Relationship between semiotic representations and student performance in the context of refraction

Cedric Linder†

Department of Physics and Astronomy, Uppsala University, Uppsala, Sweden
and Department of Physics and Astronomy, University of the Western Cape, Bellville, South Africa

Jesper Bruun‡

Department of Science Education, University of Copenhagen, Copenhagen, Denmark

Arvid Pohl§

Department of Physics and Electrical Engineering, Linnaeus University, Kalmar, Sweden

Burkhard Priemer*

Department of Physics, Humboldt-Universität zu Berlin, Berlin, Germany

(Received 9 September 2022; revised 13 November 2023; accepted 22 December 2023; published 5 February 2024)

Social semiotic discussions about the role played by representations in effective teaching and learning in areas such as physics have led to theoretical proposals that have a strong common thread: in order to acquire an appropriate understanding of a particular object of learning, access to the disciplinary relevance aspects in the representations used calls for the attainment of representational competence across a particular critical constellation of systematically used semiotic resources (which are referred to as modes, see more on this later). However, an affirming empirical investigation into the relationship between a particular object of learning and different representational formulations, particularly with large numbers of students, is missing in the literature, especially in the context of university-level physics education. To start to address this research shortfall the positioning for this article is that such studies need to embrace the complexities of student thinking and application of knowledge. To achieve this, both factor and network analyses were used. Even though both approaches are grounded in different frameworks, for the task at hand, both approaches are useful for analyzing clustering dynamics within the responses of a large number of participants. Both also facilitate an exploration of how such clusters may relate to the semiotic resource formulation of a representation. The data were obtained from a questionnaire given to 1368 students drawn from 12 universities across 7 countries. The questionnaire deals with the refraction of light in introductory-level physics and involves asking students to give their best prediction of the relative visual positioning of images and objects in different semiotically constituted situations. The results of both approaches revealed no one-to-one relationship between a particular representational formulation and a particular cluster of student responses. The factor analysis used correct answer responses to reveal clusters that brought to the fore three different complexity levels in relation to representation formulation. The network analysis used all responses (correct and incorrect) to reveal three structural patterns. What is evident from the results of both analyses is that they confirm two broad conclusions that have emerged from social semiotic explorations dealing with representations in relation to attempting to optimize teaching and learning. The first, which is linked to a facilitating-awareness perspective, is that any given disciplinary visual representation can be expected to evoke a dispersed set of knowledge structures, which is referred to as their relevance structure. Thus, the network analysis results can be seen as presenting a unique starting point for studies aiming to identify such relevance structure. The second broad conclusion is that disciplinary visual representation can and often does contain more disciplinary-relevant aspects than what

*Corresponding author: priemer@physik.hu-berlin.de
†cedric.linder@physics.uu.se
‡jbruun@ind.ku.dk
§arvid.pohl@lnu.se

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.
I. INTRODUCTION

In physics education research (PER) situated at the university level, the concept of representational competence has increasingly been seen as a crucial aspect of learning physics [1–8]. Here, representational competence refers to students’ ability to appropriately and effectively interpret and use the various social semiotic formulations of representation used in physics.¹ These are representations that the physics community and the physics teaching community have established and used to communicate the ways of knowing and doing physics. By formulation of representation, we are referring to the ensemble of systematically used semiotic resources that get incorporated into the constitution of the representation. As mentioned earlier, semiotic resources that are systematically used are referred to as modes. These modes are meaning-making systems that have, over time, been shaped and developed in ways that have made them integral parts of the discursive practices of a social community. In our contemporary physics community, it is semiotic resources that are made up of verbal, written, mathematical, graphical, schematic, pictorial, and gestural communicative actions that typically make up the modes of our disciplinary discourse (see Fig. 1) [9]. Put another way, it is this “multimodality” [10] that facilitates the specialized and effective communicative practices that are used by the physics community to share meaning vis-a-vis both existing and new physics concepts, quantities, materials, artifacts, and processes.

