Aufgabenkultur, cf. sect. 1.2 for terminology). It plays a central role in practice of and research into science and mathematics education in Germany (BLK, 1997; Prenzel, 2008; Ostermeier et al., 2010; Konsortium HarmoS Naturwissenschaften+, 2010). In this context, tasks in science education (Kircher et al., 2009) are discussed and studied with respect to their contributions to learning, evaluation, and testing; intelligent exercising, applying, transferring, and justifying/evaluating; and as the "backbone" of vertical and horizontal acquisition of competences.

---

[1] Parts of this and of some other sections are based on Müller et al. (2013) and Müller (2018a, b; 2019a, b), where further discussion can be found.

Tekkumru-Kisa et al. (2015) adapted the MTF in the form of a similar framework for science education, the task based framework in science (TBFS). This strand of research makes a strong case for the importance of tasks, based on four arguments: tasks provide (*i*) key "channels of influence" through which educational development and reform can be implemented in the classroom; (*ii*) a focus on learners' thinking; (*iii*) a fruitful interface of teachers' and students' work in the classroom and of (*iv*) research and practice (Tekkumru-Kisa et al., 2020).

We concur with Silver & Herbst (2007) regarding the "centrality" of tasks for teaching and learning: Tasks are present at the beginning (preliminary diagnosis of e.g., misconceptions), middle (learning process) and end (attainment test) of an educational sequence, and they are central educational building blocks of teaching structure and quality.

In view of this "centrality", the main purpose of the present work is to present and discuss meta analytical and a few other quantitative results concerning various features of tasks and of their use in (STEM) education. Classroom practice will play a large role in the discussion. In some cases, where we are aware of useful additional information in that respect, special text boxes (Research into Practice, RiPe) are included.

Note that this synthesis' interest lies in general results that are potentially relevant across disciplines. Of course, tasks also have important aspects specific to biology, chemistry and so on, which are not included here. In particular, aspects of interest to engineering and technology such as design, fabrication, programming and the like are not considered; however, as mathematics and physics play a large role in engineering and technology education, findings about tasks in these fields should indeed be of interest there. It is with these restrictions in mind that we write here about tasks in the STEM disciplines. To limit the scope and length of the contribution, we have also restricted our focus to the learning functions of tasks, while diagnostic and evaluative functions will only be treated implicitly. Finally, this focus implies that affective aspects of tasks will not be dealt with in their own right (which in fact would deserve an independent article), but some general background will be provided in section 1.4 of this introduction.

## 1.2 Terminology: Tasks, Questions, Exercises, Problems, Activities – A Conceptual Framework

### 1.2.1 Tasks

As stated by Doyle (1988) in a foundational work on tasks in mathematics education, the term has a "multiplicity of meanings". Indeed, with the notion of educational tasks, we are faced with a term that is ambiguous along two axes:

(i) by *level*: There is a need for a generic term for tasks, exercises, problems and so on, but the generic term chosen is often used with another meaning by the same author.
(ii) by *origin*: The meanings of generic and specific terms vary frequently by author, translation, and research tradition.

The purpose of the present contribution is not to debate definitions. We take a pragmatic position making clear the level and origin of the term in question.

For the generic term, we follow the definitions found in the MTF and TBFS, where a task is defined as "a segment of a classroom activity devoted to the development and assessment of a disciplinary idea and/or a practice" (Tekkumru-Kisa et al., 2020). Tasks as activities "incorporate the products that students are expected to produce, the processes that they are expected to use to generate those products, and the resources available to them while they are generating the products" (Tekkumru-Kisa et al., 2015).

For specific terms (questions, exercises, problems and so on) we keep the original definitions from the respective authors, without attempting to construct a unifying terminology (Müller et al., 2013 for a discussion of some aspects).

### 1.2.2 Task Culture as an Educational Term

"Culture" is the "set of shared attitudes, values, goals, conventions and practices that characterizes an institution or organization", or that are associated with a particular field or activity (Merriam-Webster, n.d). In that sense, "task culture" is a set of shared attitudes, values, goals, conventions and practices for how to formulate, use and solve tasks in an educational setting (Andersson Bakken et al., 2020). Similar ideas underpin "task systems", or "structured task systems" (Steinbring, 1991; Csapó, 2005) or the task based framework in science (TBFS; Tekkumru-Kisa et al., 2015), see section 1.1.

Given tasks' importance and broad conceptual horizon in science and mathematics education, it is to be expected that there should be multiple points of contact with discipline based education in other areas (Thonhauser, 2008). Indeed, these can be very fertile. In foreign and second language education especially, there is a large body of research and development around tasks, and "task based/supported learning" (Ellis, 2003; Adams, 2009); see Thonhauser (2010) for a conceptual and terminological analysis, and Jacquin & Müller (2013) for a comparison between science and foreign language education.

Common to all these terms and conceptualisations is the idea of a systematic or structured way of using tasks all along a learning path, and of their central role on this path. This is well encapsulated by the view (and terminology) of tasks

as the basis of “opportunity structures for learning” according to Neubrand et al. (2013) (Fig. 1). “Culture”, from its latin origin (care, tillage), additionally has the educationally positive connotation of “making grow” or “raising”.
The term “task culture” was coined in the German speaking science and mathematics education community in the context of the reaction to the TIMMS results (BLK, 1997; Prenzel & Duit, 2000; Prenzel, 2008; Konsortium HarmoS Naturwissenschaften+, 2010) and has a slight flavour of an academic Germanism (like the word “didactic”). However, “task culture” is part of the “subject culture” (Andersson Bakken et al., 2020), and as the latter, in turn, is established in the English language educational literature (Siskin, 1991), the use of “task culture”, with its positive connotation, as an adequate English term seems justified.

## 1.3 Methodology

In the present contribution, effect sizes will be reported as Cohen $d = (M_T\ M_C)/SD$, where $M_T$ and $M_C$ are the means (of some variable of interest) for the treatment and control group, respectively, and $SD$ is the either the pooled standard deviation or that of the control group (Cohen, 1988). In simple terms, $d$ thus measures the impact of an intervention in units of standard deviations of the sample under consideration. An illustration regarding questions as a specific form of tasks (cf. sect. 2.2.1) is given in Fig. 2. A corrected version of $d$ for small samples is $g = (1\text{-}3(4(n\text{-}2)\text{-}1)^{-1})d$, where $n$ is the total sample size of control and treatment group taken together (sometimes called “Hedges g”; Fritz et al., 2012); $g$ is always smaller than $d$, and the correction is small for reasonable sample sizes ($\leq$ 4% for $n > 20$).
Conversion between correlations and other reported quantities and Cohen $d$ can be carried out according to the relationships given by Fritz et al. (2012) or Lipsey and Wilson (2003). Usual effect-size levels (as established from comparison of a great many studies in different areas) are small ($0.2 < d < 0.5$), medium ($0.5 \leq d < 0.8$), or large ($0.8 \leq$ d) (Cohen, 1988). Another reference value of $d = 0.4$ is used by Hattie as a “hinge point” between influences of smaller and larger size[2].
Many modifications and refinements of the concept of “effect size” have been developed and are used in the literature, see for example Fritz et al (2012) for an overview.
Consistent with the main purpose of this contribution (1.1), the effect sizes included in the following discussion are mainly based on meta analyses and some large sample studies ($n \approx$ several thousands) selected from (international) peer reviewed research journals; results from individual studies are added as illustrative examples. Note that this work is not a meta analysis or systematic review in itself, but rather a “best evidence review” on a set of important characteristics for the use of tasks in STEM education, with a focus on teaching practice and teacher education (in the sense of e.g., Marzano and colleagues (2001) or of the Swiss project *Lernen sichtbar machen* (Beywl & Zierer, 2018)).

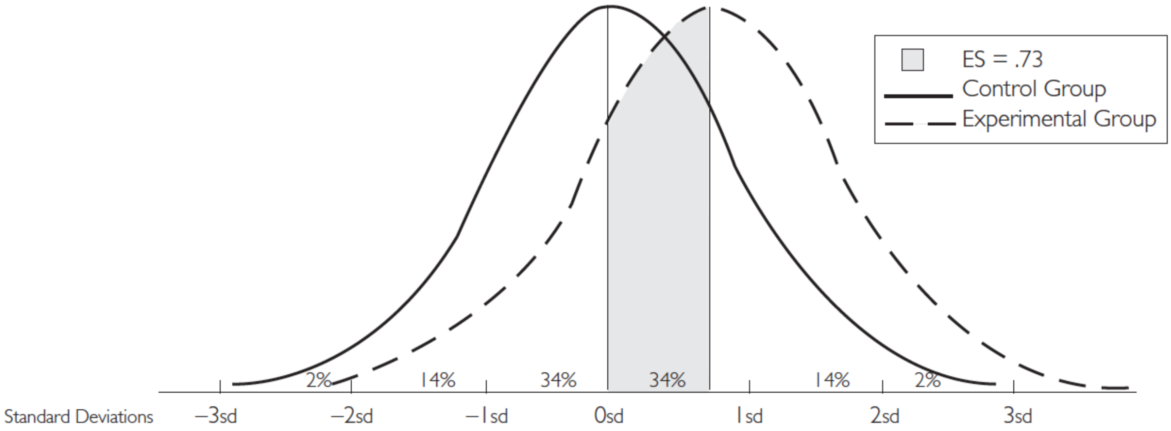

**Fig. 2.** Visualisation of Cohen $d$ as effect size for the example of higher-order questions, cf. sect. 2.2.1 (treatment vs. control group; (Marzano et al, 2001): The treatment group mean is shifted by $d$ standard deviations to the right (here: $d = 0.73$)

---

[2] We agree with Hattie (2009) that such a threshold is an element of discussion to be used with circumspection, not a value to be applied blindly. Lower effect sizes might well be worth considering, depending on available alternatives, effort, and so on, and vice versa for higher effect sizes.

## 1.4 Motivational Issues

In this section, we discuss three important challenges connected to tasks with respect to motivation. This aspect provides important conditions and constraints for their use as learning opportunities and is discussed here as a background for the results on cognitive/learning aspects presented in section 2.

### 1.4.1 Tasks with Low Interest

First, Fig. 3 shows how several classical forms of tasks in physics education are not well received by pupils, in particular tasks where calculation is required. On the other hand, experimental activities are met with a high degree of interest (high/very interest: ≥ 50% of learners, Hoffmann et al., 1998; Häussler et al., 1998). For a teacher, it is important to be aware of these findings.

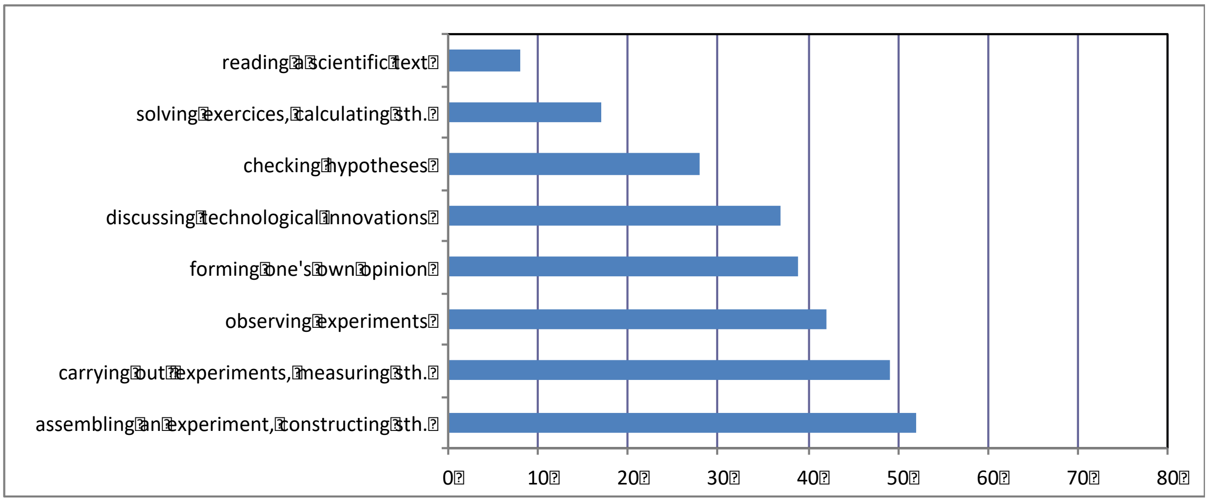

**Fig. 3.** Physics activities with high/very high interest (in %); data from the large sample IPN interest study (N ≈ 8000; ages 11–16) (Hoffmann et al., 1998, Häussler et al., 1998)

### 1.4.2 Learners' Perceptions of Tasks and Task Features

Second, difficulties may arise if learners' perceptions of tasks and their content are not adequately taken into account. Consider the decline of interest and related attitudes towards science during adolescence, which is widely documented by research (Krapp & Prenzel, 2011; Osborne et al., 2003) and which teachers know very well from facing it in their daily teaching.

One widely discussed approach to counter this is context based science education (CBSE), that is to say "using concepts and process skills in real life contexts that are relevant to students from diverse backgrounds" (Glynn & Koballa, 2005, p. 75). However, look at Fig. 4: it displays a striking "scissor" shape, with a strong discrepancy between the perceived relevance of physics, which increases over the years, and interest, which shows a marked decrease. A closely related result concerns PISA tasks, whose philosophy to ensure "relevance to students' interests and lives" is essential to PISA's understanding of scientific literacy (OECD, 2007, p. 27).

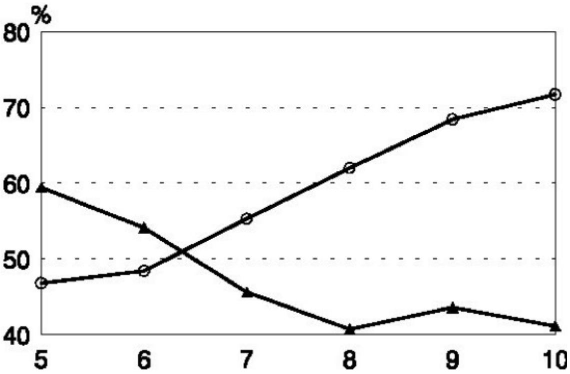

**Fig. 4.** Development of interest (▲) for and of perceived relevance (○) of physics across adolescence (Muckenfuß, 1996) (x-axis: grade; y-axis: percentage of maximum possible (POMP, Cohen et al., 1999) for the scale of interest).

