Improving learners' representational coherence ability with experiment-related representational activity tasks

Jochen Scheid,[1,*] Andreas Müller,[2] Rosa Hettmannsperger,[3] and Wolfgang Schnotz[4]

[1]*Department of Physics Education, University of Koblenz-Landau, Fortstr. 7, 76829 Landau, Germany*
[2]*Faculty of Science/Physics Section and Institute of Teacher Education, University of Geneva, Pavillon d'Uni Mail, Boulevard du Pont d'Arve 40, 1211 Geneva, Switzerland*
[3]*Faculty of Educational Sciences, Goethe-University Frankfurt am Main, Theodor-Adorno-Platz 6, 60629 Frankfurt, Germany*
[4]*Faculty of Psychology, University of Koblenz-Landau, Fortstr. 7, 76829 Landau, Germany*

(Received 2 October 2018; published 26 June 2019)

Proper understanding of and learning from physics phenomena and experiments requires—among other competencies—flexible and coherent use of multiple representations (MRs). These can include everything from the "enactive" or "operational" manipulation of the experimental devices and materials to the most abstract level of a mathematical formulation of the phenomenon investigated in a given experiment. An essential prerequisite for effective work with MRs is the ability to achieve coherence between different representations. However, research indicates that the level of representational coherence ability of learners across various age groups is low. In order to improve this state of affairs, an intervention study about the use of MRs related to physics experiments was carried out (content area geometrical optics). Specific learning tasks (representational activity tasks, RATs) were designed which explicitly require various types of coherent connections, such as comparing, completing, and correcting representations. In a quasiexperimental repeated measurement study ($N = 302$) using a multilevel analysis for measuring changes, a comparison of a treatment group learning with RATs vs a control group learning with conventional tasks was carried out (with identical content, lesson plans, and duration of the intervention in both groups; moreover, each of the four schools had corresponding classes of both groups. They were taught by the same teacher). Results showed a highly significant and practically relevant effect on students' representational coherence ability ($p < 0.001$; $d = 0.69$). The positive effect of RATs could still be found six weeks after the end of the intervention ($p < 0.001$; $d = 0.43$). Several covariates (gender, pre-instructional knowledge in physics, mathematics, three facets of intelligence) were analyzed, with no or small influence on these effects. Finally, some limitations and implications of the study for classroom practice and further research are discussed.

DOI: 10.1103/PhysRevPhysEducRes.15.010142

I. INTRODUCTION

Representations are entities or objects that stand for something else [1]. The object being represented and the representation of it have to be connected in a meaningful way ("representational connection" [2,3]). In science and science education, the use of different representational formats, often in multiple, interconnected forms (multiple representations, MRs), is well known as an essential means of domain-specific reasoning. An example in physics is the verbal description of a geometrical optics experiment, a photograph of it, a schematic description through ray diagrams, and a formal description by the magnification equation. All these representations are necessary to achieve a proper understanding of an image formation process. There is ample evidence supporting this important role of MRs across all branches of science education, in general (e.g., Ref. [4]), in biology, chemistry, physics (e.g., Refs. [5–7]), earth sciences [8], and also mathematics (e.g., Ref. [9]). This central role is also emphasized by an extended strand of research from the cognitive sciences (e.g., Refs. [10,11]).

In particular, MRs are a salient feature of scientific reasoning (and understanding) that imply the ability to build meaningful connections between different representations; they are called "referential connections" (see e.g., Ref. [3]), or to establish "representational coherence" [12]. As a result, it is not sufficient that learners are able to handle only one type of representation at a time [13,14]. Many studies in science and mathematics education have

*Corresponding author.
scheid@uni-landau.de

emphasized the essential importance of representational coherence both across scientific fields as example physics (e.g., Refs. [15–19]), other sciences (e.g., Refs. [20–22], mathematics [23], various aspects of scientific learning and cognitive processes, such as reasoning (e.g., Refs. [11,16], conceptual understanding (e.g., Refs. [24–26]), problem solving (e.g., Refs. [11,18,27–30], and creativity [31]).

Moreover, the ability of making coherent connections *within* a representational format is also of importance; they are called "*intrarepresentational connections*" [12,32,33]. Mayer [33,34] points out that these connections and their coherence are prerequisites for meaningful learning. Examples of intrarepresentational connections in geometrical optics are connections between different parts or different versions of a ray diagram.

On the one hand, one can thus certainly state an essential role of multiple representations and their various connections for physics or science learning, or as Kohl *et al.* [35] put it, "good use of multiple representations is considered key to learning physics." On the other hand, there is considerable evidence that learners have marked difficulties in using and linking several representational formats simultaneously. Thus, the ability of establishing coherence between multiple representations (students' representational coherence ability) is something to be explicitly taught and learned, beyond a given disciplinary content (or rather as a part of it). It is also important to note that this finding holds not just for school-aged learners (see e.g., Refs. [36,37]), but also for advanced university students up to their 5th and 7th semester [38].

The present contribution is specifically about the role of multiple representations as related to understanding of and learning from experiments. This is relevant both for recent research in physics and for learning physics in school. For example, a physics group at CERN [39] writes about the observation of the Higgs boson and uses verbal descriptions of the experiment, a schematic notation for reaction channels, several tables (apparatus and observation data), and diagrams (events vs energy), in particular showing the maximum at the Higgs mass, and various equations of experimental or theoretical significance. Thus, the student trying to understand (or carry out herself) an experiment in school and the reader of the article about a cutting edge experiment in physics have something in common—the necessity to use and connect a variety of various representations, with complementary content ranging from information about essential experimental features to formulation and interpretation of the findings in the theoretical framework underlying the experiment.

Can we expect from learners a coherent understanding of an experiment linked to a number of representations (text, graphs, equations, etc.), if we know that they have considerable difficulties to establish coherence of MRs in general (from school to university)? Or should we rather suppose that coherent understanding of MRs is one of the obstacles which leads to the often unsatisfactory learning effects of experiments [40–42]?

The purpose of the present article is to contribute to and answer these questions. More specifically, we present an empirical intervention study exploring the effect of theory-based tasks with a representational focus related to science experiments on students' competence of using multiple representations in the area of geometrical optics.

II. EMPIRICAL AND THEORETICAL RESEARCH BACKGROUND

A. Multiple representations and representational competence in science and science education

The ability to generate and use various specific multiple representations of a subject or problem as a problem-solving tool in a conscious, skilled, and interconnected way is called *representational competence* [36,43–45]. It includes the ability to "translate" between different forms of representations [46] and to communicate underlying, not obviously perceived, physics-related entities and processes [20,44,47]. In the following, we present the research background and theoretical basis of our study, as well as defining a number of terms that are not entirely consistent in the literature.

We follow a theoretical framework commonly used in cognitive psychology [14,48,49] and science education [50,51] which combines a referent or object, its representation, the interpretation of the latter, as well as the interactions between each of these. This tripartite relation stems from a long tradition of thought from Greek philosophy (see Ref. [52] for an overview) through to 20th century semiotics [53,54]. It bears several other names, including "triangle of reference" [54], "semiotic triangle" [50], "Peircean triangle" [51], "Ogden/Richards triangle" [48], etc. While this idea is an important background concept, it would be beyond the scope of this article to discuss it and its many applications and ramifications in various disciplines (see references mentioned above and, e.g., Ref. [55]). Here, the focus is on its profound significance for (multiple) representations in science education, and we refer to it as "triangle of meaning" [4,51,56]. We agree very much with these authors regarding their choice of "meaning" as a key term in an educational context. Not only does it encompass both cognitive and value aspects of learning, but it also links the topic to "meaningful learning," another influential line of thought in educational science [57–59] and science education [60–62].

Within this general framework, there are several conceptual and typological distinctions of (multiple) representations. First, representations can be either external to the mind (material object, such as a text) or internal (mental state; i.e., the same text as retained in working or long-term memory) [14]. Cognitive processes, such as understanding or learning, often imply an interaction of the former with the latter. However, they can also be based on internal (mental) representations only, without an external one being present (the reverse is of course impossible by definition). The intervention under investigation in the present article consists of a specific form of learning task on a given set of external representations, designed to improve the adequacy of learners' internal mental representations, and thus also their understanding.

A second important distinction is between two basic forms of representations: descriptive (language based or symbol based) and depictive (picture based). Depictive and descriptive representations are differentiated by the kind of signs they use to represent some feature of an object, which are of iconic and symbolic nature, respectively. Iconic signs "are associated with their referent by similarity or by another structural commonality," whereas symbolic signs "have no similarity [or structural commonality] with their referent" [14].

The most familiar forms of descriptive representations are verbal ones, whether written or oral. One example is a description of a scientific observation or experiment, whereas a photograph (or schematic drawing) of it is a familiar example of a depictive representation. However, there are other important forms of representations in the context of science: descriptive ones such as numbers (measured or calculated values of physical quantities) and formulas (relating these quantities on the basis of some theory), and depictive ones such as graphs (also relating quantities) or diagrams (such as electrical circuit diagrams or ray diagrams). Depictive representations can include any of the following [62]: (a) realistic pictures visually similar to what they represent (photographs, naturalistic, or realistic drawings), (b) schematic pictures bearing a visual resemblance to some aspects of interest of the object, while abstracting from all others, (c) conventional code systems (e.g., maps, engineering drawings, or blueprints), (d) logical pictures serving the visualization of some abstract structural properties (e.g., vector diagrams, function and bar graphs, tree and network graphs; see also Ref. [48]). It is worth noting that these are prototypical forms of depictive representations, which are located on a continuum of abstractness. Other examples exist of intermediate forms of these prototypes (e.g., cartoons and sketches, which are intermediate to realistic and schematic pictures). Moreover, mixtures of these prototypes can be found in certain representations. For instance, ray diagrams contain elements of schematic pictures (lenses), and logical pictures (rays, which are abstract and idealized representations of light bundles and their direction of propagation). It is useful to distinguish these representational formats, as they lead to different kinds and levels of comprehension (among others, in physics education) and associated difficulties. This has been emphasized by Leisen [63] from the practitioner's point of view, and in a theoretical account by Vosniadou [64]. The latter shows how (a) "specific scientific and mathematical domain knowledge" and (b) "substantial epistemological sophistication" lead to schematic or logical pictures being understood differently and with greater difficulty than realistic ones.

Each representational format requires a specific way of thinking and leads to a specific form of comprehension of the subject in question. When coherently combined, this multiplicity of formats leads to improved understanding [63,65]. This is also formulated more generally in the "multimedia principle" of Mayer [13]: "people learn more deeply from words and pictures than from words alone.". Mayer's research belongs to the group of dual coding approaches, together with the models of Paivio [2,14]. Within these approaches, we use the integrated model of text and picture comprehension [14,66], which allows for a concise description of the role of MRs particularly in science learning, including the approach studied in here. Using the terminology introduced above, it posits that the cognitive system contains a depictive (pictorial) and a descriptive (verbal) system with different memories, limited storage, and processing abilities [14,66]. Two separate sensory registers provide the input system for auditory and visual information, respectively. On this perceptual level, the information is stored in auditory and visual working memory. On the cognitive level, a verbal channel processes information from texts and a pictorial channel processes information from pictures. Beyond the examples for depictions given above, another concept which is critical to MRs as a reasoning tool is that of *mental (or internal) models*, in the sense "that the mind constructs models of the world that it uses to reason" [67]. The elements of mental models may come from multiple sources, viz. perception, comprehension of discourse knowledge, and imagination [67]; in particular, mental models are not sense specific and in most cases integrate information from multiple representational formats [14]. Just as with integrated formats, mental models also serve multiple purposes. These include the representation and processing of spatiotemporal, causal, structural, and other types of informational elements and the relationships between them. A crucial feature of mental models is that there is a correspondence between their structure and the structure of what they represent. As Johnson-Laird [67,68] puts it, "individuals are represented by individual tokens, properties by properties of these tokens, and relations by relations among these tokens." In other words, the construction and use of mental models are based on a kind of "structure mapping" (in other words, analogous relations) between representations and their referent [12,49,69]. There is some controversy as to whether mental models are necessarily depictive, or whether they can be of a more general nature. On the one hand, as they are "associated with their referent by similarity or by another structural commonality," they use iconic signs and are depictions by definition [14,66].

However, other authors [13] allow for internal models of a descriptive nature ("verbal models," in their terminology). Indeed, in physics, for example, a differential equation of some dynamic process (such as of a harmonic oscillator) is called a model (a very powerful one), despite its being a representation using purely symbolic signs and thus descriptive in nature. In any case, the mental models used in the present research (ray diagrams of geometric optics) are obviously depictions, and we will not attempt to resolve the general controversy here.

The representations in the descriptive (verbal) channel are called "propositional representations" in the framework of Schnotz [14,66]; by definition, they use symbols to describe their referenced object. Information from mental models can be read off, translated, and connected to propositional representations and vice versa, i.e., referential connections can be established (as opposed to intrarepresentational connections, see above). In this way, a process of coherence formation can take place and lead to a more-or-less coherent set of these mental representations; it is the purpose of this contribution to study a way to support this coherence formation.

Tasks with multiple representations may lead to *cognitive overload*, in particular in cases of high-element interactivity of information [70–72]. A student's intelligence and expertise influence her level of cognitive load. This means these parameters may influence whether the cognitive load for a student is appropriate, too high, or too low (see "aptitude treatment interaction" [73]; "expertise reversal effect" [74]; for an explanation of these effects, see also Ref. [14]). Consequently, learning tasks for the development of representational competence should be relatively easy, especially in an early stage of learning. With growing practice, learners acquire more routine in working with MRs, which reduces the cognitive load.