Attaining representational competence in a disciplinary area such as physics occurs when the ensemble of modes—that is, the ensemble of systematically used semiotic resources—“work together” to afford particular meaning. By work together, we mean how these modes get used together in physics to share meaning, for instance, the diagrams, formulas, and written or spoken language that may be used to explain how an experiment was done. The point is that individual modes working on their own can only provide a part of access to the attributes and features of a representation that are needed for someone in physics to constitute the full intended meaning of the representation. Critically for the educational endeavor is recognition that some of the access given by individual modes may overlap with that of others and some of it may be unique to a particular mode for the task at hand (for physics examples, see Ref. [12]). These partial “accounts” directly affect what can be immediately noticed by a newcomer to what is being taught. The implication of this is that some critical aspects for the formation of the intended meaning of a representation will be immediately present while others will need to become appresent for the person seeking to constitute the intended meaning. What is appresent here refers to those aspects that are not directly observable but which the “seeing” of them must be learned so these can be experienced and thus coordinated alongside what is observable to constitute an intended meaning.²

For example, one form of representation depicting Snell’s law that is found in many textbooks is given in Fig. 3(a). Here, the change in direction by a ray of light when propagating from air to water is immediately present for most students. However, no indication of why this change of direction occurs is directly visible (for an explanation that needs to become part of students’ appresent awareness of this representation, see Ref. [14]). Teaching that facilitates students developing both their awareness of what is present and their appresent awareness will optimize the match between the intended object of learning and the experienced object of learning (for a discussion on the importance of optimizing this match, see Ref. [15]).

This simultaneous “seeing” of the present and appresent in ways that facilitate the constitution of intended meaning is a useful way to appreciate the relevance of achieving representational competence. Yet, despite the fact that such representational competence is critical for the effective learning of physics [16–20], many physics educators, who are themselves “fluent” in the use of disciplinary representations, may lack a deep appreciation of the learning hurdles that students face in the appropriate interpretation and use of such representations [4,21,22]. Research into this lack of appreciation by university physics educators of the difficulties their students face in this regard has been rare but that which has been undertaken is highly illuminating (see, e.g., [4,23,24]). There is, however, a fairly extensive number of studies that can be both explicitly and tacitly related to the idea of students’ representational competence located at the

¹What is meant by social semiotics here is a perspective that has been built on seeking to study and understand communication and meaning making in relation to particular social groups and settings.

²What is meant by the appresent derives from Husserl’s writings. It was incorporated into the educational perspective known as phenomenography (see Ref. [13]) and as such refers to educationally critical aspects that are not directly observable yet need to become part of what is discursively experienced.
level of university physics. The work in the area to date can be sorted into three distinct threads:

- how representations are interpreted (see, e.g., [14,25–28]);
- how representations can affect problem solving (see, e.g., see Refs. [17,18,29–31]); and
- how representations can be used to model learning in terms of becoming competent in the discursive practices of physics (see, e.g., [3,9,11,12,32]).

The study presented here is located in the third thread and builds on this. Airey and Linder theoretically propose in their 2009 article [11] that one of the necessary conditions for optimizing physics learning experiences calls for physics educators to see an essential part of their crafting of teaching practice [33] in terms of developing students’ discursive fluency in the disciplinary discourse of physics. See Fig. 1 which illustrates the relations among physics discourse, representations, and the systems of modes used (systematically used semiotic resources) to constitute these representations (after Airey & Linder [11]).

In what follows we outline the state of the field by focusing on

(A) how disciplinary-relevant aspects are embedded in representations;
(B) what formulations of representations make up the language of physics; and,
(C) how a certain learning task may evoke the choice and connection of relevant pieces of knowledge by students (their relevance structure).

In doing so, we will use terminology to refer to constructs that are essential for the underlying framing of our article. We do this because we draw on theoretical aspects that may not be familiar to all readers. For example, (A) will be described using disciplinary-relevant aspects and appresent awareness, (B) with formulations of representations, and (C) with relevance structure. This terminology is needed and thus needs to be as precise as possible when we outline the framing of our article. To enable further exploration of these constructs, we will introduce the basic terms and refer to sources of literature that give more details.