Yet analysis shows (Fig. 5) that for the PISA items available, learners' interest and perceived relatedness to real life is at best medium, contrary to PISA's basic assumption. Moreover, there is a considerable gap between teachers' and students' perceptions (relatedness to real life: Cohen's $d = 1.34$; interest: $d = 1.60$). When asked for their assumptions about pupils' perceptions, teachers adjust their own perceptions to lower values, but still show considerable overestimation of leaners' own interest and perceived relatedness to real life (perceived relatedness to real life: $d = 0.66$, $p < 0.05$; interest: $d = 0.85$ (Weiss & Müller, 2015)).

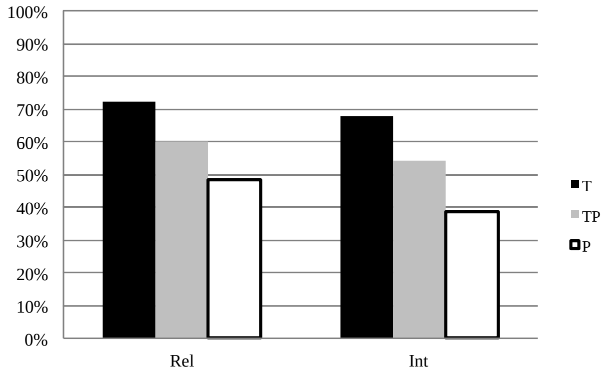

**Fig. 5.** Teachers' perceptions (T), teachers' assumptions about pupils' perceptions (TP) and pupils' perceptions (P) for the published PISA items (*x*-axis: Rel = perceived relatedness to real life, Int = interest; *y*-axis: percentage of maximum possible). (Weiss & Müller, 2015)

Both results (Fig. 4, Fig. 5) show that merely accounting for relevant contexts for science education in the sense of "making it relevant" (Nentwig & Waddington, 2005) is not sufficient to maintain (nor generate) adolescent learners' interest. Furthermore, there might be strong and unsuspected discrepancies (*i*) between the development of perceived relevance and that of interest, and (*ii*) between teachers' and students' perception of relevance and interest for a given task. Taken together, these results show the importance of an evidence-based development, selection, and use of tasks in science education also from a motivational perspective.

### 1.4.3 Self-beliefs and Fear of Failure

> *[A] sense of confidence is a most powerful precursor and outcome of schooling. It is particularly powerful in the face of adversity – when things do not go right, or when errors are made. Having high levels of confidence – "can do", "want to do"– can assist in getting through many roadblocks.* Hattie (2009, p. 47)

> *Research suggests that ways to reduce students' fear of failure [...]*
> *can be important ingredients for future success.* Martin (2012b, p. 2341)

The third issue to be discussed here is the potential for negative interactions between TBFS and self-beliefs. Self-beliefs as an umbrella term are defined "in terms of how much students believe in their own ability to handle tasks effectively and overcome difficulties (self-efficacy) and students' beliefs in their own academic abilities (self-concept)" (OECD, 2007). One the one hand, self-beliefs are strong predictors of interest, academic effort, and persistence in the face of difficulty in a given field (Möller et al., 2009, Trautwein et al., 2009). Moreover, they are known to be essential for appropriate task choice (Möller et al., 2009) and choices of an academic track or professional career in STEM fields (Gehrig et al., 2010) in particular for female students (Seymour, 1995; Murphy & Whitelegg, 2006; Stout et al., 2011). For instance, Güdel et al. (2019) have found that for 13–14-year-old youths, the influence of self-efficacy is more than three times stronger (in terms of standardized regression weights) than the influence of interest; for girls, it is even more than four times stronger. Finally, regarding learning, there is strong meta analytical evidence of sizeable positive influence of self-beliefs: Hattie (2009) reports an average effect size $d = 0.43$, with individual meta-analyses up to $d = 0.76$. Richardson et al (2012), in a more recent meta analysis on psychological correlates of academic performance of undergraduate students, find for academic self-efficacy (related to studying in general) $d = 0.65$, for performance self-efficacy (more specifically related to study outcomes) $d = 1.1$, and for intrinsic motivation $d = 0.35$; note that the ratio of the latter effect size to that of performance self-efficacy is greater than 3, comparable to the result of Güdel et al. (2019) – self-efficacy is a much stronger predictor of learning than is motivation. Finally, Aker & Ellis (2019), in another recent meta-analysis on "middle school students' science engagement" (as a combination of cognitive, affective and behavioural aspects) report for "feeling of competence" as predictor an effect size of $g = .56$.

On the other hand, if positive self-beliefs support learning, negative subject- and task-related emotions such as fear of failure (or "academic anxiety", Martin, 2012a) impede it. If – for whatever reasons, personal, situational, or social – a person's fear of failure is stronger than their motivation to achieve success, the result of the conflict is to avoid the task at hand (Zeidner, 1998). Henry et al. (2019) emphasise that fear of failure "has a strong influence on how students might approach an academic challenge" (i.e., a task or test, cf. sect. 2.2.2 for homework). Moreover, Martin (2012a) describes a sort of downward spiral or vicious circle: If, for fear of failure, a learner avoids activities and tasks in a given area or subject, "he moves down a cascade, with ever declining success and self-beliefs, and increasing fear of failure". These strong mechanisms, known to many of us from professional (or personal!) experience, are echoed in meta-analytical results. Hattie (2009) reports an effect size of reducing anxiety[3] on learning of $d = 0.4$. Within STEM disciplines, this value is even larger ($d = 0.56$ in mathematics). Note also that the considerable positive effects of supportive self-beliefs on learning discussed above ($d = 0.4$–$1.1$) become negative effects when supportive self-beliefs are absent.

Summing up, one has every reason to ensure that tasks, especially when taking the central role in the classroom as advocated by TBFS, provide an encouraging experience for learners, support the development of their self beliefs, and do not create fear of failure.

## 1.5 Evidence based (Science) Teaching and Learning

*Fundamentally, the most powerful way of thinking about a teacher's role is for teachers to see themselves as evaluators of their effects on students.* Hattie (2012, p. 14)

*[R]eform in science education should be founded on "scientific teaching", in which teaching is approached with the same rigor as science at its best.* Handelsman et al. (2004, p. 521)

The preceding examples – task types with low and high interest, learners' perceptions of tasks, and the interaction of tasks with affective variables – show how important it is to have a reliable empirical basis for the development and design of task based learning approaches, indeed for any educational decision.

Evidence based practice is the approach whereby decisions are based on the best available evidence, in the sense of the best possible – in particular, systematic! – use of existing knowledge and research. The earliest example in this respect was evidence-based medicine (EBM), that is "[t]he conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients", as one of its pioneers put it (Sackett et al., 1996). The common ground with evidence-based education is that a precious value and objective – good health, or good education – must be realised with limited means. This has led to a strong current of research and practical implementation in evidence based (science) education (EB(S)E) in the last two decades (Davies, 1999, *Educational Researcher*, 2008; Finnigan & Daly, 2014). The work of Hattie (2009; 2017) based on more than 800 meta-analyses (comprising more than 80,000,000 individuals) is considered ground breaking in this area. Other well-known sources are the *Best Evidence Encyclopedia* website from Johns Hopkins University (http://bestevidence.org), and in German speaking countries the websites *Lernen sichtbar machen* (www.lernensichtbarmachen.ch) and *Clearing House Unterricht* (www.clearing-house.edu.tum.de).

As a word of caution, it must be mentioned that there is a thorough debate about EBE, and it is certainly not to be claimed as a guarantee unto itself of a solution to all problems in science education and science teacher education, as evidenced by the following statement by Millar et al. (2006): "We need to work towards a situation in which research evidence is routinely an explicit input to teachers' decision making, but where it is also accepted that this must be weighed judiciously alongside other kinds of knowledge to reach a decision that can be rationally defended." It is to such a well-balanced point of view that we subscribe upon review of the evidence on tasks in science education. Effect sizes are an *element* of educational reflection, *not* a replacement.

# 2 Task based Learning: Meta-Analytical Results

*No one would think of getting to the Moon or of wiping out a disease without research. Likewise, one cannot expect reform efforts in education to have significant effects without research-based knowledge to guide them.* Shavelson & Towne (2002, p. 1)

## 2.1 Between Weak and Strong Guidance

### 2.1.1 Discovery-, Inquiry-, and Problem-based Learning

*Non vitae sed scholae discimus.* Seneca (ca. 62–64 BCE)

Discovery, inquiry, and problem-based learning (D/I/PBL) are three learning approaches particularly strongly advocated in science learning. They share certain similarities. A common point of departure is the criticism already made

---

[3] Note that Hattie (2009) distinguishes test anxiety and subject anxiety (e.g., towards mathematics), to which one can add "performance anxiety" (Vanden Bos, 2015), but these facets are not distinguished in the meta-analytical results.

by Seneca almost 2,000 years ago (see quote above[4]): "We do not learn for life, but for school" – in an artificial, uninteresting, and eventually useless way. To counter this, DBL, IBL and PBL share some core features, namely, to engage in learners in authentic, complex, and open tasks, most often in a collaborative setting, and with minimal or no guidance.

**Tab. 1.** Definitions of discovery-, inquiry-, and problem-based learning.

| Form of task | Definition |
|---|---|
| Discovery ( based) learning | "[D]iscovery learning occurs whenever the learner is not provided with the target information or conceptual understanding and must find it independently and with only the provided materials. [...] Common to all of the literature [...] is that the target information must be discovered by the learner within the confines of the task and its material." (Alfieri et al., 2011, p. 2) |
| Inquiry-based learning | "Focus on having students learn disciplinary knowledge, reasoning, and epistemic practices as they engage in collaborative investigations.<br>Inquiry is organized around the questions that scientists might ask or disciplinary problems that require scientific inquiry to resolve." (Eberbach & Hmelo-Silver, 2014, p. 514) |
| Problem-based learning | "Authentic, real-world situations or issues likely to be addressed in the work place." (Pepper, 2014, p. 795) |

The exact definitions differ for the three approaches (and in many cases also from author to author for the same approach), with each focussing variably on findings/results (DBL), on the process(es) of scientific inquiry (IBL), and on using real life problems as starting points (PBL) (Tab. 1). With these similarities and differences in mind, there is strong empirical support for a common, decisive condition of success for these approaches.

IBL, although strongly favoured on the political level (Rocard et al., 2007), is supported by little empirical evidence in terms of its overall effectiveness ($d = 0.31$, Hattie, 2009). Upon closer inspection, two statements can be made about IBL: First, its effects are greater on science process (understanding, skills) than they are on science content. Indeed, effect size ratios for science process/content such as 0.52/0.16 (Bredderman, 1983) or 0.40/0.26 (Shymansky et al., 1990) can be found. Second, a meta-analysis on IBL in science by Furtak et al. (2012) showed teacher guidance to be a decisive factor: Effect sizes are more than twice as large with guidance than without ($d = 0.65$ and $d = 0.25$, respectively). Similarly, DBL has been studied in comparison to conventional teaching. unguided DBL showed an effect size of $d = -0.38$, while teacher assisted conventional teaching was at $d = 0.30$ (Alfieri et al., 2011). Thus, *unguided* discovery- or inquiry based learning, as DBL and IBL are often understood, are not effective approaches for science learning. Finally, a (non-quantitative) review on a related approach, project based learning, emphasised how the benefit to learners can be enhanced through various forms of scaffolding (Thomas, 2000). An important issue for the group of learning approaches in this section is the limitation of memory resources available for the simultaneous learning of new content and the required self-guidance of the learning process, or conversely, the cognitive (over-)load created by this simultaneous requirement. The importance of the limited memory and cognitive load factors will also be discussed in sections 2.1.2 and 2.4.1.3 (Mayer, 2004; see also the debate in *Educational Psychologist* (2006, 2007)).

### 2.1.2 Worked Examples[5]

"Worked examples" (WEs) are a type of learning task that make use of model solutions, that is to say "step-by-step demonstration[s] of how to perform a task or how to solve a problem" (Clark, Nguyen & Sweller, 2006, p. 190) to facilitate complex learning while avoiding cognitive overload. They have been extensively studied and appear to be a promising teaching approach (Atkinson et al, 2004; Gauthier & Jobin, 2009 for the French *exemples ciblés*). They are usually composed of a problem statement and a detailed set of instructions for solving it. Taken together, it is expected that the problem and its solution should act as an expert model for the learner to study and re enact to solve other, similar problems (Atkinson et al, 2004).

According to McLaren et al. (2008), WEs support the development of students' problem solving skills in a high level way. By allowing learners to concentrate on understanding the solution method one step at a time, rather than having to derive it entirely themselves, WEs work by lowering the cognitive load imposed by a problem (Van Gog et al., 2004). Gauthier and Jobin (2009) list four further benefits of WEs: understanding the fundamental principle of solution methods; making sense of processes; identifying similarities and differences across examples; and anticipating the next solution step to then compare one's prediction to the WE.

By enabling learners to take ownership of the how (strategies) and why (principles) of skilled approaches, which are often implicit at the expert level (Renkl et al., 2002, Van Gog et al., 2004), WEs constitute a form of active learning,

---

[4] Note that the original quote is the one given here, meant as a *criticism* of the unsatisfactory educational practice of Roman school at that time; the converse statement "*Non vitae, sed.*

[5] Partially based on Loretan, Müller, & Weiss (2018a, b)

supporting the autonomy of learners. Several studies have established the effectiveness of WEs, notably with respect to reducing cognitive load (Sweller, 2006) with a considerable effect size ranging from $d = 0.62$ at the end of secondary to $d = 0.73$ at the start of tertiary education (Crissman, 2006). Meta analyses find $d = 0.57$ for learning in general, and $d = 0.7$ in physics (Crissman, 2006; Hattie, 2009). Loretan et al. present an example of the use of worked examples in the format of a research-based report of practice in this thematic issue of *PriSE*.