Complementary to this background in cognitive psychology, Ainsworth [46] proposed a taxonomy for the educational functions of MRs, and later a still more encompassing theoretical framework concerning educational design, functions, and tasks of MRs [design, functions, tasks (DeFT) [75,76]]. For the study presented here, the following aspects from this theory are of central importance. Regarding *functions* of MRs, Ainsworth [46] and subsequent work inspired by her approach (e.g., Refs. [12,77,78]) emphasizes the essential role of integrating information from MRs to construct understanding of a domain's key concepts and the relationships between them (the constructing function). This, of course, is completely in line with previous research about referential connections [3], information "integration" [79,80], "structure mapping" [81], the integrated model of text and picture comprehension [14,66], and especially the large body of literature regarding their significance for science education already cited above. The function "construct deeper understanding" of the DeFT model is specifically relevant in the present context. It deals with integration of information from MRs "to achieve insight that would be difficult to achieve with only a single representation" [75]. It is one purpose of this study to support this function by a specific type of "representational activity tasks (RATs)", the details of which will be presented in Sec. III C.

As for the *task* dimension in the DeFT model, the following are central for the present context [75,76]. First, the tasks to "understand the form of representations" and "how to construct an appropriate representation," used here as the understanding and appropriate use of ray diagrams, with their specific learning difficulties [82]. Second, the task of "understanding how to relate representations," which is a core element of the instructional approach of RATs presented here. Third, that "learners should understand the relation between the representation and the domain" [75,76]; this is the representation-referent connection shown in Fig. 1. The latter represents an essential and difficult step [6,41,83] when dealing with abstract scientific concepts and (multiple) representations of them that are related to experiments and observations. Finally, the design features in service of these functions and tasks (the last dimension of the DeFT framework) will be discussed in detail below (Sec. III C).

Is representational competence, in terms of cognitive psychology, part of the "procedural knowledge" of a domain, of the "knowing how" in the form of domain-specific abilities, techniques, and methods including their rules of application [84,85]? Only very few contributions explicitly treat the link of representational competence and procedural knowledge (in fact, we are only aware of work in mathematics education, e.g., Ref. [86]). However, the idea of multiple representations as powerful cognitive tools is common in the literature on MR [7,76]). The link

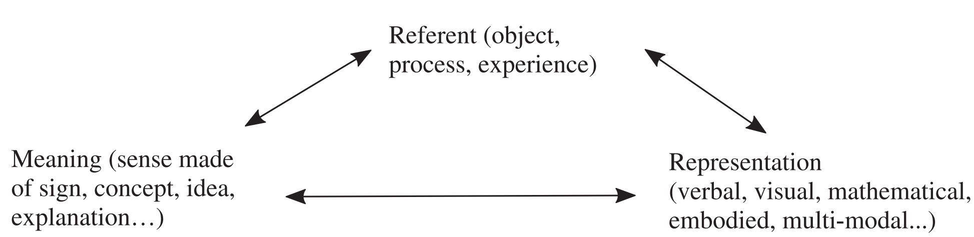


FIG. 1. Tripartite relation of referent (object), representation, and meaning (interpretation), or "triangle of meaning"; see text (according to Refs. [4,56]).

to procedural knowledge is thus almost self-evident: using tools implies "knowing how" to use them, and knowing how to use multiple representations is nothing but representational competence. In this sense, representational competence is indeed part of procedural knowledge of a given domain, for example, how to read a given form of diagram [87].

B. Representational coherence ability and science learning

An appropriate level of understanding of phenomena, experiments, and other learning content generally requires a certain level of representational competence. The ability to connect and translate between content of different representation types is also particularly important. Translating information between several representation types is inherently susceptible to misinterpretation or failure, which can lead to unnecessary contradictions and inconsistencies. Therefore, an important part of representational competence depends on a students' ability to achieve consistency across the overlapping information of a set of representations. This part is called a learner's "representational coherence ability" (RCA) (see Ref. [88]). In this contribution, we focus on RCA because it is essential for the use of multiple representations, has a fundamental connection to achievement in a given subject matter, and there is little research about how to improve it so far.

Learners show RCA to varying degrees. Experts perform significantly better than novices when translating the content of a graph, a video, or an animation about molecules into any other type of representation [20]. Learners with low representational abilities often work on the surface level of a representation [20,89,90], whereas those with high representational abilities show features of deep-level processing, such as using a higher number of formal and informal representations for problem solving, or producing more predictions and explanations about the phenomenon in question [20,91–94].

Several studies have shown that RCA of students usually is low, even for older age groups. This has been identified, for example, in chemistry [95]. Secondary school students (average age: 18 years) had to solve several tasks in their high school examination. To do this, they had to be able to connect macroscopic, submicroscopic, and symbolic levels of chemical concepts [95,96]. Students were not able to make sufficient connections. As a result, their knowledge is fragmented and often only remembered temporarily. A possible reason for this can be seen in a lack of lessons that teach students how to establish such connections. Furthermore, students not only lack the ability to interconnect different representation levels, they also have problems within the same representation level and do not clearly see the connections between the submicroscopic level and its diagrams [97]. In a study on representational coherence in mechanics ($N = 168$, 16-year-old students; [37]), it was found that only 11% of the learners produced coherent and scientifically correct representations after a teaching sequence. Even at the university level, students have been found failing to connect the meanings of formulas to phenomena and experiments (with no noticeable difference between 7th and 5th semester students [38]). First, problems occurred when students tried to explain an experiment using only one type of representation (e.g., with the symmetric form of the Coulomb law for electric point charges). Second, students were only able to give the right order of magnitude for the measurement value if they made a connection with the phenomenological level (e.g., for the case of a charge distribution, where Coulomb's law is not applicable). Third, when students were not very familiar with the topic of a representation they could not use it to solve the task, even if they had already worked on the content in the task directly before [38]. This is a typical transfer problem. One potential explanation is that students operate only on the surface-level features of representations, and fail to grasp the common underlying structure [20,89,90].

In spite of these well-documented problems, teachers do not usually focus on teaching students how to connect several representations [95,96]. Thus, the way in which teachers deal with representations in classes may be an explaining for the low level of students' RCA. A study looking at understanding of experiments (Ref. [98] $N = 344$, seventh to ninth grade, topic: mechanics and electricity) revealed that students had few opportunities to connect different types of problem-relevant representations in greater depth. Lee [99] analyzed 47 lessons in three eighth-grade classes on ray optics (age 14, secondary level I). The findings were, on the one hand, that the representations used in the classroom were in part inaccurate. On the other hand, the sources of the representations were most frequently the teachers themselves (51%), followed by the textbook (26%). Students' self-generated representations were rather rare (<10%). Moreover, the time spent on a representation was less than 3 min on average. This means that an implicit, short, and receptive way of using representations prevails in the classroom. Accordingly, the students are not taught explicitly how to coherently process representations. A similar picture emerged from an analysis of more than 800 tasks in secondary-level physics textbooks [100]: in the vast majority of cases, the number of representational formats (N_{RF}) that is needed to solve a task is low [$\bar{N}_{\mathrm{RF}} = 1.65$ (0.53) for the mean and standard deviation, respectively]. Consequently, that for the number of connections between representational formats (N_{RFC}) is low, too [$\bar{N}_{\mathrm{RFC}} = 0.66$ (0.54)]. Based on a considerable body of evidence (e.g., Refs. [91,93,94]), Kozma [47] concluded that explicit teaching of representations—in particular in the sense of fostering RCA—should be included in the (chemistry) school curriculum (as is the case in Denmark [44]).

Consistent with these findings in science education, instructional and educational psychology has also identified a lack of deliberate and systematic cognitive activation for the use of MRs as a major source of the difficulties described above (Ref. [11] Chap. 6; Ref. [101]). To counter this, a "representational focus" [4,25,102] has been proposed in science (physics) education, in the sense of actively and explicitly engaging learners with multiple representations (Ref. [5] Chap. 7; Ref. [56]; Ref. [103] Chap. 16). In this context, emphasis is placed on the need for learning activities going beyond simple cases and using more complex forms of MRs, in particular with a higher number of representational formats and connections (Ref. [5] Chap. 1; Ref. [103] Chap. 15; Ref. [104]). An important example of this is a text-image combination referring to an experiment or observation, which is a common format in science (number of representational formats >2, total number of representation connections >1). Work for explicit MR learning activities and approaches in this sense was presented by Tytler *et al.* (Ref. [4] Chap. 1: biology and Chaps. 7, 8: astronomy) and Hubber and Tytler [56] (astronomy), as well as by van Heuvelen and Zou (Ref. [104] physics). Kohl and Finkelstein [28] have found positive learning effects in a research-informed large-enrollment course of introductory physics, including, among others, deliberate variation of MRs used. However, studies about the improvement of representational competence in a well-controlled setting in science education are scarce, and none of the few intervention studies we are aware of (see references above) explicitly addressed representational coherence in the context of experiments. A study of this kind in the area of geometrical (or ray) optics is the purpose of the present contribution.

C. Learner characteristics

Previous research has shown that learning with multiple representations also depends on a series of learner characteristics (see Ref. [105] for a recent review). Indeed, prior knowledge in a domain is a strong predictor of learning in general (Cohen $d = 0.67$), and in science in particular ($d = 0.8$, Ref. [106]). Regarding the use of MRs, prior knowledge can influence how well MRs can be linked to the referent (representational connection), reduce cognitive load, and complete information not present in the available representations [75,107]. Cook [108] gives a thorough account about the effects of prior knowledge on science learning with MRs. This is essentially based on cognitive load theory, i.e., the ease with which various representations can be processed simultaneously in working memory is largely determined by the prior knowledge a learner has, and the more prone he or she is to cognitive overload. Similar arguments can be found across science education (Refs. [27,29,109,110] for physics, chemistry, and biology education, respectively). Further effects of prior knowledge specifically related to multiple representations in the physical sciences have been discussed, for example, by Bodemer and Faust [111]; [6] Chap. 7; and Ref. [105]. We thus also included prior knowledge in mathematics as a covariate in our analysis, both for its general importance [112] and for providing essential representational formats for physics and physics learning. Strong correlations between mathematics and physics or physical science achievement have been found for decades, from classical work (Ref. [113] $r = 0.77$) through meta-analysis (Ref. [114] $r = 0.48$) to recent analysis for introductory physics courses (Ref. [115] $r = 0.3$–0.46). The effect sizes corresponding to these studies (for conversion, see Ref. [116]) are medium to large throughout ($d{:}\,2.4, 1.1$, 0.6–1.0, respectively). In the last decade, more fine-grained studies have confirmed these findings, and interpreted them with increasing detail [117–122]. Moreover, as text is an all-pervading representational format in physics tasks (and physics learning in general), we included also German language grade as a covariate.

Another important factor for learning with MRs is visuospatial ability [46,123], especially for science learning (Ref. [5], Chap. 1, Chap. 11; Ref. [124]). In the physical sciences, such influences have been discussed by several authors (Ref. [18]; Ref. [6] Chaps. 7, 8, 11) and Wu and Shah [125] provide ample evidence in their review of correlational studies and other sources (albeit without reporting the correlation coefficients). However, this literature rarely provides quantitative results on the strength of the presumed association between visuospatial ability and science learning (see, e.g., Ref. [125] as just mentioned). Among the few exceptions found are studies on chemistry learning [126,127] with an effect size range ≈ 0.4–0.6 (conversion: Ref. [116]). Recently, Opfermann *et al.* [105] emphasized that spatial ability might be particularly important in an abstract domain such as physics. We follow their recommendation to take this "into account whenever research on physics learning includes (at least partly) visual multiple representations" and include it as a covariate.

Beyond visuospatial ability, two further aspects of cognitive ability related to different representational formats are considered (verbal and numerical intelligence). Finally, we consider gender as a further covariate. Gender differences for physical science and physics have also been known for decades (meta-analytic results: $d = 0.35$, [128], $d = 0.25$, [129]; large scale sample ($N \approx 8000$): $d = 0.32$, [130]). More recently a "gender gap" was discussed for physics introductory courses ($d = 0.38$ for conceptual understanding; $d \lesssim 0.2$ for course grades [131,132]).

D. Purpose and research questions

The empirical intervention study presented here aims at exploring the effect of theory-based tasks with a representational focus related to science experiments (RATs, for example; see Figs. 2 and 3 on students' RCA in the area of

Consider, in what relevant physics features do the arrangements of optical elements differ?
Hint: the sizes of the convex lens are not important.
Mark the differences in the schematic drawing. Adapt the schematic drawing so that it matches best with the realistic picture.
Describe and justfy the modifications you have made.