A. Disciplinary-relevant aspects and what is meant by the present and appresent aspects afforded by a representation

Marton and Booth [13] provide extensive authoritative detail about the anatomy of awareness perspective (phenomenography) that we draw on for our framing. Phenomenographically, they characterize the teaching
and learning challenges, such as what the physics education research literature has explicated for many years now, as follows: for any representation to be educationally valuable “the whole needs to be made more distinct, and the parts need to be found and then fitted into place, like a jigsaw puzzle that sits on the table half-finished inviting the passerby to discover more of the picture.” (p. 180)

Since our data involve visual representations and the qualitative solving of a physics problem in relation to an object of learning, a relevant theoretical framing is needed to work with how physics content is social semiotically embedded in a representation. What is particularly important here is how disciplinary represented phenomena, particularly more complex phenomena, have both critical aspects that are directly visible and aspects that are “invisible.” In other words, these aspects are not directly perceivable from the representation, they lie behind what is visible, and they need to become appresent in the learning experience. In other words, the representation requires an appresent dynamic, which physics experts will have developed but not novices. This is an integral part of an effective learning process. This means that physics education will frequently use representations that do not, and cannot, present a direct visual awareness of all the aspects needed to make an intended meaning. These aspects can only become part of a student’s appresent awareness through the use of a particular set of supplementary representations that can provide access to them (see Fig. 2); what is initially invisible to students needs to become appresent to them so that the present and appresent can be simultaneously used to access the whole meaning. In the social-semiotic PER literature, the critical-for-learning aspects needed for a given object of learning are referred to as disciplinary-relevant aspects (DRAs) [12,14]. Awareness of DRAs from any representation will be embedded in both the present and appresent, and as such “have particular relevance for carrying out a specific [physics] task” [14] (p. 2.) (two good illustrative examples of such appresent complexity in physics meaning-making can be found in Refs. [34,35]).

As part of an object of learning to deal with the refraction of light, the ray diagram is an integral representation; it is known and used widely in introductory physics contexts. As such, the ray diagram has become both iconic and ubiquitous for depicting the refraction of light as it propagates between two different optic media, apparent depth, and as an introduction to Snell’s law [see Fig. 3(a)]. To make meaning from the likes of the ray diagram given in Fig. 3(a) calls for having developed both present and appresent awareness. For example, it has important presented aspects such as the boundary between media, a light beam, different angles, and important appresented aspects such as the wavefront, the wave normal represented by a light beam, the velocity of light in different media, and the reversibility of the optical path. In the questionnaire given to students in our study, the needed appresent aspects are embedded in different representation formulations (see Sec. II B). Students’ responses are taken to be an indication of their relevance structure (Sec. II C will give more details for this) for the image-object tasks presented in the questionnaire, which relates to their ability to appropriately constitute a whole from what a particular representation formulation can present and what appresent awareness they have developed for each formulation use. These include aspects such as the bending of light rays; the amount of refraction that takes place as a function of the ratio of indices of refraction; and the predictive insights that Snell’s law offers.
B. Formulations of representation: Multimodality and semiotic resources

An essential competency that a physics student will need to develop is the capacity to appropriately understand and work with the social semiotics of physics, that is, the disciplinary communicative formulations that physics as a discipline has developed and uses ([5,36–38]). This, for example, is well characterized by the learning schema presented by McDermott and Shaffer [39] and Etkina et al. [40]. Developing this competency is an essential part of being able to become aware of the DRAs needed for a particular object of learning [1–4,9,12,16,21,41].

The essential role that representations play in both teaching and learning has also long been explored outside of PER. For example in chemistry education research (CER), see Ref. [42] and in biology education research (BER), see Ref. [43]. In PER, in contrast to the work in CER and BER, contemporary work has begun to both implicitly and explicitly use and develop a social semiotic theory to frame the work (see, for instance, Refs. [9,16] and the Ph.D. dissertations of Brookes [44], Kohl [37], Podelefsky [45], Airey [46], Eriksson [47], Fredlund [21], Weliweriya [48], and Volkwyn [49]).