## 2.2 In the Classromm and at Home

### 2.2.1 Questions, (Self-)Questioning

Questions play a crucial role in science learning. The reason for this is twofold: they are "key to active and meaningful learning", with many educational functions (Tab. 2), and they are "the cornerstone of scientific enquiry" at the same time (Chin, 2004). In the following, we consider three types of questions and forms of questioning according to the specific place in which they arise and their source within the learning process: classroom questions (or questioning), by the teacher; self questioning, by the learner; and adjunct questions, from texts or other learning media.

**Tab. 2.** Educational purposes and reasons for using classroom tasks

| **Reasons to ask classroom questions (Marzano et al., 2001; Brualdi, 1998)** |
|---|
| They help teachers "to keep students actively involved in lessons". |
| They help teachers "to pace their lessons and moderate student behavior". |
| They help teachers "to evaluate student learning". |
| They help teachers to "revise their lessons as necessary". |
| Students have the opportunity "to express their ideas and thoughts". |
| Students have the opportunity to hear different ideas, thoughts and explanations by their peers. |
| **Further purposes of classroom questions** |
| They can help to activate prior knowledge (Marzano et al., 2001). |
| They can help to diagnose misconceptions (Hodgson & Pyle, 2010). |

#### 2.2.1.1 Classroom Questions (or Questioning)

The frequency of teacher questions is around 1/min (elementary school: 45–150 questions/hr, (Marzano et al., 2001); English: 42/hr, (DESI, 2006); mathematics: 70/hr, (Heinze & Erhard, 2006); chemistry: 40/hr, (Nehring et al., 2017)). These high frequencies are mirrored by the many educational functions of classroom questions put forward in the literature (Tab. 2).

The overall effect size of questioning is $d = 0.46$, with a large range (0.26–0.82) found in the different meta analyses considered (Hattie, 2009). A strong moderator is the cognitive level of questions, with "lower order" questions merely asking for recall or recognition of factual knowledge, while "higher order" questions require learners to combine and transform elements of knowledge, yielding a considerably higher effect size of $d = 0.73$ (Marzano et al., 2001; cf. Fig. 2). A meta analysis based on more recent data and specifically for science education reported an effect size of $d = 0.74$ for questioning strategies (Schroeder et al., 2007), that is to say, strategies including more higher order questions, increased wait time (see below), or other measures.

Another strong moderator is wait time. Different authors have used a variety of relatively detailed conceptualisations of wait times (Rowe, 1974; Tobin, 1987; Ingram & Elliott, 2016). We will focus on the most frequently investigated version: the time allowed between a teacher question and a student answer (Rowe, 1974). Research spanning over 40 years has established a timing pattern that is stable and rather consistent across disciplines: wait time between question and answer is consistently < 3 s, and often < 1 s (science: Rowe, 1969, Rowe, 1974; Tobin, 1987, Black & Harrison, 2001; math: Heinze & Erhard, 2006; other subject areas: Tobin, 1987).

However, one meta-analysis (Wise & Okey, 1983) reveals a large overall effect of increased wait time (3–10 s) of $d = 0.9$ on general educational (mostly cognitive) outcomes; differentiating this result into low- and high level assessments, general achievement, and problem solving on the one hand, and critical thinking, logical thinking, creativity, and affective measures on the other hand, the effect sizes were 0.53 and 1.27, respectively. Beyond these measures of effectiveness, several benefits for classroom processes were also found (Tab. 3), with a shift from quantity to quality for teacher questions, and increased quantity and quality of student answers, and student participation.

**Tab. 3.** Consequences of longer wait time for classroom processes. References: a) Heinze & Erhard, 2006; b) Sadker, 2002; c) Caillé, 1995.

| **Consequences** | **References** |
|---|---|
| Number of teacher questions decreases (this is almost trivial). | a) |
| Cognitive level of teacher questions increases; fewer merely factual questions. | a) |
| Increased student participation (more utterances). | a) b) |
| In particular, increased participation by slower-reacting and more timid students. | c) |
| More on-task student talk. | b) |
| Student responses are longer, more correct, more complete, and at a high cognitive level. | a) b) c) |

Regarding the latter finding, Sadker (2002, p. 312) put it nicely, stating that "longer wait-time can attract low participating learners into class interactions. Students with limited English proficiency, minority students, lower-achieving students, and females are typically among those who benefit from a longer wait-time." Another noteworthy finding is the increase in the proportion of "higher order" questions (see RiPe Box.1 for examples), an interesting positive interaction with the moderator discussed above.

| **Research into Practice (RiPe) Box.1: Questioning**. Some examples of "higher-order questions" (Fonseca & Chi, 2017) and in particular for science (Barton & Jordan, 2001). |
|---|
| **Cognitively activating ("higher-order") questions** |
| "What is the main idea of…?" |
| "How does… relate to…?" |
| "What conclusions can I draw about…?" |
| "Can that be correct?" |
| "Let me have another look at that…" |
| **Examples for science** |
| "How does the author describe physical properties of matter?" |
| "Summarise the given explanation of photosynthesis." |
| "What generalisation can I make about mixtures?". |
| "What points did I learn about using electricity safely?" |
| "Restate what the author meant by renewable resource." |

### 2.2.1.2 Self-questioning

If teacher questioning has positive effects on learning, this can be also expected for self questioning, as this constitutes an even higher degree of active learning.

**Tab. 4**. Meta analytical effect sizes (Cohen *d*) for self-questioning (Hattie, 2009)

| **Self-questioning…** | **Effect size** |
|---|---|
| Overall | 0.64 |
| Pre-lesson | 0.94 |
| Post-lesson | 0.86 |
| During lesson. | 0.54 |
| With teacher modelling | 0.69 |
| Without teacher modelling | 0.47 |

Indeed, the overall effect size is $d = 0.64$ (somewhat higher than for teacher questioning; Hattie, 2009). A look at Tab. 4 reveals two further noteworthy findings: First, the effects are considerably higher for pre- and post lesson than for in lesson self questioning. Reasons for this might be limited time, increased cognitive load, or a combination of both. This provides a strong argument for self questioning as homework (sect. 2.2.2) or in inverted classroom settings (sect. 2.2.3). Second, the effects are considerably greater with teacher modelling than without. This is consistent with the necessity of guidance stated in section 2.1.

#### 2.2.1.3 Adjunct Questtions

*What are the effect sizes of higher- vs. lower-order classroom questions?*

Adjunct questions are questions asked before, during, or after a written text (or other kind of document) to be studied by a learner (Crooks, 1988). By way of example, this and the next paragraphs are formatted in a way to include a lower- and a higher order adjunct question, as you might them find in a textbook. The overall effect size by adjunct questions has been found to be $d = 0.4$ (Hattie, 2012).

*How would you compare the effects of higher- vs. lower order adjunct questions with those of classroom questions?*

Looking more closely, there are also meta-analytical comparisons of such higher- versus lower order (factual) adjunct questions, with an effect size of $d = 1$ and 0.3 for repeated and related higher order test questions, respectively (Hamaker, 1986). Note that the effect sizes on repeated and related factual test questions are negative ($d = 0.55$ and 0.4, respectively, Hamaker, 1986). As in the case of wait time, such an educational measure requiring relatively little effort can have sizable positive effects.

### 2.2.2 Homework

*I hate homework. Why can't we just learn at school and be done with it?* Student quote from Marzano et al. (2001, p. 60)

*Nous dénonçons depuis longtemps la persistance des devoirs à la maison, dont personne n'a jamais prouvé l'utilité.*
Sauvons l'école (2012)

Homework is as widespread as it is strongly debated by learners, parents (see quotes above), teachers, and researchers. First, the overall effects are not impressive ($d = 0.29$, Hattie, 2009) but a much sharper image is obtained when accounting for feedback as a moderator. The presence or absence of feedback leads to a considerable contrast between strong and weak effects ($d = 0.83$ vs. $d = 0.28$, Marzano et al., 2000). Moreover, there is also a strong dependence on age (elementary pupils: $d = 0.15$, high school students: $d = 0.64$; Hattie, 2009), probably due to the larger autonomy in learning of older pupils. An important difficulty, especially for lower achieving pupils, is frustration and fear of failure (sect. 1.4.3). Hattie (2009, p. 235) says "For too many students, homework reinforces that they cannot learn by themselves, and that they cannot do the schoolwork. For these students, homework can undermine motivation, internalize incorrect routines and strategies, and reinforce less effective study habits, especially for elementary students".

The meta-analytical results are based on hundreds of studies, involving over 100,000 students, and they reveal that the opinion whereby "*personne n'[en] a jamais prouvé l'utilité*" ("no-one has ever proven the usefulness [of homework]") is simply not true. Rather, the usefulness of homework depends crucially on feedback, and on the age of pupils.

This example serves to illustrate two general statements: (*i*) Evidence-based education and meta-analyses are only as valid as the range of important cofactors taken into account is broad; simplistic recipes are not a desirable goal. (*ii*) The cofactor in question here, feedback, is consistent with the outcomes regarding the high effectiveness of feedback in general (sect. 2.5).

| **RiPe Box 2. Homework** – some guidelines for practice (Cooper, 1989 unless specified otherwise) |
|---|
| Avoid/reduce frustration and fear of failure (Hattie, 2009) |
| Effects are greater for high than lower ability pupils. |
| Effects are greater for older rather than younger pupils. |
| Monitoring/Feedback is even more important for lower ability pupils. |
| Spend more time/effort/attention on monitoring lower ability pupils. |
| "Short is beautiful": the younger the pupils, the shorter the homework. |
| Homework time: "10 min per grade" rule of thumb (Marzano et al., 2001) = 10 min * grade level (1st grade: 10 min; 6th grade: 1hr; 9th grade: 1.5hr; 12th grade: 2hr) |

### 2.2.3 Active Learning, Interactive, and Inverted Classroom Approaches

We now turn our attention to a development from the last few decades that combines several elements of task culture as discussed above. "Active learning" or "interactive engagement" approaches (see RiPe Box below), as are they are known (Meltzer & Thornton, 2012), have increasingly been the subject of research, notably in the context of university teaching (Deslauriers et al., 2011; Freeman et al., 2014). Such approaches introduce several features to enhance students' active, intellectual participation (Deslauriers et al., 2011): (*i*) in class use of clicker questions with discussion between students; (*ii*) targeted in class instructor feedback (based on clicker question data); (*iii*) (in many cases) pre class reading assignments and quizzes (i.e., components of an inverted classroom, cf. below). Thus, such active learning combines classroom questioning, homework, and feedback – all aspects of tasks (sect. 2.2.1.1, 2.2.2, 2.5, respectively).

A large scale meta-analysis (Freeman et al., 2014) spanning multiple disciplines and universities looked at 225 studies involving approximately 46,000 and 29,000 students to assess their learning and measure their failure rate, respectively. Its outcomes strongly support the effectiveness of active learning approaches. This is illustrated in Figure 6, which compares active learning courses to traditional ones. The increase in learning found here had an overall effect size of 0.47 for examination scores and concept inventories (among other measures).

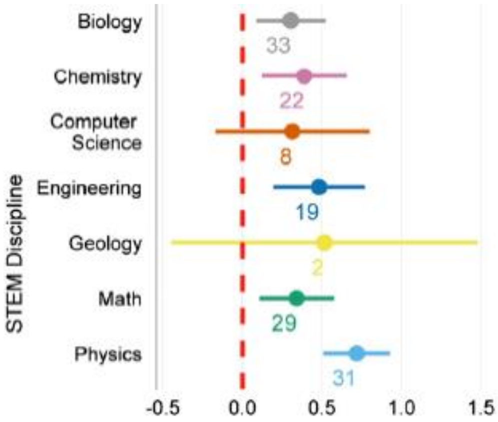

**Fig. 6.** Comparison of active learning courses vs. traditional lectures: Meta-analytical effect sizes (in SDs) for learning outcomes (course exams, concept tests, other assessments) for several STEM disciplines (Freeman et al., 2014); error bars are 95% confidence intervals, numbers below each data point are the number of studies included.

For concept tests more specifically, the effect size was 0.88, while the failure rate fell from 34% to 22%. In physics, the overall impact on learning is even greater (effect size 0.72). This latter finding is consistent with another large study on 600 classes with 25,000 students (Von Korff et al., 2016). A similar example at the high school level is the predict observe explain sequence, which has also been thoroughly examined (Yin, 2012; Yin et al., 2014; Liew & Treagust, 1995). This same sequence also works for demonstration type "experiments" in class, which are still in widespread use due to financial limitations and safety concerns. The central take away message is consistent: "Active learning increases student performance in science, engineering, and mathematics" (Freeman et al., 2014). Its effect sizes as it does so are considerable. Furthermore, it can be implemented in a variety of teaching conditions, including "frontal" ones (Meltzer & Thornton, 2012). RiPe box 3 highlights a set of practical guidelines for classroom application of active learning.

An educational format closely linked to active learning approaches is the "inverted" or "flipped" classroom, whereby traditional in class and out of-class activities are reversed[6]. Most often, it assigns what usually is in-class lecture content to out-of-class homework, providing "opportunities for students to interact and engage in instructor-monitored learning activities during face-to-face in-class" (Shi et al., 2020, p. 80). For obvious reasons, the format has gained in importance during the school closures imposed in many countries because of the Covid 19 pandemic.

Two relatively recent meta-analyses have investigated inverted classrooms. Shi et al. (2020) looked into their use in teaching at a university level in the fields of medical and social sciences and STEM. The effect size of the benefit of inverted as opposed to traditional classrooms was 0.53. In STEM specifically, this value was at 0.49. Where active learning approaches (see above) were present, the effect size doubled to 1.05. However, the studies in this meta-analysis were considerably heterogeneous. Lo et al. (2017), meanwhile, examined inverted classrooms at high school and university level in mathematics teaching. They found an overall effect size of 0.30. When formative assessment was included at the start of face to face lessons, this rose to 0.57. When it was absent, the value fell to 0.20.