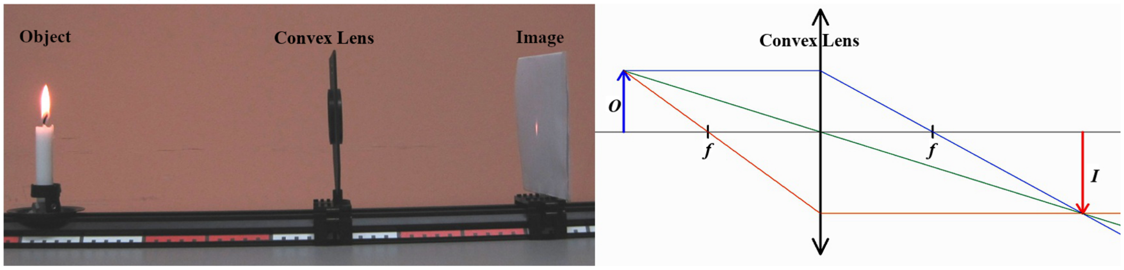
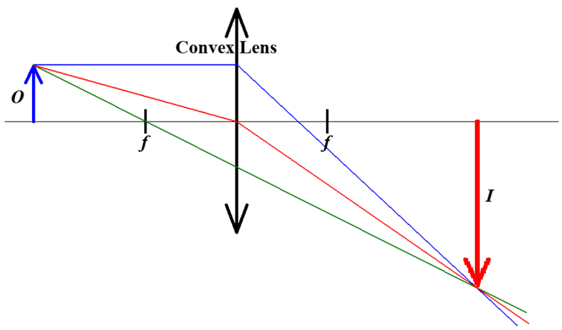


FIG. 2. RAT for fostering students' ability to connect several types of representations (O = object size; f = focus, I = image size).



Mark the failures in the ray diagram.
Correct the ray diagram.
Explain what was wrong with the rays.

____

____

____

____

FIG. 3. RAT for fostering students' ability to check the scientific correctness of a representation (correction exercise).

geometrical optics). The purpose is a contribution for the research need described above, i.e., a study with a well-specified intervention and instrument for RCA, which are described in detail in the methods Secs. III B and III C. As learning with multiple representations increases demand on learners, we also consider whether the intervention has (side-) effects on motivation. In this framework, our research questions were as follows:

(1.1) What (if any) is the effect of representational activity tasks students' representational coherence ability compared to conventional tasks?

(1.2) If there is an effect, does it also occur for the "construct deeper understanding" function in the sense of the DeFT framework?

(2) To what extent do various covariates (previous knowledge, various components of intelligence, etc.) influence learners' representational coherence ability?

(3) What is the effect of representational activity tasks on motivation compared to conventional tasks?

III. MATERIALS AND METHODS

A. Sample and study setting

The investigation took place within regular secondary level I physics lessons in the German state Rheinland-Pfalz. The age group was 7th and 8th grades in the German school system, from four academic-track schools[1] and six class groups in both the treatment (TG) and control group (CG) [TG: age $M = 13.1$ (0.72) years, $N = 175$; 70 boys, 105 girls; CG: age $M = 13.1$ (0.63) years, $N = 167$; 80 boys, 85 girls, two missings for gender].

The subject matter was geometrical optics (light sources, light propagation and rays, shadows, lenses, image formation), a standard topic according to the pertinent teaching program of this age group. The length of the intervention was about six lessons ($6 \times 45' = 4.5$ h in total). Each teacher instructed one class of the treatment group and one class of the control group. An overview for the schedule of the interventions is given in Table I.

The intervention and the learning materials were discussed and validated within an expert group of physics teachers, including those participating in the study. Every teacher received a schedule and the final version of the learning materials for every class hour of the intervention and was briefed individually on how to conduct the lessons for both the CG and the TG. An observer participated in several lessons and checked whether both groups corresponded to the intended intervention strategy.

[1]"Gymnasium", see Ref. [133] for background about the German school system.

TABLE I. Schedule of the quasiexperimental intervention study [88].

Week		CG		TG
1			Initial test (t1): RCA, motivation, covariates	
			Introduction of the convex lens, teacher experiment	
2–4	Learning phase	Conventional tasks	Students' experiment, refraction by lenses, construction of ray diagrams, geometric proof of imaging properties, image formation	Representational activity tasks
4			Post-test (t2): RCA, motivation	
5–10			Conventional lessons on optics (12 class hours)	
10			Follow-up test (t3): RCA, motivation	

B. Research design, instruments, and data analysis

The intervention study had a quasiexperimental non-equivalent comparison-group design with repeated measures and covariates [134] There were three measurement times: "t1: pre" before the treatment, "t2: post" after treatment and before conventional lessons, and "t3: follow-up" after conventional lessons. There were four weeks between pre- and post-test and six weeks between post-test and follow-up (see Table I).

The control group worked with conventional tasks, which deal with connections between representations only implicitly, with a focus on content and a problem statement related to it (such as finding the optical image in a given lens arrangement), and based on the tacit assumption that the pertinent representational means to express this content and problem, and their connections (such as ray diagrams, and relating them to the experimental situation), will be used by the learner without explicitly asking for this. In contrast, the treatment group learned with specific tasks focusing on representational coherence, based on the research expounded above. These tasks contain always more than one representation, often more than two representational connections, and explicitly ask students to elaborate on various types of coherent connections, such as translating, relating, comparing, completing, correcting and changing, or adapting representations. The design of the TG learning tasks is described in detail in Sec. III C. Beyond these differences, TG and CG were identical in their content, lesson plans, and duration of the learning sequence. Moreover, in each of the four schools, pairs of TG and CG classes were taught by the same teacher.

A comment on the quasiexperimental approach of the study is in order. As is well known, quasiexperimental research does not eliminate the problem of confounding variables and ensuing problems for internal validity. However, Ref. [135] presents research strategies dealing with these problems using specific research designs, in particular through repeated measures and covariates. Associated with that, a variety of statistical techniques (e.g., analysis of covariance and more advanced methods) can be used to try to cope with threats to internal validity, particularly useful when random assignment is not possible, practical, or adequate [136,137]. On this background, special care was used for discussion and analysis of pre-test values and covariates, in order to minimize threats for internal validity.

In order to look for potential influences of the intervention on physics motivation, a physics motivation test taken from well-validated instruments in the literature was used. [138–140]. The test assessed extrinsic [139] and intrinsic motivation as well as physics self-concept [140] as aspects of motivation together as one motivation measure. Measurement times t2 and t3 are influenced by the intervention, measurement time t1 gives the initial motivation (used for comparison with the other measurement points, and as covariate).

Based on the research background presented in Sec. II C. covariates are as follows:

- motivation (at t1)
- gender
- relevant school grades (physics, mathematics, and German language), and
- three subscales of cognitive ability: related to different representational formats taken from a standardized published instrument (verbal, numerical, and matrix reasoning as part of visuospatial intelligence [141,142]; reliabilities for all scales are satisfactory ($\alpha_C > 0.7$; [143]).
- conceptual understanding (at t1): assessed with a short concept test specific for the learning topic (ray optics, image formation; $\alpha_C = 0.78$ [144]).

For an assessment of the representational coherence ability of learners, a test for RCA was used [145,146]. RCA test items required relating physical phenomena and experiments to various types of representations and multiple representational formats to each other. Establishing representational links required comparing MRs, as well completing and correcting given incoherent MRs. Several items also asked students to explain their reasoning while resolving these questions. The RCA test contained 14 experiment- or phenomenon-related tasks, with five standard items ("S") and nine "deeper understanding" items ("DU"), see Sec. III C for this distinction (max. score of S items is 22.5 and of DU items is 21). Sample items and an

TABLE II. Instrument characteristics. Averages and standard deviations (in parentheses) of item difficulty, item discrimination, item-test correlation, internal consistency and factor loadings, as well as the recommended ranges (Ref. [147]; factor loadings see Ref. [148]).

Variable	RCA post-test	Recommended
Item difficulty $\bar{P}$	0.37 (0.17)	0.2–0.8
Item discrimination $\bar{D}$	0.54 (0.25)	≥0.3
Item-test correlation $\overline{r_{it}}$	0.46 (0.09)	≥0.3
Internal consistency $\overline{\alpha_C}$	0.79	≥0.7
Factor loadings $\overline{F_L}$	0.57 (0.10)	≥0.4

overview (Table VII) about the different combinations of representation types and their connections used in the RCA instrument are given in Appendix A. The RCA data were tested for deviations from normality, with no significant result (Kolmogorov-Smirnov test: $p = 0.78$). Test characteristics in terms of classical item and instrument analysis are given in Table II: item difficulty, item discrimination, item-test correlation [147], and α_C as a measure of internal consistency [73]. Overall internal consistency was $\alpha_C \approx 0.8$ (across different validation samples), testing for exclusion of the individual items did not lead to an improvement. Item difficulties were $0.2 < p < 0.8$, item discrimination $D > 0.5$, and item-test correlation was $r_{it} > 0.3$. All values were in the recommended range (which holds as well for all individual item characteristics, up to a few slight exceptions, [147]). A detailed description of the design and validation of the instrument is given by Refs. [145,146].

Additionally, an expert rating for the curricular validity ("the item content is conform with the curriculum") and for appropriateness as a physics test question ("the item as appropriate for a physics performance test") was carried out in the expert teacher group mentioned above (11 experienced physics teachers with an average of 21 years of teaching experience, see Refs. [100,146]). A 6-point Likert scale (1 = completely disagree, 6 = completely agree) was used. Results yielded a satisfactory rating for curricular validity and test appropriateness ["completely agree" or "agree" (6 and 5 on the Likert scale) for 10 and 11 items, respectively; "rather agree" (4) for 3 and 2 items, respectively]. One item required to derive the magnification equation in a geometric way. This is not contained in the local curriculum, leading to a low expert rating of curricular conformity. Nevertheless, it makes sense to keep this item in order to assess RCA. Interrater consistency is calculated as the intraclass correlation coefficient, which is suitable for more than two raters and equivalent to Cohen's kappa [149]). The values obtained are 0.52 for curricular conformity and 0.61 for appropriateness as a classroom test question [100], which, according to current guidelines, can be considered as satisfactory and good, respectively [146]).

To estimate the necessary sample size for the intervention study, a power analysis was carried out (with values $\alpha = 0.05$ and $1 - \beta = 0.8$ chosen according to Cohen [150]). The analysis was carried out with G ∗ Power [151], assuming a multilevel model for measurement of change (i.e., a hierarchical linear multiple regression [152]) with one assessed predictor and eight predictors in total, and an effect size $f^2 = 0.03$ (as a minimum requirement of a practical relevant effect), yielding a minimum sample size of $N \approx 270$ (present study: $N = 342$).

For RCA as a dependent variable, a multilevel analysis with two levels adapted for the measurement of change was used (level 1: measurement times, level 2: subjects [152–154], calculation performed by SPSS [155]). The advantages of multilevel models for longitudinal data compared to a usual repeated measurement analysis of variance are that they are less restrictive in their applicability assumptions, and more flexible in the data structure they can model [153,154]; [156] Chap. 5: (i) incomplete data sets (missings) can be dealt with; (ii) change is allowed to vary across subjects (in turn to be modeled by covariates); (iii) measurement points need not be equidistant, and change is allowed to vary between them (e.g., by piecewise regression); (iv) variances (and covariances) are allowed to change between measurement points. Note, that (iii) and (iv) are particularly relevant in a case of a pre-, post-, or follow-up measurement (as in this study) as changes and (co-)variances cannot be assumed to be equal for the different measurement times.

Finally, the effect size (d) to be used for this method of analysis has the general definition "difference between the means for the treatment and control groups divided by the standard deviation" [150]. In our case, the standard deviations of the compared groups were not equal and we used the pooled standard deviation. The generalization of d to multilevel analyses is given in a very readable account by Tymms [157] (see also Ref. [158]). It is used in the present paper with the usual conventional thresholds (small, medium, and large effects are $0.2 < d < 0.5$, $0.5 \leq d < 0.8$, and $0.8 \leq d$, respectively [150]).

C. Instructional design of the intervention

1. Operationalization of representational activity tasks

In line with the theory background given in Sec. II, the operationalization of RATs as learning tasks to foster representational coherence ability is based on the ability to build coherent, correct, and meaningful connections between pieces of information within one single or between several representations (intrarepresentational and referential connections). Thus, RATs can only be solved by working with intrarepresentational and referential connections (and also, as a sense giving base, representational connections to represented objects). They require students to explicitly translate, relate, map, compare, correct, and change or adapt pieces of information within and between representational formats. As descriptive representation forms, text and formulas are used, and as depictive representation forms photographs, schematic drawings, and ray diagrams (see examples below). With regard to the curricular validity of the local study program [159], only few formulas are used (five out of 29 tasks).

In terms of the DeFT design features (see Refs. [75,76] and Sec. II), RATs are designed as follows:

(a) ***Number of representations:*** RATs contain two or more different representations and one or more relations. The relations can be built between the same representational format (intrarepresentational connection) and between different representational formats (referential connection). For the RATs of this intervention, the averages are $\bar{N}_{\mathrm{RF}} = 3.6$ and $\bar{N}_{\mathrm{RFC}} = 3.1$.

(b) ***Distribution of information***: Representations complement each other with regard to the contained information (they may also partially contain redundant pieces of information).

(c) ***Format***: The representational system is monomodal (viewing) and static (on paper).

(d) ***Sequence of representations***: As RATs are supposed to foster flexible use of multiple representations as applied to various physics problems, there is no fixed sequence in which representations occur in the task. A general pattern, however, is as follows: First, the text of the task referring sometimes to an additional representation shown in the task. Second, the different representations and connections are necessary to achieve the correct solution of the task. In some cases, the sequence leading to a solution is indicated by the task. In other cases (a special case in the sense of DeFT) a possible sequence of representations has to be found by the learners themselves. Third, the answer contains usually one of the representations used in the process of the task solution, sometimes additionally a verbal description or an explanation of the answer.