The discursive modeling in social semiotics is built on how all communication is made possible through socially constructed representations, which draw on particular semiotic resources. When used systematically in communication in physics, these become particular modal forms that get used to construct the illustrations, symbols, and language of the physics and physics education communities [10,50–52]. Thus, the representation formulations typically used in physics are seen as the resources that make particular threads of disciplinary discourse possible (see Fig. 1). In PER, there is existing work that highlights the kinds of ensembles of semiotic resources that students need to be able to competently use and need to experience in order to optimize the possibility of appropriately and productively perceive the intended object of learning (for example, see Refs. [18,39,53]). This has been relatively recently characterized in terms of students needing to acquire representational competence, for example, see Refs. [4,6]. Put another way, students need to be able to engage meaningfully with the educationally relevant parts of the disciplinary discourse of physics [11] for a particular segment of the physics curriculum (see Fig. 2). Lemke [54] (p. 7) provides an exemplary illustration of the competency it takes to work appropriately and effectively within the disciplinary discourse of physics: “We can partly talk our way through a scientific event or problem in purely verbal conceptual terms, and then we can partly make sense of what is happening by combining our discourse with the drawing and interpretation of visual diagrams and graphs and other representations, and we can integrate both of these with mathematical formulas and algebraic derivations as well as quantitative calculations, and finally we can integrate all of these with actual experimental procedures and operations.” Lemke’s description illustrates how, in physics, different semiotic resources typically need to be communicatively integrated as a function of representational competence.

While acquiring such representational competence is a necessary part of effective meaning making (learning) in physics education, it is not sufficient. Thus, a theoretical model has been proposed [9,11,12,32] where students need to become “fluent in a critical constellation of semiotic resources” (modes of the disciplinary discourse of physics,
see Ref. [11], p. 2) before they can effectively experience an object of learning as intended. In this proposed model, an object of learning consists of several critical aspects. To provide access to all the needed aspects, a teacher would need to draw on a critical constellation of modes in a way that enabled their students to experience the collective disciplinary affordance needed for them to constitute the intended meaning that made up a particular object of learning.

For the concept of refraction, direction, medium, and speed of light in each of the media (also often referred to as the phase velocities of light in each of the media) are three of the DRAs need for understanding how and “why the refraction of light takes place when light propagates between two media with different refractive indices” [14] (p. 4). From a teaching perspective, insight into the appresent aspects of the mathematical formulation of Snell’s law would make up a part of the critical constellation of modalities needed for their intended meaning making of the mathematical relationship between the angles of incidence and refraction when light travels through a boundary between two different isotropic media. At the same time, for some learning tasks that some teachers might give students in connection with refraction, such as strictly learning to calculate the magnitude and direction, mathematics may be the only relevant mode, where, for instance, \( n_1 \sin \theta_1 = n_2 \sin \theta_2 \) is a possible modal formulation of representation. For example, in Fig. 2, being able to appropriately insert numbers into Snell’s law in the form just given, would in itself give (very) limited access to two aspects; direction (represented mathematically by angles) and medium (represented by refractive indices). The access is limited because it is strictly computational and does not involve how or why refraction happens. Another educational task for students could be to investigate the variation of refraction angles with respect to wavelength (thus, adding “wavelength” as well as experimental equipment and procedures as DRAs). This could involve working simultaneously with two modes: an experimental setup as well as formulating results and understandings in written and spoken language. This work could, in turn, semiotically provide combined access to DRAs such as direction, medium, and wavelength in a way that allows students to able to explain how a prism separates white light in colors. This simultaneous work of different modes is illustrated by the two arrows pointing at the south-eastern corner in Fig. 2. The task of getting to see why a ray changes direction most likely requires teaching and learning tasks that combine diagrammatic formulations with written, and spoken language (for an example, see Ref. [14]). The point being emphasized here is that while a single representational form such as language might be able to afford access to several of the aspects needed, it will not be able to provide access to all aspects needed. To achieve this, language will need to be insightfully used alongside other modes to provide access to the other needed aspects. Hence, an important dynamic of effective teaching calls for appreciating that different representation formulations cannot present equivalent meaning-making possibilities to students (for an extensive exploration of this and implications for teaching physics, see Ref. [14]). Regardless, across all parts of the illustrative example just given, the direction, medium, and speed of light in each of the media remain disciplinary-relevant aspects (DRAs). The hypothetical model depicted in Fig. 2 underpins the consequent proposal that a critical constellation of modes is required to provide access to the needed collective disciplinary affordance of the modes making up the representations used for the educational task at hand. In direct relation to the aim of our article, which involves empirically exploring this proposal with a large dataset, our study draws on schematic, pictorial, diagrammatic, and written language modes that constitute the different formulations of the representation that make up the refraction questionnaire that we used for our data collection. (This questionnaire can be found in the Supplemental Material section [55].)