Thus, inverted classrooms have been found to be somewhat superior to traditional teaching, with small to medium effect sizes across disciplines. The features most essential to its success are cognitive activation and formative assessment, which are associated with medium to large effect sizes.

| **RiPe Box 3. Guidelines for active learning approaches.** In their review on active learning approaches, Meltzer & Thornton (2012) provide a set of guidelines for how active learning can be implemented in classroom practice. While existing research is mostly about the university level, it appears clearly that these guidelines are all also applicable to science education at school: |
|---|
| Guiding instruction according to leaners' pre-instruction knowledge state, bases on existing research. |
| Eliciting and addressing students' ideas. |
| Encouraging students to figure things out for themselves. |
| Engaging in diverse problem-solving activities during class time. |

---

[6] Note, however, that they are not the same: one can have an inverted classroom without "in-class" activation of learners by appropriate tasks, and one can have the latter without the former.

| |
|---|
| Emphasising qualitative and conceptual reasoning. |
| Posing problems in a wide variety of contexts and representations. |
| Incorporating the use of actual, physical systems in problem solving. |
| Requiring students to express their reasoning explicitly. |
| Providing rapid feedback to students. |
| Incorporating student reflection on their problem-solving practice. |
| Emphasising the linking of concepts into well-organised hierarchical structures. |
| Having students work together in small groups. |
| Integrating both appropriate content and appropriate activities. |

## 2.3 Between Hands-on Learning and Cognitive Structures

### 2.3.1 Hands-on and Experimental Activities[7]

Hands on activities and manipulatives are considered by many practitioners to be strongly motivating for learners. Brunsell and Fleming (2014) describe them as one way to provide joy in a learning context. For science learning, a large-sample study (N ≈ 8000, ages 11–16) has indeed shown that they are the type of learning activities eliciting the highest degree of interest (high/very interest: ≥ 50%) in learners (Häussler et al, 1998). A further study (Swarat et al., 2012) confirms hands-on activities to be strong predictors of science interest at school.
As for cognitive effects, the RAND cooperation conducted another thorough study on the impact of hands-on learning on science achievement in a large sample of 8th graders (N ≈ 1400, ages 13–14), which yielded very strong effects on science achievement ($d = 0.91$); moreover, no strong influences of learner ability were found (Ruby, 2001). This finding is confirmed by meta-analytical results on the effects of manipulatives on science learning: $d = 0.57$ (Schroeder et al., 2007), $d = 0.56$ (Wise & Okey, 1983); note the consistency of results over more than two decades. Comparable effects have been found for several outcomes in mathematics education (effect sizes for simple application, problem solving, justification/explanation: $d = 0.59$, 0.46, 0.38 respectively; Carbonneau et al., 2013). As one can see, these effects are partially lower than those for science learning (and less consistent, Calvov & Gomila, 2008, chap. 18), a fact discussed in terms of the difficulty learners experience in linking or transferring a concrete manipulative to an abstract concept and its symbolic representation (Dörfler, 2000; Carbonneau et al., 2013).

### 2.3.2 Concept Maps

A well known task format for diagnosis and learning activities related to linking knowledge is the "concept map". Table 5 gives an overview from the meta-analysis by Nesbit and Adesope (2006).
As one can see, there are several results of interest here. The effect size covers a broad range ($d = 0.2$–$0.8$). This points to the fact that the effect of concept maps is moderated by other factors. One important such moderator is the difference between constructing as opposed to merely studying concept maps, the effect sizes being 0.52 and 0.32, respectively. This is due to the greater cognitive activation required, very much in line with the activation seen in relation to, for example, wait time (sect. 2.2.1.1) or active learning approaches (sect. 2.2.3). Table 5 discusses several other points with regard to concept maps.

**Tab. 5.** Effect sizes for concept maps (CMs) and dependence on various moderators (Nesbit & Adesope, 2006)

| Concept maps | | Effect size |
|---|---|---|
| Overall | | 0.2-0.8 |
| 1) Study | general science, non physical sciences | 0.34 |
| 2) Construct | general science, non physical sciences | 0.52 |
| 3) Construct | physical sciences | 0.28 |
| 4) Construct | group & individual work (combined) | 0.96 |
| 5) Construct | individual | 0.12 |
| 6) Study | domain ability low / high | 0.37 / 0.03 |
| 7) Study | map animated / static / hyperlinked | 0.74 / 0.40 / 0.02 |
| 1) vs. 2): "construct" more effective than "study". | | |
| 2) vs. 3): for "physical science" less effective than for "general science"; high difficulty of individual concepts as an obstacle to integration? | | |
| 5) vs. 4): including elements of group work much more effective than individual work; CMs as tool for discussion an exchange of ideas ("socio-cognitive" tool)? | | |
| 6) more effective for low ability than for high ability learners (anti Matthew effect, for once). | | |
| 7) animated CMs (much) more effective than static ones, and (much) more effective than hyperlinked ones; guiding attention (by animation) appears to be effective, while unguided exploration (hyperlinks) not. | | |

---

[7] This section is an updated version of Besse et al. (2019).

## 2.4 Learning and Instructional Characteristics

### 2.4.1 Learner Characteristics

A wide range of studies suggest that learning with tasks is also strongly influenced by a series of learner characteristics. In section 1.4.3, we discussed the influence of interest and self-beliefs as motivational characteristics. Below, we review the role of several personality traits, three cognitive factors, and gender.

#### 2.4.1.1 Conscientiousness and Other Personality Traits

The "Big Five" model of personality traits (McRae & Costa, 1997) has been well established and widely applied. It is also known as the CANOE or OCEAN model, acronyms of the five factors involved: conscientiousness, agreeableness, neuroticism, openness to experience, and extraversion. Conscientious learners seek to do well on tasks, with an obvious effect on the effort they invest. Indeed, of the Big Five, conscientiousness is the strongest predictor of academic achievement (Hattie (2009): $d = 0.44$; even higher in a more recent meta-analysis: $d = 0.52$ (Vedel, 2014)). There is little evidence concerning an influence of the remaining factors among the Big Five, showing considerably smaller effects at best, if any. The only factor among them which is sometimes found to have a strong effect is openness to experience, $d = 0.63$ in the study by Lee & Stankov (2016). They reviewed personal traits and other non-cognitive influences on achievement based on TIMSS and PISA data. The main factors of influence they found were self-beliefs (section 1.4.3) and the two Big Five traits discussed above (conscientiousness and openness to experience).

#### 2.4.1.2 Prior Knowledge

David Ausubel coined the phrase: "The single most important factor influencing learning is what the learner already knows" (1978, p. 85). Indeed, learning in general, and science learning especially, is strongly predicted by prior knowledge ($d = 0.67$ and 0.8, respectively, Hattie, 2009).

Moreover, a considerable interaction with self beliefs is to be expected, in turn having a strong influence on task relevant characteristics, such as academic effort, persistence in the face of difficulty, task choice and more (sect. 1.4.3). Beyond domain specific prior knowledge in science, prior knowledge in mathematics is also expected to have an important role, especially in physics (Schwartz et al., 2005; Karam, 2015; Uhden et al., 2012), but also in biology (Bialek & Botstein, 2004; Gross et al., 2004; Ableitinger, 2011) and chemistry (Leopold & Edgar, 2008; Shallcross & Yates, 2014; Hoban et al., 2013). Work over several decades has in fact established strong achievement correlations between mathematics and physics. This research spans traditional studies (Thorndike, 1946; $r = 0.77$), meta-analysis (Fleming & Malone, 1983, $r = 0.48$), and an analysis of introductory physics courses (Meltzer, 2002, $r = 0.3–0.46$). A meta analytical value for general science is $r = 0.63$. These studies yield medium to large effect sizes ($d = 2.4$; 1.1; 0.6–1.0; 1.6, respectively – cf. sect. 1.3 for the conversion), which have been confirmed more recently by more fine grained studies and been interpreted with increasing detail (Torigoe & Gladding, 2007; 2011; Uhden et al., 2012).

#### 2.4.1.3 Memory-Capacity

Table 6 summarises several studies covering a range of age groups and science disciplines. These show substantial correlations between learners' working memory and their achievement in STEM subjects. The values are all comparable to, and in some cases up to a twice as large as, their counterpart correlations between motivation and achievement[8]. These latter are given considerable importance by, for instance, university science departments' outreach efforts.

Note that these findings concern working (short term) memory, not long term memory. By way of a metaphor, the level of complexity and interrelatedness of information that can be treated at any given moment is limited by the capacity of the CPU (working memory), not the hard drive (long term memory). These factors are determinants of how well a learner can come to know and understand something new, especially in complex subjects such as science. This may seem self evident, but the implications for reducing cognitive load (the demand for working memory) for teaching and learning are considerable (a detailed discussion is beyond the scope of the present contribution; cf. e.g., Sweller et al. (2011) for more).

---

[8] This also holds for the only study regarding physics, yielding a somewhat lower value than most other studies.

**Tab. 6**. Multi study summary of correlations between working memory and STEM achievement across disciplines and ages (sec I: secondary level 1; uni: university), to be compared with the correlation between motivation and science achievement.

| discipline (age/age group) | r (correlation coefficient) | Comments | References |
|---|---|---|---|
| **Working memory - achievement** | | | |
| Chemistry (uni. 1st year 1) | 0.28–0.75 | 9 independent studies | Tsaparlis, 2005 |
| Physics (sec. I) | 0.30 | | Chen & Whitehead, 2009 |
| Biology (sec. I) | 0.62 | | Chu, 2008 |
| Science* (age 14) | 0.50 | | Gathercole et al., 2004 |
| Mathematics (age 14) | 0.54 | | |
| Science* (age 11) | 0.46 | Large sample study, n ≈ 5000 - 10000 depending on test and age group. | Donati et al., 2019 |
| Science* (age 11) | 0.60 | | |
| Mathematics (age 11) | 0.58 | | |
| Mathematics (age 16) | 0.63 | | |
| **Motivation - achievement** | | | |
| Comparison value | 0.3–0.4 | n > 500,000 (!) | Wild et al., 2001;<br>Uguroglu & Walberg, 1979 |

*Note that in many countries, "science" is taught as a single, undifferentiated discipline.

#### 2.4.1.4 Visuo-spatial Ability

Visuo-spatial ability is another important factor for the use of many types of tasks in STEM education, especially for tasks involving multiple representations (Ainsworth, 1999; Mayer & Moreno, 2003; Tsui & Treagust, 2013, ch. 1 & 11; Khine, 2017). In the physical sciences, such influences have been discussed by Meltzer (2005), and Gilbert and Treagust (2009, chap. 7.7, 8, 11).

A review by Wu & Shah (2014) looking into correlational studies provides broad evidence about the influence of visuo-spatial ability (although without reporting the correlation coefficients). A few quantitative studies can be found in chemistry education (Stieff, 2004; Carter et al., 1987) where the effect size of visuo-spatial ability is in the range 0.4–0.6 (for the conversion, cf. sect. 1.3). Physics, with its notable degree of abstractness, may be especially demanding of learners' visuo-spatial ability, according to Opfermann and colleagues (2017, p. 16). We echo their recommendation to take it "into account whenever research on physics learning includes (at least partly) visual multiple representations" (ibid, p. 16) and to include it as an important learner characteristic when dealing with tasks.

#### 2.4.1.5 Gender

It can also be expected that learning with tasks should also be influenced by gender differences in STEM education, which are widely discussed for both cognitive and affective variables. For achievement, meta-analyses and large-scale studies have found no or only negligible differences for biology, chemistry, and general science (meta-analyses: Becker, 1989, Lee & Burkam, 1996; large scale study: Louis & Mistele, 2012). In the latter study cited, effect sizes for general science and biology of $d = 0.14$ and 0.08 were found. This is a nice example showing that with a large sample size ($N \approx 8{,}000$), small effects can become statistically significant, yet still be of little practical relevance. For physics, small, but non-zero effects in favour of boys have been found ($d = 0.35$, Becker, 1989; $d = 0.25$, Lee & Burkam, 1996; $d = 0.32$, Louis & Mistele, 2012). Likewise, for interest and self-concept, similar small but non-zero effects exist (interest: $d = 0.4$, from the 5,000-pupil UPMAP study in England, Mujtaba & Reiss, 2013; self-concept: $d = 0.27$, OECD average for science in general, OECD, 2007). It should be noted, however, that it is not the subject of physics per se which is uninteresting for girls, but that there are specific contexts which they perceive as much more interesting than others.

Figure 7 offers a revealing finding in this respect: When asking about the same physics topic presented in different contexts, it turns out that girls' interest for a context in the life sciences is much larger than it is for a technical context. In fact, it even runs against the general adolescent decline in interest for science discussed above (sect. 1.4.2). For boys, the interest level for both contexts is similar and rather high (70%). Note that Hoffmann (2002) showed that context (kind of applications and activities) has a much stronger influence on physics interest than the specific content (subject matter): 80% of the variance of physics interest across the items of this study can be attributed to context, and only 20% to the content itself. An example of successful implementation of medical contexts into physics teaching practice can be found in works by Colicchia (2002) and Wiesner and Colicchia (2010), with effect sizes for interest development before and after the sequence of = +0.45 with a medical context context and –0.52 without. Another context creating high levels of interest for both girls and boys is astronomy (ROSE, n.d.; Baram-Tsabari & Yarden, 2005), and a contribution in this thematic issue (Müller et al.) presents other examples of that type based on newspaper story problems, advertisements, and other sources.

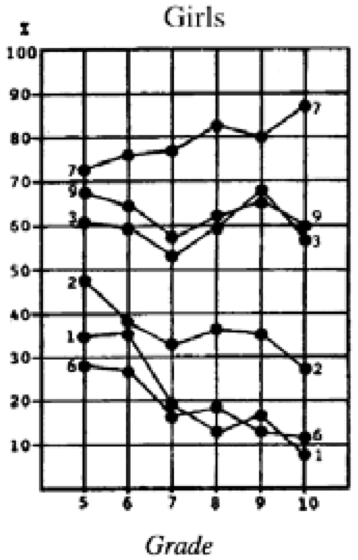

**Fig. 7.** Percentage of girls with "great" and "very great" interest in selected contexts for the topic of mechanics (motion, force, pressure; Hoffmann, 2002).
Curve 7: an artificial heart as a blood pump. Curve 2: pumping petrol from great depths (cf. Hoffmann, 2002 for the other contexts).