(e) ***Support for translation***: This has a special status for RATs, as the translations (or connections) are asked for in the tasks, and are not given by them. A "support" on a general level would occur if repeated solving of RATs as learning tasks leads to an increase of RCA in the sense of the main research question of this contribution.

Another attribute of RATs is the type of connection that has to be made for solving the task: The first type of RATs requires only to relate, translate, or complete information that is contained in representations as is common in some of the conventional tasks in physics education. As mentioned in the theory Sec. II, RCA is an important component of physics understanding in general. This means that there are domain-specific physics questions and problems that cannot be answered without implicitly using MRs and RCA. The kind of connections in these standard tasks usually only require relating pieces of information of two representational formats or translating them from one to another [100]. Thus, this kind of representational tasks close to conventional ones only implies a low number of representations and of their connections close to the lower limits (2 and 1, respectively). This type of RATs are called standard (S) RATs, because their design is similar to that of tasks occurring routinely in standard physics teaching, both with regard to the *kind* of connections between representations (only translating and relating representations, see above) and the *number* of involved representations and connections. An example of this type would be to take the conventional task of drawing a ray diagram from givens in the task text, and to ask additionally to relate explicitly corresponding elements in both representational formats.

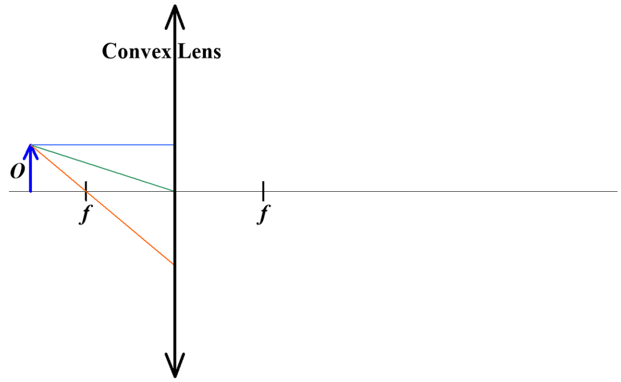
The DeFT model contains an additional important function that goes beyond that of standard RATs. The function is "construct deeper understanding" [75], which leads to a further RAT type. In accordance with the DeFT model (see above), constructing deeper understanding is implemented by RATs which require comparing, changing, correcting, or adapting representations in addition to the demands of S RATs. For these purposes, the mental model (e.g., a ray diagram) has to be further developed than it has to be for S RATs. Moreover, the number of representations ($\bar{N}_{\mathrm{RF}} = 4.2$) and their connections ($\bar{N}_{\mathrm{RFC}} = 3.9$) is quite high for these RATs.

We sum up as follows: There were 10 "standard" RATs and 19 "deeper understanding" RATs in the intervention in total. RATs and conventional tasks are different qualitatively and quantitatively. Qualitatively, RATs use *different types* of representational activities (translate, relate, compare, correct, and change or adapt), whereas conventional tasks only use a single one (translate, in the typical situation to draw a ray diagram with givens in a text). Moreover, RATs ask for *explicit* representational connections (e.g., comparing of corresponding elements, or describing differences between ray diagrams), while conventional tasks only ask for implicit connections (translate from text to diagram). Quantitatively, RATs use more representational formats, and more representational connections ($\bar{N}_{\mathrm{RF}} = 3.6$, $\bar{N}_{\mathrm{RFC}} = 3.1$) than conventional tasks ($\bar{N}_{\mathrm{RF}} = 1.7$, $\bar{N}_{\mathrm{RFC}} = 0.66$).

Details of RAT design features in contrast to conventional tasks are illustrated in the following section for several examples.

2. *Examples of representational activity tasks*

Figure 2 shows a RAT for fostering students' ability to connect several representation types with each other (RAT design point *e*, see above: support translation is part of the task). Students were asked to compare the two pictorial representation types (deeper understanding RCA demand; RAT-design points *a*, *d*) and to mark relevant differences with respect to the image formation (object-image distance, object-image size, magnification factor).



Without carrying out an experiment, find the image size. Complete the ray diagram to find the answer.

Focal length = 2.0 cm
Object distance = 3.3 cm
Object size = 1.0 cm

FIG. 4. Conventional task of the CG for practicing the construction of ray diagrams for an image-forming experiment with a convex lens.

They had to identify the parameters that are relevant for the image formation process on their own (the size of the lens is not important; in the second picture the magnification factor was different because of different object and image distance ratios). Then students needed to bring together these relevant pieces of information from the two different sources (translation from instructional text to realistic image and from instructional text to ray diagram: standard RCA demand; RAT-design points *b*, *c*, *d*) and mentally process the information. In a second step, the schematic drawing had to be adapted to the realistic picture (deeper understanding RCA demand), so that both showed the same magnification. Finally, students were asked to describe the relevant differences between both pictures verbally (translation from realistic picture or adapted ray diagram to answer text: standard RCA demand; RAT-design point *d*). In short, they were asked to analyze the differences between two similar experimental settings, which were shown by two different types of representations. As a result, they had to connect mentally the two experiments with each other and change the schematic drawing. After that, the students needed to transfer their solution into a textual representation. Thus, they practiced translating information between three different types of representation to achieve a scientifically correct solution.

Another example of a RAT is shown in Fig. 3. Here the students were asked to mark the errors of the given ray diagram and to correct them (deeper understanding RCA demand; RAT-design points *a*, *b*, *c*, *d*, *e*; to understand the task it is needed before to relate the instructional text to the given ray diagram: standard RCA demand).

After that, the students had to explain verbally what was wrong with the rays when passing through the convex lens (translation: standard RCA demand). So they had to work actively on two different types of representations and practice translating information from the pictorial into the textual type of representation. In sum, this was an exercise given to the TG to develop students' ability to detect scientifically incorrect representations and to change them into scientifically correct ones.

In contrast, the CG, on the other hand, received no RATs but practiced the usual constructing of ray diagrams (see Fig. 4). One has to read some pieces of information from the text and to relate it to the ray diagram, but relating, comparing, changing or adapting different representation formats is missing. Tasks of the CG work mainly with one single representation, not two or more.

IV. RESULTS

The results of a descriptive and a multilevel analysis of students' RCA are presented. Furthermore, the influences of control variables that affect students' RCA are reported. The psychometric values of the test instruments are already provided in Sec. III B.

Table III shows the descriptive values of the RCA instrument for the different item types and measurement times. When the intervention study began, both groups were approximately on the same level of RCA (nevertheless the small difference is taken into account by the statistical analysis). However, the RCA levels were very low, and consistently so across the sample, as expected. At the other measurement times, they scored between 24% and 70% of the maximum value. Among measurement times t2 and t3, the highest value was reached by the TG with S assessment items (at t2), the lowest the CG with DU assessment items (at t3). In general, students reached higher levels for S items than for DU items. With respect to motivation, the levels of the CG decrease slightly over time whereas the levels of TG were stable between measurement times t1 and t2, and only decrease slightly between t2 and t3.

In all item groups, the TG attained higher scores than the CG on the descriptive level. In the following, the differences between the assessment item types, the groups, and measurement times as well as influences by covariates are analyzed by a multilevel model for measuring changes (see above, Refs. [153,154]). For that, we consider the

TABLE III. Descriptive values (averages and standard deviations in percent of max. score) for S (standard) and DU (deeper understanding) assessment items separately and for the whole test, for motivation and for the different measurement times (t1: pre-intervention, t2: postintervention, t3: follow-up, 6 weeks after intervention).

		Control group		Treatment group	
Measurement time	Item type	*N*	Mean (SD)	*N*	Mean (SD)
t1	S	152	4.49 (7.96)	174	4.89 (6.62)
	DU	152	6.96 (8.67)	174	8.24 (9.48)
	S and DU	153	5.63 (6.51)	174	6.51 (6.78)
	Motivation	126	64.75 (11.60)	138	67.24 (10.31)
t2	S	155	57.78 (14.93)	171	70.13 (15.20)
	DU	155	28.10 (16.81)	171	38.19 (18.24)
	S and DU	153	43.45 (13.49)	171	54.71 (13.95)
	Motivation	156	59.38 (13.40)	140	67.91 (9.68)
t3	S	144	56.49 (19.16)	156	63.60 (14.04)
	DU	144	23.86 (16.19)	156	31.29 (16.14)
	S and DU	143	41.01 (15.03)	156	48.00 (12.23)
	Motivation	123	55.35 (17.19)	161	60.40 (16.56)

TABLE IV. Estimated changes of RCA between pre- and post-test, and pre- and follow-up test (multilevel model for measuring changes).

			Estimated change of RCA (SD)	
Group	Item type	Comparison of measurement times	Absolute	In % of max.
Control group	S and DU	t1-t2	16.92 (4.62)	38.89 (10.62)
		t1-t3	15.41 (4.94)	35.43 (11.37)
Treatment group	S and DU	t1-t2	21.06 (4.76)	48.42 (10.95)
		t1-t3	17.97 (4.24)	41.31 (9.74)

changes between different measurement times for each group, which are the main interest of the study. The changes of RCA, estimated by the multilevel analyses, are shown in Table IV for the whole instrument and in Table V for S and DU items separately.

For both item types together, the CG improved its RCA between measurement times t1 and t2 by 38% of the maximum score and the TG by 48%. Thus, the students of the TG developed RCA by about 10% more than students of the CG. Between the start of the intervention and additional six weeks of conventional teaching (measurement times 1 and 3), the TG performed better than the CG by about 6% of the maximum score. For S items, where all increases were larger, the CG had a 54% increase and the TG a 65% increase (t1–t2). For DU items, the CG increased 23% and the TG 30%.

TABLE V. Estimated changes of RCA between pre- and post-test, and pre- and follow-up test, for S and DU items separately (multilevel model for measuring changes).

			Estimated change of RCA (SD)	
Group	Item type	Comparison of measurement times	Absolute	In % of max.
Control group	S	t1-t2	12.14 (2.70)	53.96 (12.01)
		t1-t3	11.56 (3.27)	51.39 (14.53)
Treatment group	S	t1-t2	14.72 (2.69)	65.43 (11.96)
		t1-t3	13.18 (2.49)	58.58 (11.05)
Control group	DU	t1-t2	4.73 (3.52)	22.51 (13.37)
		t1-t3	3.69 (2.71)	17.59 (12.89)
Treatment group	DU	t1-t2	6.34 (4.95)	30.19 (14.50)
		t1-t3	4.77 (2.74)	22.73 (13.04)

TABLE VI. Results of the multilevel analysis for the S- and DU-assessment items together: influence of variables on students' RCA (not shown: nonsignificant variables, viz. verbal and numeric intelligence, and German language grade). The full table of all items and separated tables for S and DU assessment items are shown in Appendix B.

Parameter	β	Standard error	Degrees of freedom	t statistics	Significance	Effect size d
Motivation pretest	−0.02	0.01	294.3	−2.18	0.030	−0.12
Visuospatial ability	2.97	1.14	294.1	2.61	0.009	0.13
Grade mathematics	0.56	0.19	294.9	2.87	0.004	0.18
Grade physics	0.53	0.19	291.3	2.76	0.006	0.18
Optics concept pretest	0.13	0.04	292.9	3.38	0.001	0.18
Gender	0.99	0.30	294.2	3,28	0.001	0.17
Interaction: Difference between RCA pre- and post-test, TG vs CG	4.15	0.68	295.0	6.06	<0.000	0.69
Interaction: Difference between RCA pre- and posttest, TG vs CG	2.56	0.72	277.3	3.53	<0.000	0.43

Table VI and Fig. 5 show that the RAT intervention had a highly significant positive influence on the change of learners' RCA both between the pre- and post-test ($d = 0.69$ compared to CG) and between the pre- and follow-up test ($d = 0.43$, compared to CG). This corresponds to medium to large and small to medium effect sizes, respectively [150]. For S items and DU items taken separately, the effects between pre- and post-test are $d = 0.77$ and $d = 0.47$, respectively (compared to CG).

Moreover, Table VI and Fig. 5 indicate an influence of the following control variables: pre-instructional grades in mathematics and physics, and conceptual knowledge in geometrical optics. Higher values in these variables led to a higher RCA, regardless of whether the student was in the TG or CG, with small effect sizes (all around $d \approx 0.2$). There was a gender effect in favor of boys, appearing only for the DU assessment items ($d = 0.17^{***}$). Visuospatial ability ($d = 0.13^{***}$) had a very low effect, as well as initial motivation, (in the opposite direction, but with very small $d = -0.12^{*}$). Verbal and numeric intelligence and grade in the German language had no influence on RCA (therefore was not included in Table VI; for full results see Table VIII in Appendix B).