## 2.4.2 Classroom and Teaching Characteristics

Just as is the case for learner characteristics, there is broad evidence for the importance of the influence of the instructional setting on learning with tasks. Above, the beneficial effects of reducing anxiety and fear of failure in the classroom were discussed in section 1.4.3, and those of increasing wait time after questions in section 2.2.1.1. Note that these two factors share a reduction in pressure and stress in the classroom atmosphere. In view of its singular importance, the role of feedback will be reviewed in its own section (2.5). Below, we review the effect of two other temporal parameters (time-on-task, spaced/massed practice) and of collaborative learning.

### 2.4.2.1 Time-on-task

Insight and mastery require practice and active work with learning tasks, and an important part of classroom management is therefore time management, with the objective being to maximize effective learning time, or time-on-task. Classroom time management is understood as the interplay of four nested categories of allocated time, instructional time, engaged time, and effective or productive time; the latter is sometimes also called "academic learning time", the time where a learner is "actively, successfully, and productively engaged in learning relevant academic content" (Brodhagen & Gettinger, 2012, p. 33). Meta-analyses support this influence of time-on-task with an average effect size of $d = 0.38$ (Hattie, 2009), while individual studies find a broad range from 0.3 to 2.0 (Brodhagen & Getting, 2012). One thus can concur with Brodhagen and Gettinger (2012) that it has "a strong influence on academic learning and student achievement." However, individual studies find a broad range of effect sizes from 0.3 to 2.0 (Hattie, 2009; Brodhagen & Gettinger, 2012). As in other cases, the breadth of this range draws attention to the fact that the effects of time-on-task are moderated by other factors, (particularly depending on how exactly time on task was operationalized in terms of the above categories).

### 2.4.2.2 Developing Automaticity and Routines

The development of routines by repeated, deliberative practice is considered as one of the key elements of expertise (Hattie & Yates, 2013, chap. 13). Bloom (1986) went so far to call automaticity in the sense of skilled practice the "hands and feet of genius". In the STEM disciplines, routines are needed in many situations: experimental routines, mathematical routines, routines for writing programs, and so forth. The need for practice and appropriate practice tasks thus cannot be doubted, but just what type of practice and tasks? Several teaching characteristics seen in meta-analyses come into play here.

First, the question of the ratio of known to unknown elements in exercise tasks ("drill ratio"[9]). The meta-analysis carried out by Burns (2004) found a clear superiority of higher ratios ($d = 1.2$ for 70-90% known to unknown elements, $d = 0.5$ for < 50%; effect sizes for pre-post comparisons). Second, the question of the learning phase was also investigated in this study, with a clear advantage for the use of drill tasks in the acquisition stage ($d = 1.1$) over the proficiency

---

[9] "Drill" often has a negative connotation of dullness and mindless repetition, but is used by some authors in a more neutral sense of practice tasks for acquiring automaticity and routines (see RiPe Box 4 for some features relevant for this distinction).

development stage (*d* = 0.4). This is consistent with educational recommendations from Comenius (1657; "principle of facility": to proceed from what is easy to what is more difficult, from the more familiar to the less familiar) to Ausubel (1963; "progressive differentiation") and represents in some sense a quantitative element of the concept of the "zone of proximal development" (Vygotsky, 1978). A limitation of Burns' meta-analysis (2004) is that the interaction of "drill ratio" and "learning stage" is not considered. Another limitation is that the above effect sizes are for academic outcomes and learner preferences taken together; when looking at these separately the effect sizes are twice as large for academic outcomes (0.8) than for student satisfaction (0.4), but again not differentiated by "drill ratio" nor by stage.

These limitations notwithstanding, one can conclude in agreement with Hattie (2009, p. 224) that "despite the moans by many adults, students need much drill and practice". For an explanatory framework in cognitive science, Hattie and Yates (2014) discuss the interplay of cognitive automaticity and routine and non-routine thinking in terms of dual system (or dual process) approaches to thought, with a fast-operating, automated system 1, and a slow-operating, non-automated system 2 (as proposed e.g., by Kahneman (2011) in his book *Thinking, Fast and Slow*).

A third relevant teaching characteristic is that of spaced (distributed) versus massed practice. "[W]ith any considerable number of repetitions a suitable distribution of them over a space of time is decidedly more advantageous than the massing of them at a single time". These were the words of Hermann , Ebbinghaus (1885/1964), a German pioneer of experimental psychology, as far back as 1885. He was referring to what is now variably known as a "distributed-" or "spaced practice" effect, or "spacing" effect, whereby long-term recall of previously learned content can be improved by distributing or "spacing" a certain set of learning content across multiple time periods compared to the same amount of time being dedicated to the same content in a single session (Wiseheart et al., 2019). The meta-analytical effect size has been reported as *d* = 0.71 (Hattie, 2009). In practical terms, three to four exposure sessions are often found to be needed to ensure stable learning, with an interval of more than 24 hours between sessions or more complex tasks. The start of a lesson is a suitable opportunity for such repetition (Hattie, 2009).

A fourth classroom characteristic of interest here is the use of computers and other ICT. Arguments in favour of this are the increased autonomy and individualisation, as well as the possibility of gamification for practice tasks (Hattie, 2009; Cheung et al., 2017). Indeed, a recent meta-analysis reported an effect size of *g* = 0.58 for computer-based practice tasks (when compared to control groups taught without the use of digital tools; Hillmayr et al, 2020). Interestingly, meta-analytic effect sizes have increased over the years (Hattie, 2009: 0.34; Burns & Bozeman, 1981: 0.17). This can be seen as an indication of an improved instructional quality for that matter, both on the technological level, and in view of Hattie's (2009) exhortation that practice "does not need to be dull and boring, but can be, and indeed should be, engaging and informative" (see also RiPe Box 4).

| **RiPe Box 4. Practice tasks: "drill & skill" versus "drill & kill".** Some guidelines for practice (Hattie, 2009, 185pp unless specified otherwise). |
|---|
| Provide opportunities to enhance mastery and fluency. |
| Avoid dull and repetitive tasks. |
| Avoid massed practice. |
| Avoid isolated practice of skills and routines (Goulding & Kyriacou, 2008) |
| Embed within/connect to tasks with deeper and conceptual understanding. |
| Provide multiple different contexts and connections. |
| Proceed from familiar/simple to less familiar/more complex. |
| Provide feedback. |
| Make success criteria and their attainment visible. |

#### 2.4.2.3 Collaborative Learning Tasks

A strong research basis exists for organising work on learning tasks in a collaborative format (Topping et al., 2017). Cooperative forms of learning are currently understood to be important elements of a systematic perspective on teaching quality, as an important contextual factor at the social level, as a means of cognitive activation (verbalisation, communication, confrontation with other points of view), and as a motivational factor through social relatedness (in the sense of "basic needs" according to Deci & Ryan, 2000).

Meta-analyses support the idea that collaborative learning has strong positive effects for many learning domains (collaborative vs. individual learning, achievement: $d \approx 0.6$; quality of reasoning: *d* = 0.97; Hattie, 2009; Johnson & Johnson, 2009). For science specifically, results for achievement are even more positive (*d* = 0.95, Schroeder et al., 2007). While it is not trivial to find effective and practically feasible collaborative task arrangements for each teaching situation, existing research provides many examples of successful features and learning arrangements (Topping et al., 2017).

### 2.5 Feedback

According to Hattie (2009), feedback on an individual's learning is one of the 10 most influential factors for success overall (*d* = 0.73). It is also consistently found to be a major factor for the efficacy and efficiency of various forms of tasks. In the case of homework (sect. 2.2.2), feedback leads to a threefold increase in learning (with: $d = 0.83$; without: $d = 0.28$). For an inverted mathematics classroom (sect. 2.2.3), formative assessment corresponds to an almost twofold increase in learning (with: $d = 0.57$, without: $d = 0.3$). Note that enhancing self-beliefs (sect. 1.3.4), and (thereby) reducing anxiety, also originate from a certain kind of feedback. Beyond the immediate impact on the affective level, this can also have considerable effects on learning (e.g., reducing anxiety on learning, across all disciplines: $d = 0.4$; mathematics: $d = 0.56$).

## 3 Conclusions and Perspectives

The purpose of the present contribution is to provide an account of quantitative, mainly meta-analytical evidence on the effects of different forms of tasks in science education, and about the influences of several learner and classroom characteristics.

Several important limitations must be pointed out. First, it has already been stated that this work focuses (i) on *general* results and therefore does not discuss important aspects specific to STEM disciplines taken individually; (ii) on *cognitive* aspects, discussing affective aspects only insofar as they represent an important background for this focus. It has also been emphasised that this contribution is not a meta-analysis or systematic review in itself, but rather an account of best (meta-analytical) evidence related to the use of tasks, with a focus on teaching practice and teacher education.

Second, due to space restrictions, certain important general aspects and topics related to tasks could not be included. Notably, diagnosis and evaluation are of course essential components of "task culture" and would require a systematic account of meta-analytical evidence of their own. As an example, we cite Liu (2012) for a comprehensive review on concept tests with their special importance for science education.

Third, the bridging between research and practice offered here (e.g., in the RiPe boxes) is far from being complete. Fullan and Scott (2009, p. 73) say "Failed implementation is the bane of all change aspirations". To prevent this from ruining the great educational potential of task culture, systematic effort is needed to implement it into teacher education.

Fourth, the perspective of evidence-based (science) education (using effect sizes, meta-analyses, etc.) clearly has its limitations. By no means is it intended to replace educational reflection, but it can be a useful starting point or element of this reflection, and a good reason to take a closer look at individual studies.

In that sense, we believe that the evidence-based perspective on tasks taken here can be helpful, and in some respects healthy.

One beneficial effect to be underlined is connected to the problem of "classroom" or "educational illusions", by which we mean the overestimation of features of educational quality (by analogy to optical illusions; Müller & Weiss, 2018). For instance, with respect to certain features supposed to motivate pupils for a given task, a very large gap between teachers' and students' perceptions must be noted (relatedness to real life: $d = 1.34$; interest: $d = 1.60$; sect. 1.4.2). Moreover, it has been shown that it can be illusory to believe that mere relevance is sufficient to create interest (Fig. 4). Regarding classroom questioning, wait time after a question is found to be < 3 s, often 1 s, but teachers are of course convinced of giving their students enough time to reflect (Rowe, 1974, Tobin, 1987, Heinze & Erhard, 2006; sect. 2.2.1.1). They are also convinced that teaching questions should require reflection and promote understanding, but in reality, there are 60–80% "surface thinking questions" (repetition, completion, one-word answers, etc; across the subjects, unchanged findings documented for 20 years, Hattie, 2009). A closely related finding concerns the share of teacher speaking time, empirically found to be between 60% and 80% for STEM disciplines, while teachers tend to rate their share much lower (physics: 60% (Seidel, 2003); chemistry: ≥ 65% (Sumfleth & Pitton, 1998), mathematics: 78%, (DESI, 2006). We are convinced that it is important to take account of these "educational illusions" in a more conscientious and systematic way. This is not at all about criticising teachers, but with a view to improving teacher education and classroom practice.

Another positive effect of an evidence-based approach is to counter simplistic, sometimes ideological convictions. Regarding tasks, a first example is homework, where one can see that such convictions ("homework is necessary"; "homework is punishment of families") do not lead anywhere; it is feedback that turns homework into an educationally useful practice. Different and new modes of feedback that also work for homework (such as "rapid feedback techniques" (Kennedy et al., 2016), or formative assessment quizzes, sect. 2.2.3) certainly merit strong attention from science education research and development.

Inquiry-based learning is another highly relevant example for science education. Again, the teacher can have a decisive influence: in the form of guidance (effect sizes with and without 0.65 and 0.25, respectively, sect. 2.1.1). As in the case of homework, a differentiated, quantitatively based stance should be preferred.

We close with some perspectives on two aspects considered to be central for science and science education. First, multiple representations, well known as essential means of domain-specific reasoning across all scientific disciplines, and mathematics (Tsui & Treagust, 2013; Gilbert & Treagust, 2009; Treagust et al., 2017; Janvier, 1987). Representational competence, that is to say the flexible and coherent use of multiple representations, is not easy to achieve. Recent work on specific learning tasks (representational activity tasks, RATs), asking learners to explicitly work on various types of coherent connections, such as comparing, completing, and correcting representations, has found a considerable supporting effect ($d = 0.69$) when comparing to a control group without RATs (but otherwise identical content, lesson plan, duration, Scheid et al., 2019).

Second, experimentation. The beneficial effects of experimental and hands-on activities are discussed in section 2.3.1. A recent development is the use of mobile devices (smartphones, tablets) as experimental tools, allowing for a variety of educational tasks (observation, measurement, analysis, reflection, creativity) across many scientific topics (Kuhn & Vogt, since 2012; Mathevet et al., 2019; Brenner, 2017; Science on Stage Europe, 2014; Darmendrail et al, 2019; Kuhn & Vogt, 2019); with many applications to science in everyday settings or to interdisciplinary phenomena (in the sense of sect. 1.4.2: Müller et al, 2016; Thoms et al., 2019; Haglund & Schönborn, 2019; Darmendrail & Müller, 2020; Kasper & Vogt, 2020); with applications to environmental issues, for example for radioactivity monitoring, sometimes with an overlap to citizen science projects (Keller et al., 2019); and often including elements of creativity and serendipity (Müller et al., 2016; Varra et al., 2020; Kasper & Vogt, 2020). In addition to this broad range of educational content and purposes, Swarat et al. (2012) make an important link to affective variables, stating that "technology may enhance student interest by connecting students with real data and thus promoting a sense of authenticity". Empirical investigations of this approach for experimental tasks are underway, providing increasing evidence for its positive effects on interest (Hochberg et al., 2018) and learning (Klein et al., 2018; Hochberg et al., 2020).