For S assessment items taken separately, the effect sizes of control variables were in the same range, but slightly smaller than those given above. In addition, numerical IQ had a very small effect, and gender no effect. For DU assessment items taken separately, only the effect sizes of

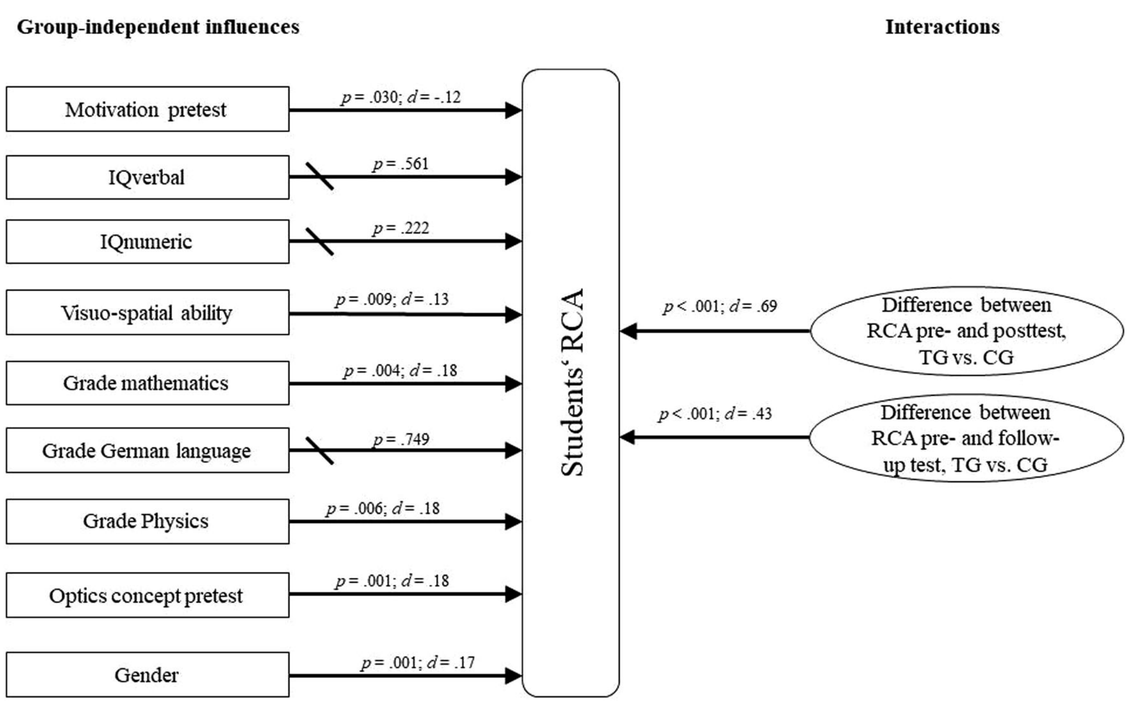


FIG. 5. Influence of variables on students' RCA, (multilevel model for measuring changes).

the control variables mathematics, physics, and initial conceptual knowledge were in the same range as those given above. The gender effect was slightly larger. However, there were no effects of motivation and intelligence (for result overviews see Table XI and Table X in Appendix B).

V. DISCUSSION

Research question 1.1. is answered as follows: Students who learned with RATs showed a significantly higher increase of representational coherence ability than students of the control group, with a sizable effect size of $d = 0.69$ (RCA post-test). These positive effects were still present at the follow-up test ($d = 0.43$), six weeks after the end of the intervention, and thus showed at least a certain degree of medium-term stability.

For answering research question 1.2. the learning effects for S items and DU items are considered separately: First, positive effects on S items were found, with effect sizes comparable to that of the complete test ($d = 0.77$ at post-test). Thus, RATs can develop that part of RCA that is implicitly required also in the standard requirements of physics learning (e.g., relating text and image) better than conventional tasks. Second, positive effects were also found for the more demanding DU items ($d = 0.47$). It is not surprising, that effects for a more demanding task are smaller than for a less demanding one, and we consider an effect of this size as encouraging for an ability known from the literature to be hard to achieve, and after only a few hours of learning time. All these results were obtained with a series of control measures to ensure comparability between the TG and CG (in particular same teacher, comparable initial situation, and control for remaining differences, in particular initial competencies, motivation and intelligence).

With regard to research question 2, no influences of verbal and numerical intelligence or of German language grades were found. The following covariates showed influences with quite low effect sizes (mostly ≈ 0.2 or smaller): prior knowledge (pre-instructional grades in mathematics and physics, initial conceptual knowledge in geometrical optics), visuospatial ability, and gender.

These findings are in line with existing research (see Sec. II C "learner characteristics"): Prior knowledge in the domain, well known as a predictor of learning in general, also has an influence on representational learning. The same holds for prior knowledge in mathematics: its influence, well known for general physics achievement, is also found for representational learning. We note that the effect size found here is smaller than the one found in previous studies. An explanation for this could be the small part of formula-related problems in this study, chosen in accord with its age group and the relevant curriculum. It seems plausible that the influence of mathematical abilities becomes more pronounced when leaners have to deal more with (symbolic) formulas, and indeed Torigoe and Gladding [117,118] have found indications for such an effect. We thus hypothesize, for further study, that the influence of prior knowledge in mathematics on representational competence (in particular RCA) might increase with the degree to which understanding of and dealing with formulas is necessary in a given content area.

We also find an influence of visuospatial ability, however, and somehow unexpectedly, with a quite small size (smaller than few values reported in the literature, see Sec. II C). A reason for this quite small effect of visuospatial ability on learning of ray optics could be the fact that the latter is considered to be a science topic which can be quite well visualized, as compared, e.g., to mechanics and electromagnetism (involving vectors, fields, etc.). In fact, this is even a reason to treat ray optics as one of the first topics in many physics curricula (there are, in particular, nice experiments to visualize rays, as also used in this work, see, e.g., Ref. [160]). Such a moderating influence of the level of abstraction is discussed in Ref. [161]. We conclude that in order to substantiate such a moderating effect on the strength of association between visuospatial ability and RCA, studies on content areas other than ray optics would be of interest, and a more systematic inclusion of effect sizes as quantitative measures of these strengths appears necessary.

Finally, we also find gender influences on RCA (with an advantage for boys), consistent with existing research. The effect size is somewhat smaller than those found previously in older age groups (see Sec. II C). A tentative explanation for this could be the so-called "Matthew effect" (or "cumulative advantage effect") in education [162]; the expression comes from the biblical saying that the rich get richer and the poorer get poorer (gospel according to Matthew). It denotes a situation where those who initially score higher on some desirable variable also gain more during learning, i.e., with the initial value also the slope as function of time is higher. Thus, a gender difference of RCA at the secondary level I (this work) would be smaller than one accumulated at university freshman level [163]. However, this question is beyond the scope of our study and cannot be answered here.

In view of potentially increased cognitive demands in the TG, the following findings concerning motivation are useful to know. First (belongs to research question 1.2.), there was no effect of initial motivation on RCA for the more demanding items (DU items): less motivated learners can develop RCA as well as more motivated ones. We do not have an explanation for the opposite effect on S items, but it was lower in significance level than all other effects and among the lowest in effect size. Second (research question 3), there was a decrease of motivation for the CG, whereas there was a slight increase of motivation from pre- to post-test for TG (the difference was however not significant). This means that the RAT intervention (with its more demanding tasks) does at least not diminish motivation.

VI. LIMITATIONS AND PERSPECTIVES

A. Limitations and research perspectives

Our findings supporting positive effects of the RAT approach on the increase of RCA are limited in the following ways by the educational setting of the study: First, it is a study about a given age group (secondary level I), and studies, for example, on the tertiary level (with increased representational demands) would be interesting. Moreover, the study was carried in the academic-track schools within the German school system[1]. Concerning the latter point, existing research points to an appreciable association of academic success and working memory [164], and as cognitive load is one of the main problems with MRs (see Sec. II A), our findings about the positive effects of RATs would have to be checked for learner groups with lower cognitive abilities. Note, however, that for a series of cognitive variables related to learning with multiple representations no or relatively weak effects and no interaction with the main effect were found (verbal and numerical intelligence, German language grades, visuospatial intelligence, mathematics and physics grades, initial conceptual knowledge, respectively). So, cognitive factors should only have a moderate influence on the positive impact of learning with RATs.

Second, the RAT effects were found for a specific topic rich in representations (ray optics). An extension to other areas in physics (mechanics) or other sciences (the micro-macro-symbolic relationship important in chemistry [6]) would be of interest. Third, the present work considered only RATs as an educational approach, in order to study their effects in a well-controlled way. In the future, it could be combined with other educational approaches of interest in the area, such as for joint gain of representational and conceptual competence [165,166], reducing cognitive load [71,72,109], for example, by worked examples, explicitly promoting self-generated representations (Ref. [4] Chap. 5; Ref. [7] Chap. 7), specific forms of collaborative learning [167–169], and others.

The following limitations on the methodological level have to be mentioned. First, there were some minor points in the RCA instrument to be improved in the future. The study provided information on how to do this, for example, either removing too difficult items, or enhancing the learning opportunities related to the content in question (see Ref. [146] for details). Consideration of covariates in this study was limited to control for their influences on students' RCA, and checking for potential interaction effects known from the literature. In view of the main research questions of the study, no in-depth discussion of the causal mechanisms behind potential interaction effects was intended. For instance, existing studies [170,171] have shown that gender differences in spatial ability are largely due to experience and can be reduced by appropriate learning measures. There is also a discussion of gender differences, and possible reasons and remedies in particular on the introductory university level (see Refs. [131,132, 172,173]). These and other aspects regarding a more fine-grained analysis of covariates would have to be considered for a more complete understanding of RCA and its increase in future studies. Finally, the joint increase of representational and other abilities should be considered (e.g., conceptual abilities [82,165]).

B. Conclusions and classroom perspectives

The present study describes the design and the outcomes of a theory-based intervention (RATs) for the improvement of the RCA of learners in the domain of geometrical optics. A multilevel analysis revealed a positive impact of the RAT intervention, with a practical relevant effect size and lasting at least partially beyond a mere short-term effect.

With regard to classroom practice, the effect size of RATs would situate at rank 13 of the 800 values presented in the research synthesis of Hattie [106]. Of course, this is an individual study, not a meta-analysis, but still the size for the RAT effects appear to in the interesting range and it was obtained after a few hours of total learning time (4.5 h, in accord with the applicable curriculum). Moreover, the overall weak influences of different covariates are also of practical interest. The RAT effect size directly after the intervention is more than 3 times larger than those of prior knowledge, visuospatial ability, and gender. Other covariates did not have an influence. This means that the intervention works for diverse kinds of learners, and that, in particular, it is not restricted to learners of a higher initial level. We thus feel that at the given stage RATs already offer a useful approach for an important component of physics learning.

In order to apply the approach in a broader context, it is necessary to develop RATs for different topics of physics and also for other parts of science education and we see this as a useful topic in initial or continuous teacher education. In this respect, the detailed description of the instructional design of the intervention, including a quantitative characterization of RATs might help for transfer to other teaching topics. Of course, the possible extensions for future research (other learner groups, other science topics, combination with other educational approaches) will also help to inform and improve practice, and work along these lines is under way.

ACKNOWLEDGMENTS

This research was funded by the German Research Association (DFG, Graduate School GK1561). The statements of this contribution are made by the authors, not by the funding body. The authors thank also Mario Gollwitzer and Alexander Kauertz for continuous support. The publication was funded by the Open Access Fund of the University of Koblenz-Landau.

APPENDIX A: SAMPLE ITEMS AND REPRESENTATIONAL FORMATS OF THE REPRESENTATIONAL COHERENCE ABILITY TEST

Fig. 6 shows an example of a standard RCA test item. It assesses an RCA involving the two representational formats text and a ray diagram, as well as relations and translations between them (see Table 8, DeFT *a*, *b*, *c*, *d*, *e*). To solve the item, the following coherence formation processes need to be performed: information given by the text has to be related with and translated in a ray diagram (standard RCA; DeFT *e*). The propagation of the rays has to be constructed following their underlying representational and conceptual rules. The focal length of the lens has to be read from the self-constructed external ray diagram and translated in the answer text. Incoherent RCA processes lead to scientifically incorrect answers.

Figure 7 shows examples of deeper understanding RCA test items.

In particular, the following coherence formation processes (standard and deeper understanding) need to be performed to solve the item 1(b): Information of the given text has to be related and translated to the given ray diagram (standard RCA; DeFT *a*, *b*, *c*, *d*, *e*). The given ray diagram has to be changed into a self-generated internal ray diagram (deeper understanding RCA). This internal ray diagram has to be compared to the given external ray diagram (i.e., if the new image-lens distance and image size is larger or shorter than the given distance or image size, respectively; deeper understanding RCA; DeFT *d*). The differences need to be read and have to be translated into an answer text (standard RCA; DeFT *e*).

In item 1(c) is needed to be verbalized, how the tasks were solved. This delivered insight into the thinking processes and RCA use of the student in connection to necessary task solving work for task 1(b). In particular, the student had to process the solution of task 1(b) again or remember it, but this time it was necessary to describe the representational reasoning itself.

An overview about the different combinations of representation types and their connections used in the RCA instrument is given in Table VII.

Standard RCA test items have only "relating and translating" as connection type. Second, they have the minimum numbers possible for an item about multiple representations for both the number of representations and the number of connections between them ($N_{RF} = 2$, $N_{RFC} = 1$).

Deeper understanding RCA test items have additionally "comparing, changing, or adapting" as connection type. Second, they have higher numbers than standard RCA test items for both the number of representations and the number of connections between them ($N_{RF} \geq 3$, $N_{RFC} \geq 2$). This is necessary to assess higher levels of students' RCA (for details see Ref. [146]).