In view of the evidence and discussion presented in this contribution, we summarise "task culture" as a deliberative practice that:

- encourages mastery and fluency (including conceptual and deep understanding);
- involves feedback;
- takes the form of multiple different experiences and contextual embeddings (to facilitate learning transfer);
- is embedded within a larger framework of learning;
- fosters motivation and self-concept.

In any case, the rich variety of forms, aims, and evidence-based outcomes of tasks goes far beyond dull, repetitive "drill and practice", which is restricted to surface knowledge. This said, tasks may very well sometimes "only" promote understanding and interest, and not always "competencies".

We hope that this review of meta-analytical and other empirical results related to the topic of this thematic issue is helpful and stimulating for researchers and practitioners. In any case, we are convinced of the importance of tasks and task culture in science education. Karl Popper (1999) says "all life is problem solving", and the latter is certainly a large part of education.

## References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1. http://doi.org/ 10.1037/a0021017

Ableitinger, C. (2011). *Biomathematische Modelle im Unterricht* [Biomathematical models in teaching]. Vieweg & Teubner.

Adams, R. (2009). Recent publications on task-based language teaching: a review. *International Journal of Applied Linguistics*, *19*(3), 339–355. https://doi.org/10.1111/j.1473-4192.2009.00239.x

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33, 131–152. https://doi.org/10.1016/S0360-1315(99)00029-9

Aker, L. B., & Ellis, A. K. (2019). A meta-analysis of middle school students' science engagement. *International Dialogues on Education: Past and Present*, *6*(1), 9–24.

Andersson-Bakken, E., Jegstad, K. M., & Bakken, J. (2020). Textbook tasks in the Norwegian school subject natural sciences: what views of science do they mediate? *International Journal of Science Education*, *42*(8), 1320–1338. https://doi.org/10.1080/09500693.2020.1756516

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214. https://doi.org/10.3102/00346543070002181

Ausubel, D. (1963). *The psychology of meaningful verbal learning*. Grune & Stratton.

Ausubel, D. (1978). *Educational Psychology: A Cognitive View*. Holt, Rinehart & Winston.

Baram-Tsabari, A., & Yarden, A. (2005). Characterizing children's spontaneous interests in science and technology. *International Journal of Science Education*, *27*(7), 803–826. https://doi.org/10.1080/09500690500038389

Barton, M. L., & Jordan, D. L. (2001). Teaching reading in science. In: *Teaching Reading in the Content Areas: If Not Me, Then Who?* Association for Supervision and Curriculum Development. http://www.ascd.org/ASCD/pdf/books/barton2001_sample_chapters.pdf

Becker, B. J. (1989). Gender and science achievement: a reanalysis of studies from two meta-analyses. *Journal of Research in Science Teaching*, *26*(2), 141–169. https://doi.org/10.1002/tea.3660260206

Besse, M., Fragnière, S., Müller, A., Piguet, M., Dubois, L., Miéville, D., Schoeb, S., & Schumacher, D. (2019). Learning about archaeology and prehistoric life: the effects of two workshops in primary education. *Science & Education*, *28*(6-7), 759–795.

Beywl, W., & Zierer, K. (2018). 10 Jahre "Visible Learning" – 10 Jahre "Lernen sichtbar machen". *Pädagogik*, *70*(9), 36–41.

Bialek, W., & Botstein, D. (2004). Introductory science and mathematics education for 21st century biologists. *Science, 303*(5659), 788–790. https://doi.org/10.1126/science.1095480

BLK (Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung). (1997). Gutachten zur Vorbereitung des Programms "Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts" [Report on the preparation of the programme "Increase of the efficiency of mathematics and science teaching"]. BLK. http://www.blk-bonn.de/papers/heft60.pdf

Black, P., & Harrison, C. (2001). Feedback in questioning and marking: the science teacher's role in formative assessment. *School Science Review, 82(*301), 55–61.

Bloom, B. S. (1986). Automaticity: The hands and feet of genius. *Educational Leadership, 43*(5), 70–77.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: a quantitative synthesis. *Review of Educational Research, 53*(4), 499–518. https://doi.org/10.3102/00346543053004499

Brenner, E. D. (2017). Smartphones for teaching plant movement. *The American Biology Teacher, 79*(9), 740–745. https://doi.org/10.1525/abt.2017.79.9.740

Brodhagen, E. M., & Gettinger, M. (2012). Academic learning time. In: N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning*. Springer. https://doi.org/10.1007/978-1-4419-1428-6_5460

Brualdi, A.C. (1998). Classroom questions (ED422407). ERIC. https://files.eric.ed.gov/fulltext/ED422407.pdf

Brunsell, E., Fleming, M. A. (2014). *Engaging Minds in Science and Math Classrooms: The Surprising Power of Joy.* Association for Supervision and Curriculum Development

Burns, M. K. (2004). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education, 25*(3), 167–173.

Bums, P. K., & Bozeman, W. C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship? *Educational Technology*, 21, 37–38.

Caillé, A. (1995). L'enseignement des sciences de la nature au primaire [Natural sciences teaching in primary]. Presses de l'Université du Québec.

Calvo, P., & Gomila, T. (Eds.). (2008). Handbook of Cognitive Science: An Embodied Approach. Elsevier.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology, 105*(2), 380. https://doi.org/10.1037/a0031084

Carter, C. S., Larussa, M. A., & Bodner, G. M. (1987). A study of two measures of spatial ability as predictors of success in different levels of general chemistry. *Journal of Research in Science Teaching, 24*(7), 645–657. https://doi.org/10.1002/tea.3660240705

Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load.* San Francisco: Pfeiffer.

Chen, W. C., & Whitehead, R. (2009). Understanding physics in relation to working memory. *Research in Science & Technological Education, 27*(2), 151–160.

Cheung, A., Slavin, R. E., Kim, E., & Lake, C. (2017). Effective secondary science programs: A best evidence synthesis. *Journal of Research in Science Teaching, 54*(1), 58–81.

Chin, C. (2004). Students' questions: fostering a culture of inquisitiveness in science classrooms. *School Science Review,* 86, 314, 107–112.

Chu, Y.-C. (2008). Learning difficulties in genetics and the development of related attitudes in Taiwanese junior high schools [Doctoral dissertation, University of Glasgow]. http://theses.gla.ac.uk/168/

Crissman, J. K. (2006). The design and utilization of effective worked examples: a meta-analysis [Doctoral dissertation, University of Nebraska-Lincoln]. https://digitalcommons.unl.edu/dissertations/AAI3208114/

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Erlbaum.

Cohen, P., Cohen, J., Aiken, L. S. & West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315–346. https://doi.org/10.1207/S15327906MBR3403_2

Colicchia, G. (2002). Physikunterricht im Kontext von Medizin und Biologie: Entwicklung und Erprobung von Unterrichtseinheiten zur Steigerung des Interesses und für den fachübergreifenden Physikunterricht [Physics teaching in the context of medicine and biology: development and testing of teaching units to increase interest and for beyond-subject physics teaching] [Doctoral dissertation, LMU München]. https://doi.org/10.5282/edoc.274

Comenius, J. A. (1657). Didactica Magna. Part I of Opera didactica omnia. Amsterdam: Laurentius de Geer; in German: Flitner, A. (Ed.) (1970). Comenius – Grosse Didaktik. Kuÿpper; in English: Comenius, J. A., & Keatinge, M. W. (1997). The great didactic of John Amos Comenius. A. & C. Black

Cooper, H. (1989). Synthesis of research on homework. Educational Leadership, 47 (3), 85–91.

Crahay, M. (2006). Un bilan des recherches processus-produit. L'enseignement peut-il contribuer à l'apprentissage des élèves et, si oui, comment ? [A summary of process-product research: can teaching contribute to pupil learning, and if so, how?]. Université de Genève. https://archive-ouverte.unige.ch/unige:21882

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438–481. https://doi.org/10.3102/00346543058004438

Csapó, B. (1992). Improving operational abilities in children. In: A. Demetriou, M. Shayer, A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development. Implications and applications for education.* Routledge.

CTGV (Cognition and Technology Group at Vanderbilt). (1992). The Jasper experiment: an exploration of issues in learning and instructional design. *Educational Technology Research and Development,* 40, 65–80. https://doi.org/10.1007/BF02296707

Davies, P. (1999). What is evidence based education? *British Journal of Educational Studies, 47*(2), 108–121. https://doi.org/10.1111/1467-8527.00106

Darmendrail, L., Keller, O., & Müller, A. (2019). Data, Data Everywhere, and Quite a Bit(e) to Learn. In: A. F. Koch, S. Kruse, & P. Labudde. (Eds.). *Zur Bedeutung der technischen Bildung in Fächerverbünden: Multiperspektivische und interdisziplinäre Beiträge aus Europa*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-25623-4_3

Darmendrail, L., & Author. (2020). Helmholtz in the kitchen: a frying pan as a volume resonator. *European Journal of Physics, 41*(3), 035804. https://doi.org/10.1088/1361-6404/ab6fba

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry,* 11, 227–268. https://doi.org/10.1207/S15327965PLI1104_01

DESI Consortium (Eds.) (2006). Unterricht und Kompetenzerwerb in Deutsch und Englisch: Zentrale Befunde der DESI Studie [Teaching and competence acquisition in German and English: main findings of the DESI study]. Deutsches Institut für Internationale Pädagogische Forschung

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332, 862–864. https://doi.org/10.1126/science.1201783

Donati, G., Meaburn, E. L., & Dumontheil, I. (2019). The specificity of associations between cognition and attainment in English, maths and science during adolescence. *Learning and Individual Differences,* 69, 84–93.

Dörfler, W. (2000). Means for meaning. In P. Cobb, E. Yackel, & K. McClain (Eds.), Symbolizing and Communicating in Mathematics Classrooms. Perspectives on Discourse, Tools and Instructional Design. Erlbaum.

Doyle, W. (1988). Work in mathematics classes: the context of students' thinking during instruction. *Educational Psychologist,* 23, 167–180.

Ebbinghaus, H. (1964). Memory: A Contribution to Experimental Psychology (H. A. Ruger, C. E. Bussenius, & E. R. Hilgar, Trans.). Dover Publications. (Original work published 1885)

Eccles, J. S., Fredricks, J. A., & Epstein, A. (2015). Understanding well developed interests and activity commitment. In: K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), Interest, the Self, and K-16 Mathematics and Science Learning. American Educational Research Association. https://dx.doi.org/10.3102/978-0-935302-42-4_3

Educational Psychologist (Ed.) (2006, 2007). Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why unguided learning does not work: An analysis of the failure of discovery learning, problem-based learning, experiential learning and inquiry-based learning. *Educational Psychologist, 41*(2), 75–86; Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist,* 42, 99–107; Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational psychologist, 42*(2), 115–121.

Educational Researcher (Ed.). (2008). Perspectives on Evidence-Based Research in Education (special issue). Educational Researcher 37(1).

Eberbach, C., & Hmelo-Silver, C. (2014) Inquiry, learning through. In: R. Gunstone (Ed.). *Encyclopedia of Science Education.* Springer. https://doi.org/10.1007/978-94-007-6165-0_192-3

Ellis, R. (2003). Task-based Language Learning and Teaching. Oxford University Press.

Finnigan, K. S., & Daly, A.J. (Eds.). (2014). Using Research Evidence in Education: From the Schoolhouse Door to Capitol Hill. Springer.

Fleming, M. L., & Malone, M. R. (1983). The relationship of student characteristics and student performance in science as viewed by meta analysis research. *Journal of Research in Science Teaching,* 20, 481-495.

Fonseca, B. A., & Chi, M. T. H. (2017). In: R. E. Mayer, & P. A. Alexander (Eds.), Handbook of Research on Learning and Instruction. Routledge.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wanderoth, M. P. (2014) Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences,* 111, 8410–8415. https://doi.org/10.1073/pnas.1319030111

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of Experimental Psychology:* General, *141*(1), 2. https://doi.org/10.1037/a0024338

Furtak, E.M., Seidel, T., Iverson, H., Briggs, D. (2012). Experimental and quasi experimental studies of inquiry-based science teaching: a meta-analysis. *Review of Educational Research, 82*(3), 300–329. https://doi.org/10.3102/0034654312457206

Fullan, M., & Scott, G. (2009). Turnaround Leadership for Higher Education. Jossey-Bass.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18*(1), 1–16. http://doi.org/10.1002/acp.934

Gauthier, C., & Jobin, V. (2009). Moins c'est souvent mieux : guide des principes d'enseignement multimédia élaborés à partir de recherches en psychologie cognitive [Less is Often Better: A Guide to Multimedia Teaching Principles Based on Cognitive Psychology Research]. Presses de l'Université Laval.

Gehrig, M., Gardiol, L. & Schaerrer, M. (2010). Der MINT-Fachkräftemangel in der Schweiz [The lack of STEM skills in Switzerland]. Staatssekretariat für Bildung und Forschung. https://www.sbfi.admin.ch/dam/sbfi/it/dokumente/webshop/2010/der_mint-fachkraeftemangelinderschweiz.pdf

Gilbert, J. K., & Treagust, D. F. (Eds.). (2009). Multiple Representations in Chemical Education. Springer.

Glynn, S., & Koballa, T.R. (2005). The contextual teaching and learning approach. In: R.E. Yager (Ed.), Exemplary Science: Best Practices in Professional Development. National Science Teacher Association Press.

Goulding M., & Kyriacou, C. (2008). A systematic review of the use of ICTs in developing pupils' understanding of algebraic ideas. Report. In: Research Evidence in Education Library. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2380i

Gross L. J., Brent, R., Hoy, R. (2004). Points of view: the interface of mathematics and biology: interdisciplinarity and the undergraduate biology curriculum: finding a balance. *Cell Biology Education*, 3, 85–87; 88–89; 90–92.