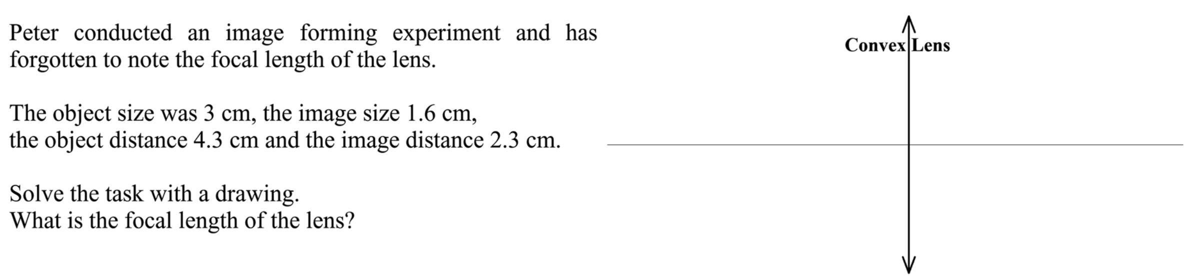


FIG. 6. Example of a standard item for assessment of RCA (item 7 [145,146]): building a ray diagram with given values.

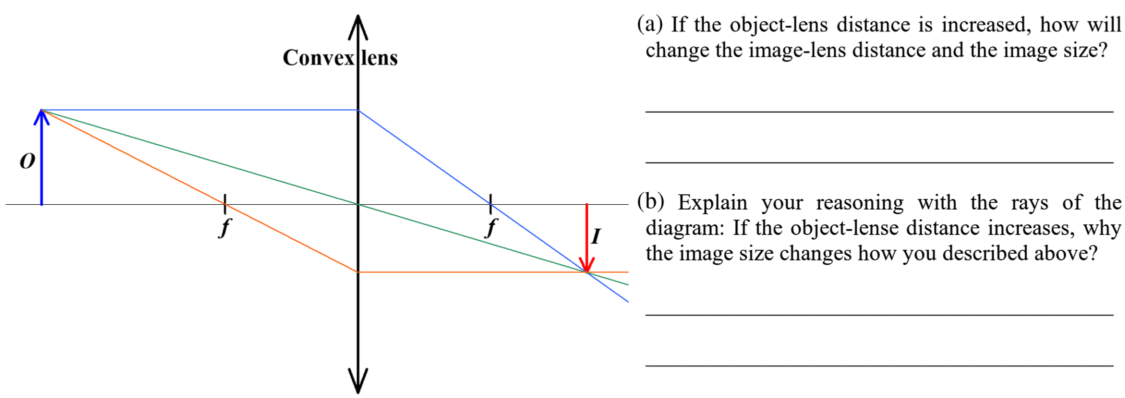


FIG. 7. Examples of deeper understanding items for assessment of RCA [items 1(b) and 1(c) [100,145,146]].

TABLE VII. Overview of the representational formats and their connections used in RCA assessment items (see Ref. [146]; N_{RF} = number of different representations, N_{RFC} = number of connections between different representations, rep. = repeated, i.e., an already constructed connection of representations is used again with another variable and counted only once, $*$ = a variable integrated in the text (e.g., focal length = 4 cm) is not counted as separate representation, $**$ = table and text combined).

				Representational formats					Quantity and type of connections between representations	
Item	Type	N_{RF}	N_{RFC}	Text	Table	Formula	Ray diagram	Realistic drawing or photography	Relate, translate	Compare, change or adapt
1(a)	Standard (S)	2	1	1*			1		1	0 (per
2		2	1	1	1				1 (rep.)	construction)
3		2	1	1**		1			1 (rep.)	
4(a)		2	1	1		1			1	
7		2	1	1			1		1 (rep.)	
N		2	1						1.0	
1(b)	Deeper	4	3	2			2		2	1 (rep.)
1(c)	understanding	4	3	2			2		2	1
4(b)	(DU)	3	2	2			1		1	1
5(a)		4	2	2				2	1	1
5(b)		4	2	2				2	1	1
5(ca)		4	3	2				2	2	1
5(cb)		4	3	2				2	2	1
6		5	5	2		1	2		2	3
8		4	4	2				2	3	1
$\bar{N}$(SD)		4.0 (0.5)	3.0 (1.0)						1.8 (0.7)	1.2 (0.7)

APPENDIX B: FULL RESULT TABLES

This section provides the significant and also the non-significant results of the multilevel model for measuring changes of the intervention study (see Table VIII for all assessment items, Table IX only for S assessment items, and Table X only for DU assessment items).

TABLE VIII. Results of the multilevel analysis for S and DU assessment items together: influence of variables on the changes of students' RCA between pre- and post-test, pre- and follow-up test, and the groups, fixed effects (multilevel model for measuring changes).

Parameter	β	Standard error	Degrees of freedom	t statistics	Significance
Intercept	−6.36	1.69	295.7	−3.76	0.000
Motivation pretest	−0.02	0.01	294.3	−2.18	0.030
IQ verbal	0.87	1.50	294.2	0.58	0.561
IQ numeric	1.49	1.22	293.0	1.22	0.222
Visuospatial ability	2.97	1.14	294.1	2.61	0.009
Grade mathematics	0.56	0.19	294.9	2.87	0.004
Grade physics	0.53	0.19	291.3	2.76	0.006
Grade German language	0.06	0.19	292.6	0.32	0.749
Optics concept pretest	0.13	0.04	292.9	3.38	0.001
Gender (male = 1)	0.99	0.30	294.2	3.28	0.001
Dummy1 ($t2 = 1$)	16.92	0.50	295.8	33.58	0.000
Dummy2 ($t3 = 1$)	15.41	0.53	277.1	29.04	0.000
Condition (TG = 1)	0.38	0.33	284.3	1.17	0.242
Interaction: Difference between RCA pre- and post-test, TG vs CG	4.15	0.68	295.0	6.06	0.000
Interaction: Difference between RCA pre- and post-test, TG vs CG	2.56	0.72	277.3	3.53	0.000

TABLE IX. Results of the multilevel analysis for S items only: influence of variables on the change of students' RCA between pre- and post-test, pre- and follow-up test and the groups, fixed effects (multilevel model for measuring changes).

Parameter	β	Standard error	Degrees of freedom	t statistics	Significance
Intercept	−2.97	0.97	303,2	−3, 05	0.002
Motivation pretest	−0.01	0.01	297.4	−2.00	0.047
IQ verbal	−0.82	0.85	298.9	−0.96	0.338
IQ numeric	1.40	0.69	296.2	2.02	0.044
Visuospatial ability	1.63	0.65	296.6	2.52	0.012
Grade mathematics	0.24	0.11	297.8	2.13	0.034
Grade German language	0.10	0.11	295.3	0.91	0.364
Grade physics	0.25	0.11	294.2	2.27	0.024
Optics concept pre test	0.07	0.02	296.3	3.21	0.001
Gender (male = 1)	0.19	0.17	296.9	1.09	0.275
Dummy1 ($t2 = 1$)	12.14	0.30	287.1	41.00	0.000
Dummy2 ($t3 = 1$)	11.56	0.36	279.8	32.18	0.000
Condition (TG = 1)	0.04	0.19	292.6	0.22	0.825
Interaction: Difference between RCA pre- and post-test, TG vs CG	2.58	0.40	285.9	6.42	0.000
Interaction: Difference between RCA pre- and post-test, TG vs CG	1.62	0.49	279.9	3.30	0.001

TABLE X. Results of the multilevel analysis for DU assessment items only: influence of variables on students' RCA, fixed effects (multilevel model for measuring changes).

Parameter	β	Standard error	Degrees of freedom	t statistics	Significance
Intercept	−3.51	1.10	306.3	−3.18	0.002
Motivation pretest	−0.01	0.01	300.4	−1.73	0.085
IQ verbal	1.72	0.97	301.7	1.78	0.077
IQ numeric	0.14	0.79	299.8	0.17	0.864
Visuospatial ability	1.32	0.73	300.1	1.80	0.073
Grade mathematics	0.33	0.13	301.0	2.63	0.009
Grade German language	−0.02	0.12	299.1	−0.16	0.876
Grade physics	0.27	0.12	298.3	2.21	0.028
Optics concept pretest	0.07	0.03	299.9	2.64	0.009
Gender (male = 1)	0.77	0.20	300.1	3.93	0.000
Dummy1 ($t2 = 1$)	4.73	0.32	298.3	14.70	0.000
Dummy2 ($t3 = 1$)	3.69	0.30	284.0	12.27	0.000
Condition (TG = 1)	0.35	0.21	288.9	1.65	0.100
Interaction: Difference between RCA pre- and post-test, TG vs CG	1.61	0.44	296.8	3.69	0.000
Interaction: Difference between RCA pre- and post-test, TG vs CG	1.08	0.41	284.1	2.63	0.009

[1] *Forms of Representation*, edited by D. Peterson (Intellect Books, Exeter, 1996).
[2] A. Paivio, *Mental Representations: A Dual Coding Approach* (Oxford University Press, New York, 1986).
[3] R. E. Mayer and R. B. Anderson, Animations need narrations: An experimental test of a dual-coding hypothesis, J. Educ. Psychol. **83**, 484 (1991).
[4] *Constructing Representations to Learn in Science*, edited by R. Tytler, V. Prain, P Hubber, and B. Waldrip (Sense, Rotterdam, 2013).
[5] *Multiple Representations in Biological Education*, edited by C. Tsui and D. Treagust (Springer, Dordrecht, Netherlands, 2013).
[6] *Multiple Representations in Chemical Education,* edited by J. K. Gilbert and D. Treagust (Springer, Netherlands, 2009).

[7] *Multiple Representations in Physics Education,* edited by D. Treagust, R. Duit, and H. E. Fischer (Springer, Dordrecht, Netherlands, 2017).
[8] K. S. Sell, B. E. Herbert, C. L. Stuessy, and J. Schielack, Supporting student conceptual model development of complex earth system through the use of multiple representations and inquiry, J. Geosci. Educ. **54**, 396 (2006).
[9] *Problems of Representation in the Teaching and Learning of Mathematics*, edited by C. Janvier (Erlbaum, Hillsdale, NJ, 1987).
[10] *Mental Models and the Mind: Current Developments in Cognitive Psychology, Neuroscience, and Philosophy of Mind, Advances in Psychology, 138*, edited by C. Held, M. Knauff, and G. Vosgerau (Elsevier, Amsterdam, 2006).
[11] L. Verschaffel, E. De Corte, T. de Jong, and J. Elen, *Use of External Representations in Reasoning and Problem Solving* (Routledge, New York, 2010).
[12] T. Seufert, Supporting coherence formation in learning from multiple representations, Learn. Instr. **13**, 227 (2003).
[13] R. E. Mayer, Cognitive theory of multimedia learning, in *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, New York, 2005), pp. 31–48.
[14] W. Schnotz, An integrated model of text and picture comprehension, in *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, New York, 2005), pp. 49–70.
[15] A. Van Heuvelen, Learning to think like a physicist: A review of research-based instructional strategies, Am. J. Phys. **59**, 891 (1991).
[16] L. C. McDermott, Oersted Medal Lecture 2001: Physics Education Research—The Key to Student Learning, Am. J. Phys. **69**, 1127 (2001).
[17] R. Plötzner and H. Spada, Inhalt, Struktur und Anwendung von Physikwissen: Eine psychologische Perspektive, Z. Didaktik der Naturwissenschaften **4**, 81 (1998).
[18] D. E. Meltzer, Relation between students' problem-solving performance and representational format, Am. J. Phys. **73**, 463 (2005).
[19] J. L. Docktor and J. P. Mestre, Synthesis of discipline-based education research in physics, Phys. Rev. ST Phys. Educ. Res. **10**, 020119 (2014).
[20] R. B. Kozma and J. Russell, Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena, J. Res. Sci. Teach. **34**, 949 (1997).
[21] H. K. Wu, J. S. Krajcik, and E. Soloway, Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom, J. Res. Sci. Teach. **38**, 821 (2001).
[22] K. J. Schönborn and S. Bögeholz, Experts' Views on Translation Across Multiple External Representations in Acquiring Biological Knowledge About Ecology, Genetics, and Evolution, in *Multiple Representations in Biological Education (Models and Modeling in Science Education)*, edited by D. F. Treagust and C. Y. Tsui (Springer, Dordrecht, Netherlands, 2013).
[23] R. Lesh, T. Post, and M. Behr, Representations and translations among representations in Mathematics learning and problem solving, in *Problems of Representation in the Teaching and Learning of Mathematics*, edited by C. Janvier (Erlbaum, Hillsdale, NJ, 1987), pp. 33–40.
[24] B. Waldrip, V. Prain, and J. Carolan, Using Multi-Modal Representations to Improve Learning in Junior Secondary Science, Res. Sci. Educ. **40**, 65 (2010).
[25] P. Hubber, R. Tytler, and F. Haslam, Teaching and Learning about Force with a Representational Focus: Pedagogy and Teacher Change, Res. Sci. Educ. **40**, 5 (2010).
[26] P. Nieminen, A. Savinainen, and J. Viiri, Relations between representational consistency, conceptual understanding of the force concept, and scientific reasoning, Phys. Rev. ST Phys. Educ. Res. **8**, 010123 (2012).
[27] P. B. Kohl and N. D. Finkelstein, Student representational competence and self-assessment when solving physics problems, Phys. Rev. ST Phys. Educ. Res. **1**, 010104 (2005).
[28] P. B. Kohl and N. D. Finkelstein, Effect of instructional environment on physics students' representational skills, Phys. Rev. ST Phys. Educ. Res. **2**, 010102 (2006).
[29] P. B. Kohl and N. D. Finkelstein, Effects of representation on students solving physics problems: A fine-grained characterization, Phys. Rev. ST Phys. Educ. Res. **2**, 010106 (2006).
[30] D. Rosengrant, A. Van Heuvelen, and E. Etkina, Do students use and understand free-body diagrams?, Phys. Rev. ST Phys. Educ. Res. **5**, 010108 (2009).
[31] W. Schnotz, C. Baadte, A. Müller, and R. Rasch, Creative thinking and problem solving with depictive and descriptive representations, in *Use or Representations in Reasoning and Problem Solving: Analysis and Improvement*, edited by L. Verschaffel, E. De Corte, T. de Jong, and J. Elen (Routledge, Milton Park, UK, 2010), pp. 11–35.
[32] T. Seufert and R. Brünken, Cognitive load and the format of instructional aids for coherence formation, Appl. Cogn. Psychol. **20**, 321 (2006).
[33] R. E. Mayer, Systematic thinking fostered by illustrations in scientific text, J. Educ. Psychol. **81**, 240 (1989).
[34] R. E. Mayer, Rote versus meaningful learning, Theory Into Practice **41**, 226 (2002).
[35] P. B. Kohl, D. Rosengrant, and N. D. Finkelstein, Strongly and weakly directed approaches to teaching multiple representation use in physics, Phys. Rev. ST Phys. Educ. Res. **3**, 010108 (2007).
[36] R. B. Kozma and J. Russell, Students becoming chemists: Developing representational competence, in *Visualization in Science Education*, edited by J. Gilbert (Springer, Dordrecht, 2005).
[37] P. Nieminen, A. Savinainen, and J. Viiri, Force concept inventory-based multiple-choice test for investigating students' representational consistency, Phys. Rev. ST Phys. Educ. Res. **6**, 020109 (2010).
[38] A. Saniter, Spezifika der Verhaltensmuster fortgeschrittener Studierender der Physik, in *Studien zum Physiklernen*, edited by H. Niedderer and H. Fischler (Logos, Berlin, 2003), Vol. 28.

[39] ATLAS, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B **716**, 1 (2012).

[40] D. Hodson, Practical work in science: Time for a reappraisal, Studies Sci. Educ. **19**, 175 (1991).

[41] I. Abrahams and R. Millar, Does practical work really work? A study of the effectiveness of practical work as a teaching and learning method in school science, Int. J. Sci. Educ. **30**, 1945 (2008).

[42] K. Miller, N. Lasry, K. Chu, and E. Mazur, Role of physics lecture demonstrations in conceptual learning, Phys. Rev. ST Phys. Educ. Res. **9**, 020113 (2013).

[43] A. diSessa, Science learning, in *Encyclopedia of Education*, edited by J. W. Guthrie (Macmillan, New York, 2002).

[44] J. Dolin, Science education standards and science assessment in Denmark, in *Making it Comparable, Standards in Science Education*, edited by D. Waddington, P. Nentwig, and S. Schanze (Waxmann, Münster, 2007), p 77.

[45] R. Justi, J. K. Gilbert, and P. F. Ferreira, The application of a ‘model of modelling’ to illustrate the importance of metavisualisation in respect of the three types of representation, in *Multiple Representations in Chemical Education*, edited by J. K. Gilbert and D. F. Treagust (Springer, London, 2009), pp. 285–307.

[46] S. Ainsworth, The functions of multiple representations, Comput. Educ. **33**, 131 (1999).

[47] R. Kozma, Representation and language: The case for representational competence in the chemistry curriculum, *Proceedings of the Biennial Conference on Chemical Education, Ann Arbor, Michigan* (ACS, Washington, DC, 2000).

[48] A. Fenk, Symbols and icons in diagrammatic representation, Pragmatics Cogn. **6**, 301 (1998).

[49] W. Schnotz, Towards an integrated view of learning from text and visual displays, Educ. Psychol. Rev. **14**, 101 (2002).

[50] W. M. Roth and G. M. Bowen, Complexities of graphical representations during lectures: A phenomenological approach, Learn. Instr. **9**, 235 (1999).

[51] V. Prain and R. Tytler. Learning Through Constructing Representations in Science: A framework of representational construction affordances, Int. J. Sci. Educ. **34**, 2751 (2012).

[52] K. Hülser, Zeichenkonzeptionen in der Philosophie der griechischen und römischen Antike, in *A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, edited by R. Posner, K. Robering, and T. A. Sebeok (de Gruyter, Berlin, New York, 1997), Vol. 1.

[53] C. Peirce, Logic as semiotic: The theory of signs, in *Philosophical Writings of Peirce (1893-1910)*, edited by J. Buchler (Dover, New York, Reprint 1955, 1931), pp. 98–119.

[54] C. K. Ogden and I. A. Richards, *The Meaning of Meaning* (Routledge, London, New York, 1923).

[55] U. Eco, *A Theory of Semiotics* (Indiana University, Bloomington, 1997).

[56] P. Hubber and R. Tytler, Enacting a Representation Construction Approach to Teaching and Learning Astronomy, in *Multiple Representations in Physics Education*, edited by D. Treagust, R. Duit, and H. E. Fischer (Springer, Dordrecht, Netherlands, 2017).

[57] D. P. Ausubel, *The Psychology of Meaningful Verbal Learning* (Grune and Stratton, New York, 1963).

[58] D. P. Ausubel, *The Acquisition and Retention of Knowledge: A Cognitive View* (Kluwer Academic Publishers, Dordrecht, 2000).

[59] M. C. Wittrock, Learning science by generating new conceptions from old ideas, in *Cognitive Structure and Conceptual Change*, edited by L. H. T. West and A. L. Pines (Academic Press, Orlando, 1985).

[60] J. D. Novak and D. B. Gowin, *Learning How to Learn* (Cambridge University Press, Cambridge, England, 1984).

[61] H. Muckenfuß, *Lernen im sinnstiftenden Kontext, Entwurf einer zeitgemäßen Didaktik des Physikunterrichts* (Cornelsen, Berlin, 1995).

[62] *Teaching Science for Understanding*, edited by J. J. Mintzes, J. H. Wandersee, and J. D. Novak (Academic Press, San Diego, 1998).

[63] J. Leisen, Physikalische Begriffe und Sachverhalte: Repräsentationen auf verschiedenen Ebenen, Praxis der Naturwissenschaften Physik **47**, 14 (1998).

[64] S. Vosniadou, Instructional considerations in the use of external representations: The distinction between perceptually based depictions and pictures that represent conceptual models, in *Use of Representations in Reasoning and Problem Solving—Analysis and Improvement*, edited by L. Verschaffel, E. d. Corte, T. d. Jong, and J. Elen (Routledge, London, 2010), pp. 36–54.

[65] J. Leisen, Förderung des Sprachlernens durch den Wechsel von Symbolisierungsformen im Physikunterricht, Praxis der Naturwissenschaften Physik **47**, 9 (1998).

[66] W. Schnotz and M. Bannert, Construction and interference in learning from multiple representations, Learn. Instr. **13**, 141 (2003).

[67] P. N. Johnson-Laird, Mental Models, Sentential Reasoning, and Illusory Inferences, in *Mental models and the mind: Current Developments in cognitive psychology, neuroscience, and philosophy of mind, Advances in Psychology*, edited by C. Held, M. Knauff, and G. Vosgerau (Elsevier, Amsterdam, 2006).

[68] P. Johnson-Laird, *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness* (Cambridge University Press; Harvard University Press, Cambridge, England; Cambridge, MA, 1983).

[69] D. Gentner, The mechanisms of analogical learning, in *Similarity and Analogical Reasoning*, edited by S. Vosniadou and A. Ortony (Cambridge University Press, Cambridge, England, 1989), pp. 197–241.

[70] J. Sweller, *Instructional Design in Technical Areas* (ACER Press Melbourne, Australia, 1999).

[71] J. Sweller, Implications of cognitive load theory for multimedia learning, in *The Cambridge Handbook of Multimedia Learning*, edited by R. E. Mayer (Cambridge University Press, New York, 2005), pp. 19–30.

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

[73] L. Cronbach and R. Snow, Aptitudes and Instructional Methods: A Handbook for Research on Interactions (Irvington New York 1977).

[74] S. Kalyuga, P. Ayres, P. Chandler, and J. Sweller, The expertise reversal effect, Educ. Psychol. **38**, 23 (2003).

[75] S. Ainsworth, DeFT: A conceptual framework for considering learning with multiple representations, Learn. Instr. **16**, 183 (2006).

[76] S. Ainsworth, The educational value of multiple representations when learning complex scientific concepts, in *Visualization: Theory and Practice in Science Education*, edited by J. Gilbert, M. Reiner, and M. Nakhleh (Springer, Dordrecht, 2008), pp. 191–208.

[77] D. Bodemer, R. Ploetzner, I. Feuerlein, and H. Spada, The active integration of information during learning with dynamic and interactive visualisations, Learn. Instr. **14**, 325 (2004).

[78] K. Berthold and A. Renkl, Instructional aids to support a conceptual understanding of multiple representations, J. Educ. Psychol. **101**, 70 (2009).

[79] A. M. Glenberg and W. F. Langston, Comprehension of illustrated text: Pictures help to build mental models, J. Memory Language **31**, 129 (1992).

[80] R. E. Mayer, Multimedia learning: Are we asking the right questions?, Educ. Psychol. **32**, 1 (1997).

[81] W. Schnotz, E. Picard, and A. Hron, How do successful and unsuccessful learners use texts and graphics?, Learn. Instr. **3**, 181 (1993).

[82] A. Müller, R. Hettmannsperger, J. Scheid, and W. Schnotz, Representational competence, understanding of experiments, phenomena and basic concepts in geometrical optics—A representational approach, in *Multiple Representations in Physics Education*, edited by D. Treagust, R. Duit, and H. E. Fischer (Springer Dordrecht, Netherlands, 2017).

[83] V. N. Lunetta, A. Hofstein, and M. P. Clough, Learning and teaching in the school science laboratory: An analysis of research, theory, and practice, in *Handbook of Research on Science Education*, edited by S. K. Abell and N. G. Lederman (Lawrence Erlbaum Mahwah, NJ, 2007), pp. 393–441.

[84] D. R. Krathwohl, A Revision of Bloom's taxonomy: An overview, Theory Into Practice **41**, 212 (2002).

[85] L. F. Koziol and D. L. Budding, Procedural learning, in *Encyclopedia of the Sciences of Learning*, edited by N. M. Seel (Springer, New York, 2012), pp. 2694–2696.

[86] L. Haapsalo and D. Kadjievich, Two types of mathematical knowledge and their relation, J. Mathematik-Didaktik **21**, 139 (2000).

[87] R. Plötzner, Pictorial representations and learning, in *Encyclopedia of the Sciences of Learning*, edited by N. M. Seel (Springer, New York, 2012).

[88] J. Scheid, A. Müller, R. Hettmannsperger, and W. Schnotz, Scientific experiments, multiple representations, and their coherence: A task based elaboration strategy for ray optics, in *Multiple Perspectives on Teaching and Learning*, edited by W. Schnotz, A. Kauertz, H. Ludwig, A. Müller, and J. Pretsch (Palgrave Macmillan, Basingstoke, UK, 2015).

[89] M. Chi, P. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5**, 121 (1981).

[90] A. diSessa, D. Hammer, B. Sherin, and T. Kolpakowski, Inventing graphing: Metarepresentational expertise in children, J. Math. Behav. **10**, 117 (1991).

[91] K. Dunbar, How scientists really reason: Scientific reasoning in real-world laboratories, in *The Nature of Insight*, edited by R. Sternberg and J. Davidson (MIT Press, Cambridge, MA, 1997), pp. 365–396.

[92] C. Goodwin, Seeing in depth, Soc. Studies Sci. **25**, 237 (1995).

[93] R. B. Kozma, E. Chin, J. Russell, and N. Marx, The role of representations and tools in the chemistry laboratory and their implications for chemistry learning, J. Learn. Sci. **9**, 105 (2000).

[94] W. M. Roth and M. McGinn, Inscriptions: Toward a theory of representing as social practice, Rev. Educ. Res. **68**, 35 (1998).

[95] I. Devetak, M. W. Urbančič, K. S. Grm, D. Krnel, and S. A. Glažar, Submicroscopic representations as a tool for evaluating students' chemical conceptions, Acta Chim. Slov. **51**, 799 (2004); http://acta-arhiv.chem-soc.si/51/51-4-799.pdf.

[96] R. B. Thiele and D. F. Treagust, An interpretive examination of high school chemistry teachers' analogical explanations, J. Res. Sci. Teach. **31**, 227 (1994).

[97] D. F. Treagust, G. D. Chittleborough, and T. L. Mamiala, The role of sub-microscopic and symbolic representations in chemical explanations, Int. J. Sci. Educ. **25**, 1353 (2003).

[98] M. Tesch, Das Experiment im Physikunterricht, in *Studien zum Physik- und Chemielernen*, edited by H. Niedderer, H. Fischler, and E. Sumfleth (Logos, Berlin, 2005), Vol. 42.

[99] V. Lee, Examining patterns of visual representation use in middle school science classrooms, *Proceedings of the National Association of Research in Science Teaching (NARST) Annual Meeting Compact Disc* (Omnipress, Garden Grove, CA, 2009).