Güdel, K., Heitzmann, A., & Müller, A. (2019). Self-efficacy and (vocational) interest in technology and design: an empirical study in seventh and eighth-grade classrooms. *International Journal of Technology and Design Education,* 29, 1053–1081. https://doi.org/10.1007/s10798-018-9475-y

Haglund, J., & Schönborn, K. J. (2019). The pedagogical potential of infrared cameras in biology education. *The American Biology Teacher, 81*(7), 520–523. https://doi.org/10.1525/abt.2019.81.7.520

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentille, J., Lauffer, S., Stewart, J., Tilghman, S. M., Wood, W. B. (2004). Scientific teaching. Science, 304, 521–522. https://doi.org/ 10.1126/science.1096022

Hattie, J., (2009). Visible Learning – A Synthesis of Over 800 Meta analyses relating to achievement. Routledge.

Hattie, J. (2012). Visible Learning for Teachers. Routledge.

Hattie, J. (2017). L'apprentissage visible pour les enseignants : Connaître son impact pour maximiser le rendement des élèves [Visible Learning for Teachers: Understanding Impact to Maximise Learner Achievement]. Presses de l'Université du Québec.

Hattie, J., & Yates, G. C. R. (2014). Visible learning and the science of how we learn. Corwin.

Häussler, P., Bünder, W., Duit, R., Gräber, W., Mayer, J. (1998). Naturwissenschaftsdidaktische Forschung: Perspektiven für die Unterrichtspraxis. [Science Education Research: Perspectives for Teaching Practice] Institut für die Pädagogik der Naturwissenschaften.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*(2), 212–242. https://doi.org/10.3102/00346543056002212

Heinze, A., & Erhard, M. (2006). How much time do students have to think about teacher questions? An investigation of the quick succession of teacher questions and student responses in the German mathematics classroom. *Zentralblatt für Didaktik der Mathematik, 38*(5), 388–398. https://doi.org/10.1007/BF02652800

Henry, M. A., Shorter, S., Charkoudian, L., Heemstra, J. M., & Corwin, L. A. (2019). FAIL is not a four-letter word: a theoretical framework for exploring undergraduate students' approaches to academic challenge and responses to failure in STEM learning environments. *CBE—Life Sciences Education, 18*(1), 1-17.

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897.

Hoban, R. A., Finlayson, O. B., Nolan, B. C. (2013). Transfer in chemistry: a study of students' abilities in transferring mathematical knowledge to chemistry. *International Journal of Mathematical Education in Science and Technology, 44*(1), 14–35. https://doi.org/10.1080/0020739X.2012.690895

Hochberg, K., Becker, S. Louis, M., Klein, P. & Author. (2020). Using smartphones as experimental tools – a follow-up: cognitive effects by video analysis and reduction of cognitive load by multiple representations. *Journal of Science Education and Technology,* 29(2), 303–317. https://doi.org/10.1007/s10956-020-09816-w

Hochberg, K., Kuhn, J., & Müller, A. (2018). Using smartphones as experimental tools—effects on interest, curiosity, and learning in physics education. *Journal of Science Education and Technology, 27*(5), 385–403.

Hodgson, C., & Pyle, K. (2010). A Literature Review of Assessment for Learning in Science. National Foundation for Educational Research. https://www.nfer.ac.uk/media/1556/aas01.pdf

Hoffmann, L. (2002). Promoting girls' interest and achievement in physics classes for beginners. *Learning and Instruction, 12*(4), 447-465. https://doi.org/10.1016/S0959-4752(01)00010-X

Hoffmann, L., Häussler, P., & Lehrke, M. (1998). Die IPN-Interessenstudie Physik [The IPN Interest Study: Physics]. Institute for Science Education.

Ingram, J., & Elliott, V. (2016). A critical analysis of the role of wait time in classroom interactions and the effects on student and teacher interactional behaviours. *Cambridge Journal of Education, 46*(1), 37–53. https://doi.org/10.1080/0305764X.2015.1009365.

Jacquin, M., & Müller, A. (2013). La tâche scolaire comme lieu d'intégration des savoirs et compétences ? Deux études dans l'enseignement des langues étrangères et des sciences [The school task as a place to integrate kowledge and compétences? Two studies in foreign language and science teaching]. In: Savoirs, compétences. Approches comparatives de l'organisation des formes et des contenus de l'étude : variations et constantes disciplinaires, institutionnelles, culturelles. Troisième Colloque International de l'Association pour des Recherches Comparatistes en Didactique, 9-12 January 2013. Université de Marseille.

Janvier, C. (Ed.). (1987) Problems of Representation in the Teaching and Learning of Mathematics. Erlbaum.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher,* 38, 365–379. https://doi.org/10.3102/0013189X09339057

Kahneman, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.

Karam, R. (2015). Thematic issue: The interplay of physics and mathematics: Historical, philosophical and pedagogical considerations. *Science & Education,* 24, 487–805.

Kasper, L., & Vogt, P. (2020). Corkscrewing and speed of sound: A surprisingly simple experiment. *The Physics Teacher, 58*(4), 278–279. https://doi.org/10.1119/1.5145480

Keller, O., Benoit, M., Author, & Schmeling, S. (2019). Smartphone and tablet-based sensing of environmental radioactivity: mobile low-cost measurements for monitoring, citizen science, and educational purposes. Sensors, 19(19), 4264. https://doi.org/10.3390/s19194264

Kennedy, I. G., Latham, G., Jacinto, H. (2016). Education Skills for 21st Century Teachers: Voices From a Global Online Educators' Forum. Springer.

Khine, M. S. (2017). Visual spatial Ability in STEM Education. Springer.

Kircher, E., Girwidz, R., & Häußler, P. (2009). Physikdidaktik: Theorie und Praxis [Physics Teaching: Theory and Practice]. Springer.

Konsortium HarmoS Naturwissenschaften+ (2010). Naturwissenschaften. Wissenschaftlicher Kurzbericht und Kompetenzmodell. Bern: Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).

Krapp, A., & Prenzel, M. (2011). Research on interest in science: theories, methods, and findings. International Journal of Science Education, 33(1), 27–50. https://doi.org/10.1080/09500693.2010.518645

Kuhn, J., & Vogt, P. (since 2012). iPhysicsLabs (Series), Column Editors' Note. The Physics Teacher, 50, 372–373.

Lee, V. E., & Burkam, D. T. (1996). Gender differences in middle grade science achievement: subject domain, ability level, and course emphasis. Science Education, 80, 613–650.

Lee, J., & Stankov, L. (2016). Non-cognitive influences on academic achievement: evidence from PISA and TIMSS. In: M. S. Khine, & S. Areepattamannil. (2016), Non-cognitive Skills and Factors in Educational Attainment. SensePublishers, 153–169.

Leopold, D. G., & Edgar, B. (2008). Degree of mathematics fluency and success in second-semester introductory chemistry. *Journal of Chemical Education*, 85, 724–731. https://doi.org/10.1021/ed085p724

Levin, A., Arnold, K.-H. (2006). Selbstgesteuertes und Selbstreguliertes Lernen [Self-supported and self-regulated learning], (p. 206-214). In: K.-H¨. Arnold, J. Wiechmann, & U. Sandfuchs. (Eds.), Handbuch Unterricht. Klinkhardt.

Liew, C. W., & Treagust, D. F. (1995). A predict-observe-explain teaching sequence for learning about students' understanding of heat and expansion of liquids. *Australian Science Teachers' Journal, 41*(1), 68–71.

Lipsey, M., & Wilson, D. (2003). Practical Meta-Analysis, Applied Social Research Methods Series (Vol. 49). Sage.

Liu, X. (2012). Developing measurement instruments for science education research. In: B. J. Fraser, K. Tobin, C. J. McRobbie. (Eds.), Second International Handbook of Science Education. Springer. https://doi.org/10.1007/978-1-4020-9041-7_43

Lo, C. K., Hew, K. F., & Chen, G. (2017). Toward a set of design principles for mathematics flipped classrooms: a synthesis of research in mathematics education. *Educational Research Review*, 22, 50-73. https://doi.org/10.1016/j.edurev.2017.08.002

Loretan, C., Müller, A., Weiss, L. (2018a). Order-of-magnitude reasoning as part of science literacy: An intervention study. In: O.E. Finlayson, E. McLoughlin, S. Erduran, & P. Childs (Eds.), Electronic Proceedings of the ESERA 2017 Conference. Research, Practice and Collaboration in Science Education. Dublin City University.

Loretan, C., Müller, A., & Weiss, L. (2018b). Fostering semi-quantitative reasoning in the classroom with worked examples as educational scaffold. *Progress in Science Education, 1*(1).

Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. International *Journal of Science and Mathematics Education, 10*(5), 1163–1190. https://doi.org/10.1007/s10763-011-9325-9

Löwen, K., Baumert, J., Kunter, M., Krauss, S., & Brunner, M. (2013). The COACTIV research program: methodological framework. In: M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), Cognitive activation in mathematics classrooms and professional competence of teachers: Results from the COACTIV project. Springer. https://doi.org/10.1007/978-1-4614-5149-5_5

Martin, A. J. (2012a). Fear of failure in learning. In: N. Seel. (Ed.), Encyclopedia of the Sciences of Learning. (p. 1276–1278). Springer,

Martin, A. J. (2012b) Motivation Enhancement. In: N. Seel. (Ed.), Encyclopedia of the Sciences of Learning. (p.2339–2342). Springer,

Martin, A. J. (2004). School motivation of boys and girls: Differences of degree, differences of kind, or both?. *Australian Journal of psychology, 56*(3), 133-146.

Marzano, R., Gaddy, B., Dean, C. (2000). What Works In Classroom Instruction. Association for Supervision and Curriculum Development

Marzano, R. J., Pickering, D. J., and Pollock, J. E. (2001). Classroom Instruction That Works: Research-Based Strategies for Increasing Student Achievement. Alexandria (VA, USA): Association for Supervision and Curriculum Development.

Mathevet, R., James, E., Fabre, C., Lamrani, N., Martin, S., Castro, J., Staacks, S., Marchou, P. (2019). Une approche quantitative de la loi de Beer-Lambert avec un smartphone, parties I & II [A quantitative approach to Beer-Lambert's law using a smartphone, parts I & II]. Bulletin de l'Union des Physiciens, 1017, 1079–1089 & 1019, 1357–1366.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, 59, 14–19.

McLaren, B.M., Lim, S., & Koedinger, K.R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In: B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society, 2176-2181.

Meltzer, D. E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics: a possible "hidden variable" in diagnostic pretest scores. *American Journal of Physics*, 70, 1259.

Meltzer, D. E. (2005). Relation between students' problem-solving performance and representational format. *American Journal of Physics,* 73, 463–478. https://doi.org/10.1119/1.1862636

Meltzer, D. E., & Thornton, R. K. (2012). Resource letter ALIP–1: active-learning instruction in physics. *American Journal of Physics, 80*(6), 478–496. https://doi.org/10.1119/1.3678299

Merriam-Webster. (n.d.). Merriam-Webster.com dictionary. Retrieved 28 February 2021, from https://www.merriam-webster.com/dictionary/culture

Millar, R., Leach, J., Osborne J., & Ratcliffe, M. (2006). Improving Subject Teaching: Lessons from Research in Science Education. Routledge.

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, *79*(3), 1129–1167. https://doi.org/10.3102%2F0034654309337522

Muckenfuß, H. (1996) Orientierungswissen und Verfügungswissen. Zur Ablehnung des Physikunterrichts durch die Mädchen [orientation knowledge and availability knowledge. On girls' tendency away from physics class]. *Unterricht Physik, 31*(7), 20–25.

Mujtaba, T., & Reiss, M. J. (2013). Inequality in experiences of physics education: secondary school girls' and boys' perceptions of their physics education and intentions to continue with physics after the age of 16. *International Journal of Science Education, 35*(11), 1824–1845. https://doi.org/10.1080/09500693.2012.762699

Müller, A., Del Notaro, C., Dubois, L., Jacquin, M., & Lombard, F. (2013). Tâches, exercices, problèmes : dispositifs didactiques et éléments d'une culture d'apprentissage. Un regard transdisciplinaire et interdidactique [Tasks, exercises, problems: teaching tools and éléments of a learning culture. A transdisciplinary and interdidactic view]. In: J. Dorier, F. Leutenegger, & B. Schneuwly. Didactique en construction, constructions en didactiques Université de Genève.

Müller, A. (2018a). (Science) Teacher Education in a Knowledge Society. Good Practice at the University of Geneva. https://arxiv.org/abs/1807.00974

Müller, A. (2018b) Evidenzbasiertes Lehren und Lernen von Physik - Was sonst? [Evidence-based physics teaching and learning – what else?]. In: D. Brovelli. (Ed.), *Wirksamer Physikunterricht.* Schneider Verlag, 138–150.

Müller, A. (2019a). Physics Education Research - An Applied Science (part 1). Progress in Physics no. 68. Communications of the Swiss Physical Society, 58, 28–33.

Müller, A. (2019b). Physics Education Research - An Applied Science (part 2). Progress in Physics no. 68. Communications of the Swiss Physical Society, 59, 25–29.

Müller, A., & Weiss, L. (2018). Educational illusions. In: Pixel Associazione Culturale (Ed.), Proceedings of the 7th Conference on New Perspectives in Science Education, 475

Müller, A., Hirth, M., & Kuhn, J. (2016). Tunnel pressure waves - a smartphone inquiry on rail travel. *The Physics Teacher, 54*(2), 118–119. https://doi.org/10.1119/1.4940180

Murphy, P., & Whitelegg, E. (2006). Girls in the physics classroom: a review of the research on the participation of girls in physics. Institute of Physics. http://oro.open.ac.uk/6499/

Neidorf, T., Arora, A., Erberber, E., Tsokodayi, Y., & Mai, T. (2020). Student Misconceptions and Errors in Physics and Mathematics: Exploring Data from TIMSS and TIMSS Advanced. Springer.