[100] J. Scheid. Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: Ein Beitrag zur Entwicklung der Aufgabenkultur, in *Studien zum Physik- und Chemielernen*, edited by H. Niedderer, H. Fischler, and E. Sumfleth (Logos, Berlin, 2013).

[101] R. Cox and P. Brna, Supporting the use of external representations in problem solving: The need for flexible learning environments, J. Artificial Intell. Educ. **6**, 239 (1995).

[102] R. Tytler, V. Prain, P. Hubber, and F.Haslam, Reasoning in science through representation, in *Constructing Representations to Learn in Science*, edited by R. Tytler, V. Prain, P Hubber, and B. Waldrip (Sense, Rotterdam, 2013), Chap. 6.

[103] M. Van Someren, P. Reimann, H. P. A. Boshuizen, and T. de Jong, *Learning with Multiple Representation* (Elsevier, Amsterdam, 1998).

[104] A. Van Heuvelen and X. Zou, Multiple representations of work-energy processes, Am. J. Phys. **69**, 184 (2001).

[105] M. Opfermann, A. Schmeck, and H. E. Fischer, Multiple representations in physics and science education–why

should we use them?, in *Multiple Representations in Physics Education*, edited by D. Treagust, R. Duit, and H. E. Fischer (Springer, Dordrecht, Netherlands, 2017).

[106] A. C. Hattie, *Visible Learning, A Synthesis of over 800 Meta-Analyses Relating to Achievement* (Routledge, London, New York, 2009).

[107] G. Molinari and I. Tapiero, Integration of new domain-related states and events from texts and illustrations by subjects with high and low prior knowledge, Learn. Instr. **17**, 304 (2007).

[108] M. P. Cook, Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles, Sci. Educ. **90**, 1073 (2006).

[109] J. M. Nyachwaya and M. Gillaspie, Features of representations in general chemistry textbooks: A peek through the lens of the cognitive load theory, Chem. Educ. Res. Pract. **17**, 58 (2016).

[110] H. Yarden and A. Yarden, Learning and teaching biotechnological methods using animations, in *Multiple Representations in Biological Education*, edited by C. Tsui and D. Treagust (Springer, Dordrecht, 2013).

[111] D. Bodemer and U. Faust, External and mental referencing of multiple representations, Comput. Human Behav. **22**, 27 (2006).

[112] R. Karam, Thematic issue: The interplay of physics and mathematics: Historical, philosophical and pedagogical considerations, Sci. Educ. **24**, 487 (2015).

[113] A. Thorndike, Correlation between physics and mathematics grades, Schol Sci. Math. **46**, 650 (1946).

[114] M. L. Fleming and Malone, The relationship of student characteristics and student performance in science as viewed by meta-analysis research, J. Res. Sci. Teach. **20**, 481 (1983).

[115] D. E. Meltzer, The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores, Am. J. Phys. **70**, 1259 (2002).

[116] M. Lipsey and D. Wilson, *Practical Meta-Analysis, Applied Social Research Methods Series* (Sage, Thousand Oaks, CA, 2003), Vol. 49.

[117] E. Torigoe and G. Gladding, Symbols: Weapons of math destruction, AIP Conf. Proc. **951**, 200 (2007).

[118] E. Torigoe and G. Gladding, Connecting symbolic difficulties with failure in physics, Am. J. Phys. **79**, 133 (2011).

[119] R. E. Pepper, S. V. Chasteen, S. J. Pollock, and K. K. Perkins, Observations on student difficulties with mathematics in upper-division electricity and magnetism, Phys. Rev. ST Phys. Educ. Res. **8**, 010111 (2012).

[120] O. Uhden, R. Karam, M. Pietrocola, and G. Pospiech, Modelling mathematical reasoning in physics education, Sci. Educ. **21**, 485 (2012).

[121] B. R. Wilcox, M. D. Caballero, D. A. Rehn, and S. J. Pollock, Analytic framework for students' use of mathematics in upper-division physics, Phys. Rev. ST Phys. Educ. Res. **9**, 020119 (2013).

[122] L. Bollen, P. van Kampen, and M. De Cock, Students' difficulties with vector calculus in electrodynamics, Phys. Rev. ST Phys. Educ. Res. **11**, 020129 (2015).

[123] R. E. Mayer and R. Moreno, Nine ways to reduce cognitive load in multimedia learning, Educ. Psychol. **38**, 43 (2003).

[124] M. S. Khine, *Visual-spatial Ability in STEM Education* (Springer International Publishing, Cham, CH, 2017).

[125] H. K. Wu and P. Shah, Exploring visuospatial thinking in chemistry learning, Sci. Educ. **88**, 465 (2004).

[126] M. Stieff. A localized model of spatial cognition in chemistry, Ph.D. thesis, Northwestern University, Evanston, 2004.

[127] C. S. Carter, M. A. LaRussa, and G. M. Bodner, A study of two measures of spatial ability as predictors of success in different levels of general chemistry, J. Res. Sci. Teach. **24**, 645 (1987).

[128] B. Becker, Gender and science achievement: A reanalysis of studies from two meta-analyses, J. Res. Sci. Teach. **26**, 141 (1989).

[129] V. E. Lee and D. T. Burkam, Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis, Sci. Educ. **80**, 613 (1996).

[130] R. A. Louis and J. M. Mistele, The differences in scores and self-efficacy by student gender in mathematics and science, Int. J. Sci. Math. Educ. **10**, 1163 (2012).

[131] L. Kost, S. Pollock, and N. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[132] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, Phys. Rev. ST Phys. Educ. Res. **9**, 020121 (2013).

[133] I. V. S. Mullis, M. O. Martin, C. A. Minnich, G. M. Stanco, A. Arora, V. A. S. Centurino, and C. E. Castle, *TIMSS 2011 encyclopedia: Education policy and curriculum in mathematics and science (Vol. 1 & Vol. 2)* (TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA, 2012).

[134] W. R. Shadish, T. D. Cook, and D. T Campbell, *Experimental and Quasi- Experimental Designs for Generalized Causal Inference* (Houghton Mifflin, Boston, 2002).

[135] C. S. Taylor, *Validity and Validation (Understanding Statistics)* (Oxford University Press, Oxford, 2013).

[136] *Research Methods in Psychology*, edited by J. A. Schinka and W. F Velicer (Wiley, New York, 2003), Vol. 2, pp. 87–114.

[137] G. Marczyk, D. DeMatteo, and D. Festinger, *Essentials of research design and methodology.* (Wiley, New York, 2005).

[138] L. Hoffmann, P. Häußler, and S. Peters-Haft, *An den Interessen von Mädchen und Jungen orientierter Physikunterricht. Ergebnisse eines BLK-Modellversuches* (Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN), University of Kiel, Kiel, Germany, 1997).

[139] F. Rheinberg and M. Wendland, *Veränderung der Lernmotivation in Mathematik und Physik: Eine Komponentenanalyse und der Einfluss elterlicher sowie schulischer Kontextfaktoren (Rh 14/8-1), Abschlussbericht zum DFG-Projekt, Dokumentation* (Universität Potsdam, Potsdam, Germany, 2003).

[140] J. Kuhn, *Authentische Aufgaben im theoretischen Rahmen von Instruktions- und Lehr-Lern-Forschung: Effektivität und Optimierung von Ankermedien für eine*

neue Aufgabenkultur im Physikunterricht (Vieweg & Teubner Verlag, Wiesbaden, 2010).

[141] D. Liepmann, A. Beauducel, B. Brocke, and R. Amthauer, *Intelligenz-Struktur-Test 2000 R* (I-S-T 2000 R, 2nd ed.) (Hogrefe, Göttingen, 2007).

[142] D. Liepmann, A. Beauducel, B. Brocke, and W. Nettelnstroth, *IST-Screening: Intelligenz-Struktur-Test—Screening* (Hogrefe, Göttingen, 2012).

[143] L. Schmidt-Atzert and W. Rauch, TBS-TK Rezension: Intelligenz-Struktur-Test 2000 R (I-S-T 2000 R, 2nd ed.), Report Psychologie **33**, 6 (2008).

[144] R. Hettmannsperger, W. Schnotz, A. Müller, and J. Scheid, Understanding basic concepts in ray optics: A representational approach, in *Multidisciplinary Research on Teaching and Learning*, edited by W. Schnotz, A. Kauertz, H. Ludwig, A. Müller, and J. Pretsch (Palgrave Macmillan, Basingstoke, UK, 2015).

[145] J. Scheid, A. Müller, R. Hettmannsperger, and W. Schnotz, Representational competence in science education–from theory to assessment, in *Towards a Framework for Representational Competence in Science Education. Models and Modeling in Science Education*, edited by K. Daniel (Springer, Cham, Switzerland, 2018).

[146] J. Scheid, A. Müller, R. Hettmannsperger, and J. Kuhn. Erhebung von repräsentationaler Kohärenzfähigkeit von Schülerinnen und Schülern im Themenbereich Strahlenoptik, Z. Didaktik der Naturwissenschaften **23**, 181 (2017).

[147] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5**, 020103 (2009).

[148] E. Guadagnoli and W. F. Velicer, Relation of sample size to the stability of component patterns, Psychol. Bull. **103**, 265 (1988).

[149] D. V. Cicchetti, Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology, Psychol. Assess. **6**, 284 (1994).

[150] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Erlbaum Hillsdale, NJ, 1988).

[151] F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner, G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, Behav. Res. Meth. Instrum. Comput. **39**, 175 (2007).

[152] R. Heck, S. Thomas, and L. Tabata, *Multilevel and Longitudinal Modeling with SPSS* (Routledge, New York, 2014).

[153] R. Göllner, M. Gollwitzer, J. Heider, A. Zaby, and A. Schröder, Auswertung von Längsschnittdaten mit hierarchischen linearen Modellen, Zeitschrift für Klinische Psychologie und Psychotherapie **39**, 179 (2010).

[154] J. D. Singer and J. B. Willett, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (Oxford University Press, New York, 2003).

[155] IBM Corp. IBM SPSS Statistics for Windows, Version 24.0. (IBM Corp, Armonk, NY, 2016).

[156] *Handbook of Advanced Multilevel Analysis*, edited by J. J. Hox and J. K. Roberts (Routledge, New York, NY, 2010).

[157] P. Tymms, Effect sizes in multilevel models, in *But What does it Mean? The Use of Effect Sizes in Educational Research*, edited by I. Schagen and K. Elliot (Institute of Education, University of London, 2004), pp. 55–66.

[158] A. Feingold, A Regression Framework for effect size assessments in longitudinal modeling of group differences, Rev. Gen. Psychol. **17**, 111 (2013).

[159] MBWW—Ministerium für Bildung, Wissenschaft und Weiterbildung Rheinland-Pfalz, *Lehrplanentwurf Physik–Realschule, Gymnasium (Klassen 7/8–10)*, (Sommer Druck und Verlag, Grünstadt, 1998).

[160] K. Kaewkhong, N. Emarat, K. Arayathanitkul, and A. Mazzolini, A Simple Apparatus for Demonstrating How Light Rays Emanate From a Point Source, Phys. Teach. **47**, 243 (2009).

[161] R. Ploetzner and R. A. Lowe, Systematic characterisation of expository animations, Comput. Human Behav. **28**, 781 (2012).

[162] H. J. Walberg and S. Tsai, Matthew effects in education, Am. Educ. Res. J. **20**, 359 (1983).

[163] P. Nieminen, A. Savinainen, and J. Viiri, Gender differences in learning of the concept of force, representational consistency, and scientific reasoning, Int. J. Sci. Math. Educ. **11**, 1137 (2013).

[164] S. E. Gathercole, S. J. Pickering, C. Knight, and Z. Stegmann, Working memory skills and educational attainment: Evidence from National Curriculum assessments at 7 and 14 years of age, Appl. Cogn. Psychol. **18**, 1 (2004).

[165] A. diSessa, Metarepresentation: Native competence and targets for instruction, Cognit. Instr. **22**, 293 (2004).

[166] R. Hettmannsperger, A. Mueller, J. Scheid, and W. Schnotz, Developing conceptual understanding in ray optics via learning with multiple representations, Z. Erziehungswiss. **19**, 235 (2016).

[167] T. White and R. Pea, Distributed by design: On the promises and pitfalls of collaborative learning with multiple representations, J. Learn. Sci. **20**, 489 (2011).

[168] L. Liu and C. E. Hmelo-Silver, Promoting the collaborative use of cognitive and metacognitive skills through conceptual representations in hypermedia, in *Multiple Representations in Biological Education*, edited by C. Tsui and D. Treagust (Springer, Dordrecht, 2013), Chap. 2.

[169] K. J. Topping, C. Buchs, D. Duran, and K. H. Van, *Effective Peer Learning: From Principles to Practical Implementation* (Routledge, Taylor & Francis Group, London, New York, 2017).

[170] J. E. Roberts and M. A. Bell, Sex differences on a computerized mental rotation task disappear with computer familiarization, Percept. Motor Skil1s **91**, 1027 (2000).

[171] R. M. Scali, S. Brownlow, and J. L Hicks, Gender differences in spatial task performance as a function of speed or accuracy orientation, Sex Roles **43**, 359 (2000).

[172] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. **74**, 118 (2006).

[173] K. Kreutzer and A. Boudreaux, Preliminary investigation of instructor effects on gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **8**, 010120 (2012).