Nentwig, P., & Waddington, D. (Eds.). (2005). Making it Relevant: Context based Learning of Science. Waxmann.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: a meta analysis. *Review of Educational Research, 76*(3), 413–448. https://doi.org/10.3102%2F00346543076003413

Neubrand, M., Jordan, A., Krauss, S., Blum, W., & Löwen, K. (2013). Task analysis in COACTIV: examining the potential for cognitive activation in German mathematics classrooms. In: M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), Cognitive Activation in Mathematics Classrooms and Professional Competence of Teachers: Results from the COACTIV Project. Springer, 125-144.

OECD (Organisation for Economic Co-operation and Development). (2007). PISA 2006: Science Competencies for Tomorrow's World, Volume 1: Analysis. OECD. http://www.oecd.org/unitedstates/39722597.pdf

Opfermann, M., Schmeck, A., Fischer, H. E. (2017). Multiple representations in physics and science education – why should we use them? In: D. F. Treagust, R. Duit, & H. E. Fischer (Eds.), Multiple Representations in Physics Education. Springer, 1–22.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Ostermeier, C., Prenzel, M., & Duit, R. (2010). Improving science and mathematics instruction: The SINUS project as an example for reform as teacher professional development. International journal of science education, 32(3), 303-327.

Pepper, C. (2014). Problem-based learning (PBL). In: R. Gunstone (Ed.), Encyclopedia of Science Education. Springer. https://doi.org/10.1007/978-94-007-6165-0_128-2

Popper, K. (1999). All Life is Problem Solving. Routledge.

Prenzel, M. (2008). What can be learnt from model experiments? Thoughts on the dissemination of results and findings. In: S. Mikelskis-Seifert, U. Ringelband, M. Brückmann (Eds.), *Four Decades of Research in Science Education - from Curriculum Development to Quality Improvement.* Waxmann, 239–252.

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *The Journal of Experimental Education, 70*(4), 293–315.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg Henriksson, H., Hemmo, V. (2007). Science Education Now: A Renewed Pedagogy for the Future. European Commission. http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_en.pdf

ROSE (Relevance in Science Education). (n.d.). http://roseproject.no/. Cf. e.g., Elster, D. (2007). Student interests—the German and Austrian ROSE survey. *Journal of Biological Education, 42*(1), 5-10.

Sjøberg, S., Schreiner, C. (2007). Reaching the minds and hearts of young people: what do we know about their interests, attitudes, values and priorities? What about the interest for space science? International Space Science Institute. http://roseproject.no./network/countries/norway/eng/nor-sjoberg-issi-2007.pdf

Rowe, M. B. (1969). Science, silence, and sanctions. Science and Children, 6(6), 11–13.

Rowe, M. B. (1974). Wait-time and rewards as instructional variables, their influence on language, logic and fate control. Part one. Wait-time. *Journal of Research in Science Teaching*, 11, 81–94.

Ruby, A. (2001). Hands-on science and student achievement. Santa Monica (CA, USA): RAND Cooperation.

Sackett, D.L, Rosernberg, W. M, Gray, J. M., Haynes, R. B. & Richardson, W. S. (1996). Evidence-based medicine: what it is and what it isn't. British Medical Journal, 312, 71–72. https://doi.org/10.1136/bmj.312.7023.71

Sadker, D. (2002). Classroom questions. In: J. W. Guthrie (Ed.), Encyclopedia of Education. Macmillan.

Sauvons l'école. (2012). Ce soir pas de devoirs ! [No homework tonight]. http://sauvons-lecole.over-blog.com/article-ce-soir-pas-de-devoirs-fcpe-102319313.html

Scheid, J., Müller, A., Hettmannsperger, R., & Schnotz, W. (2019). Improving learners' representational coherence ability with experiment-related representational activity tasks. *Physical Review Physics Education Research, 15*(1), 010142. https://doi.org/10.1103/PhysRevPhysEducRes.15.010142

Schroeder, C., Scott, T., Tolson, H., Huang, T., & Lee, Y. (2007). A meta-analysis of national research: effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching,* 44, 1436–1460. https://doi.org/10.1002/tea.20212

Schwartz, D. L., Martin, T., & Pfaffman, J. (2005). How mathematics propels the development of physical knowledge. *Journal of Cognition and Development, 6*(1), 65–88.

Seidel, T. (2003). Lehr-Lernskripts im Unterricht. Freiräume und Einschränkungen für kognitive und motivationale Prozesse beim Lernen – eine Videostudie im Physikunterricht [Teaching and learning scripts in the classsroom. Opportunities and restrictions for cognitive and motivational processes during learning – a video study in physics class]. Waxmann.

Seneca. (ca. 62–64 BCE) Epistulae morales ad Lucilium [Moral Letters to Lucilius].

Seymour, E. (1995). The loss of women from science, mathematics, and engineering undergraduate majors: An explanatory account. Science Education, 79, 437–473.

Shavelson, R. J., & Towne, L. (Eds.). (2002). Scientific Research in Education. National Academy Press.

Shallcross, D. E., & Yates, P. (2014). Skills in Mathematics & Statistics in Chemistry & Tackling Transition. Higher Education Academy.

Shi, Y., Ma, Y., MacLeod, J., & Yang, H. H. (2020). College students' cognitive learning outcomes in flipped classroom instruction: a meta-analysis of the empirical literature. *Journal of Computers in Education, 7*(1), 79–103. https://doi.org/10.1177%2F0735633119881477

Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A reassessment of the effects of inquiry based science curricula of the 60's on student performance. *Journal of Research in Science Teaching, 27*(2), 127–144.

Silver, E. A., & Herbst, P. (2007). Theory in mathematics education scholarship. In: F. K. Lester (Ed.), Second Handbook of Research on Mathematics Teaching and Learning (2nd ed.). National Council of Teachers of Mathematics.

Siskin, L. S. (1991). Departments as different worlds: subject subcultures in secondary schools. *Educational Administration Quarterly*, 17, 134–160. https://doi.org/10.1177%2F0013161X91027002003

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: an analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488. https://doi.org/10.3102%2F00028312033002455

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). Implementing Standards-based Mathematics Instruction: A Casebook for Professional Development. Teachers College Press.

Steinbring, H. (1989). Routine and meaning in the mathematics classroom. For the Learning of Mathematics, 9(1), 24–33.

Steinbring, H. (1991). The theoretical nature of probability in the classroom. In R. Kapadia, & M. Borovcnik (Eds.), Chance Encounters: Probability in education. Kluwer, 135-167.

Stieff, M. (2004). A localized model of spatial cognition in chemistry [Doctoral dissertation, Northwestern University].

Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept and professional goals in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, 100, 255–270.

Sumfleth, E., & Pitton, A. (1998). Sprachliche Kommunikation im Chemieunterricht – Schülervorstellungen und ihre Bedeutung im Unterrichtsalltag [Verbal communication in chemistry teaching – pupil notions and their meaning in daily teaching]. *Zeitschrift für Didaktik de Naturwissenschaften, 4*(2), 4-20.

Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: understanding student interest in school science. *Journal of Research in Science Teaching, 49*(4), 515–537. https://doi.org/10.1002/tea.21010

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction, 16*(2), 165–169. https://doi.org/10.1016/j.learninstruc.2006.02.005

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Cognitive Load Theory: Explorations in the Learning Sciences, Instructional Systems and Performance Technologies. Springer.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). A framework for analyzing cognitive demand and content-practices integration: task analysis guide in science. *Journal of Research in Science Teaching, 52*(5), 659–685.

Tekkumru-Kisa, M., Stein, M. K., & Doyle, W. (2020). Theory and research on tasks revisited: task as a context for students' thinking in the era of ambitious reforms in mathematics and science. *Educational Researcher*, 606-617. https://doi.org/10.3102%2F0013189X20932480

Thomas, J. W. (2000). A Review of Research on Project-based Learning. Autodesk Foundation.

Thoms, L.-J., Collichia, G., Girwidz, R. (2019). Real-life physics: phonocardiography, electrocardiography, and audiometry with a smartphone. *Journal of Physics: Conference Series, 1223*(1). https://doi.org/10.1088/1742-6596/1223/1/012007

Thonhauser, J. (Ed.). (2008). Aufgaben als Katalysatoren von Lernprozessen [Tasks as catalysts of learning processes]. Waxmann.

Thonhauser, I. (2010). Was ist neu an den Aufgaben im aufgabenorientierten Fremdsprachenunterricht? [What is new in tasks in task-oriented foreign language teaching?]. Babylonia, 3/10, 8–16. http://hdl.handle.net/20.500.12162/303

Thorndike, A. (1946). Correlation between physics and mathematics grades. *School Science and Mathematics, 46*(7), 650.

Tobin, K. G. (1987). The role of wait time in higher cognitive level learning. *Review of Educational Research*, 57, 69–95. https://doi.org/10.3102%2F00346543057001069

Topping, K. J., Buchs, C., Duran, D., & Van, K. H. (2017). Effective Peer Learning: From Principles to Practical Implementation. Routledge.

Torigoe, E. T., & Gladding, G. E. (2007). Symbols: Weapons of math destruction. AIP Conference Proceedings, 951, 200. https://doi.org/10.1063/1.2820933

Torigoe, E. T. & Gladding, G. E. (2011). Connecting symbolic difficulties with failure in physics. *American Journal of Physics*, 79, 133. https://doi.org/10.1119/1.3487941

Trautwein, U., Lüdtke, O., Roberts, B. W., Schnyder, I., & Niggli, A. (2009). Different forces, same consequence: conscientiousness and competence beliefs are independent predictors of academic effort and achievement. *Journal of Personality and Social Psychology, 97*(6), 1115–1128. https://doi.org/10.1037/a0017048

Treagust, D. F., Duit, R., & Fischer, H. E. (Eds.). (2017). Multiple Representations in Physics Education. Springer.

Tsaparlis, G. (2005). Non-algorithmic quantitative problem solving in university physical chemistry: a correlation study of the role of selective cognitive factors. *Research in Science & Technological Education, 23*(2), 125–148. https://doi.org/10.1080/02635140500266369

Tsui, C. & Treagust, D. F. (Eds.). (2013). Multiple Representations in Biological Education. Springer.

Uhden, O., Karam, R., Pietrocola, M., & Pospiech, G. (2012). Modelling mathematical reasoning in physics education. *Science & Education*. 21, 485–506. https://doi.org/10.1007/s11191-011-9396-6

Uguroglu, M. E., & Walberg, H. J. (1979). Motivation and achievement: a quantitative synthesis. American Educational Research Journal, 16(4), 375–389. https://doi.org/10.3102%2F00028312016004375

Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2004). Process-oriented worked examples: improving transfer performance through enhanced understanding. *Instructional Science, 32*(1), 83-98. https://doi.org/10.1023/B:TRUC.0000021810.70784.b0

Vanden Bos, G. R. (Ed.). (2015). APA Dictionary of Psychology. American Psychological Association.

Varra, T. Simpson, A., Roesler, B., Nilsson, Z., Ryan, D., Van Erdewyck, M., Schuttlefield Christus, J. D., Sambur, J. B. (2020). A homemade smart phone microscope for single-particle fluorescence microscopy. *Journal of Chemical Education, 97*(2), 471–478. https://doi.org/10.1021/acs.jchemed.9b00670

Vedel, A. (2014). The big five and tertiary academic performance: a systematic review and meta analysis. Personality and Individual Differences, 71, 66–76. https://doi.org/10.1016/j.paid.2014.07.011

Von Korff, J. S., Archibeque, B., Gomez, K. A., Heckendorf, T., McKagan, S. B., Sayre, E. C., Schenk, E., Shepher, C., & Sorell, L. (2016). Secondary analysis of teaching methods in introductory physics: A 50 k-student study. *American Journal of Physics, 84*(12), 969–974. https://doi.org/10.1119/1.4964354

Viennot, L. (1996). Raisonner en physique, la part du sens commun [Thinking in physics, the role of common sense]. De Boeck.

Vosniadou, S. (2013). Reframing the classical approach to conceptual change: preconceptions, misconceptions and synthetic models. In: B. Fraser, K. Tobin, & C. J. McRobbie (Eds.), Second International Handbook of Science Education. Springer, 119-130. https://doi.org/10.1007/978-1-4020-9041-7_10

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Weiss, L., & Müller, A. (2015). The notion of authenticity in the PISA units in physical science: an empirical analysis. *Zeitschrift für Didaktik der Naturwissenschaften, 21*(1), 87–97. https://doi.org/10.1007/s40573-015-0025-9

Wiesner, H. & Colicchia, G. (2010). Improving students' interest – medical and biological contexts in physics education. In: C. P. Constantinou (Ed.), Physics Curriculum Design, Development and Validation (GIREP Conference 2008). University of Cyprus.

Wild, E., Hofer, M. & Pekrun, R. (2001). Psychologie des Lerners [Psychology of the pupil]. In: A. Krapp & B. Weidenmann (Ed.), Pädagogische Psychologie – Ein Lehrbuch (4th ed.). Beltz Psychologie Verlag, 207–270.

Wise, K. C., & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. Journal of Research in Science Teaching, 20(5), 419–435. https://doi.org/10.1002/tea.3660200506

Wiseheart, M., Küpper-Tetzel, C. E., Weston, T., Kim, A. S. N., Kapler, I. V., & Foot-Seymour, V. (2019). Enhancing the quality of student learning using distributed practice. In: Dunlosky, J., Rawson, K. (Eds.), The Cambridge Handbook of Cognition and Education. Cambridge University Press, 550–583.

Wu, H.-K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning, *Science Education*, 88, 465–492. https://doi.org/10.1002/sce.10126

Yin, Y. (2012). Applying scientific principles to resolve student misconceptions. *Science Scope, 35*(8), 48–53.

Yin, Y., Tomita, M. K., & Shavelson, R. J. (2014). Using formal embedded formative assessments aligned with a short-term learning progression to promote conceptual change and achievement in science. *International Journal of Science Education, 36*(4), 531–552. https://doi.org/10.1080/09500693.2013.787556

Zeidner, M. (2005). Test Anxiety: The State of the Art. Kluwer Academic/Plenum Publishers.