C. Relevance structure

For any physics education task, students will spontaneously respond to the contents of the task as a function of their perception of these in terms of what they feel is relevant to the task at hand [56]. For example, when presented with an object submerged in water and asked to predict where, from a situated point of view, the image and object are to be located, what gets taken to be relevant to make the asked-for prediction is solely that the image is lifted in relation to the object, another may be that it is lifted and shifted toward the situated point of view, and another may be that a ray diagram such as that given in Fig. 3(b) needs to be carefully constructed, see, for example, Ref. [57]. Drawing on phenomenology and Székely’s work [56], Marton and Booth [13] (p. 143) characterized such contextual responses as representing the relevance structure for that person for that given task—what the perception of a situation “calls for, what it demands.” Marton and Booth portray relevance structure as an integral component of the phenomenographic model of learning that they present: that is, becoming “capable of being simultaneously and focally aware of other aspects or more aspects of a phenomenon than was previously the case” [13] (p. 142, emphasis added). The form and structure of relevance structure(s) evoked by an educational task then becomes a highly valuable educational consideration in the crafting of effective teaching practice. This is because the educational aim behind such crafting of practice becomes principally about developing capabilities that are essential for making the intended meaning making (learning) possible: “[…] people act not in relation to situations as such, but in relation to situations as they perceive, experience, and understand them. […] If we want learners to develop certain capabilities, we must make it possible for them to develop a certain way of seeing or experiencing.
Consequently, arranging for learning implies arranging for developing learners’ ways of seeing or experiencing […]” [15] (p. 8, emphasis added). Relevance structure is an important complementary framing feature for our study because, in effect, our analysis centered on looking for identifiable relationship(s) between relevance structure and the modality of the situations presented. To elaborate on this point, we return to the ray diagram given in Fig. 3. In order to predict image and object positions, our questionnaire called for students to generate a personal visual extrapolation of light ray paths for propagation from one medium to another. A starting assumption was that different formulations of representation could evoke qualitatively different relevance structures and that the analysis would reveal important attributes of these relevance structures.

III. RESEARCH QUESTIONS

As outlined in the previous section, it is so far only theoretically clear precisely what role the modal formulation of representations may play in relation to making the intended meaning making possible for particular physics objects of learning. The applicable social semiotic theory is that different modes have different possible affordances. And the communicative sharing of a DRA almost always needs more affordance than any mode can provide on its own. This means, as illustrated in Fig. 1, that combinations of different modes are needed to semiotically work together to provide the collective affordance that is needed. Thus, using this theoretical lens, to facilitate intended meaning making, a particular object of learning will require a particular set of representations that are constituted from a particular set of modes (i.e., a critical constellation of representations is needed). For the case of refraction, the ray diagram [Fig. 3(a)] is the visual representation that is most commonly used in lectures and textbooks. However, there are other formulations that may provide a different affordance (see Sec. II B) allowing other DRAs to become part of students’ apperceive awareness (see Sec. II A), and which may evoke a particular relevance structure when embedded in a particular problem (see Sec. II C). The aim of the research, as reflected in the research questions, is to use a large sample to empirically explore this theoretical background.

The research questions are aimed at exploring the above theory by using the object of learning area introductory refraction and a questionnaire that contains modal formulations of representation that consist of a variety of visual image-object projections:

Research question 1 (RQ1): What relational factors between ways of answering and modal formulations of representation of the visual representation used to present physics information can be identified in physics problems (using our questionnaire questions as a particular example)?

Research question 2 (RQ2): What kinds of relational structures can be identified within the ways of answering data when analyzed using two different cluster analysis approaches? In what ways are the sets of clusters gained by the two methods related?

The approach to addressing the research aim was to ground it in looking at how responses to different items for one object of learning can be seen to be clustered in a way that takes into account that there is complexity in the data. To not rely on only one specific research approach and run the risk of being too specific, we used two approaches. These are described next.

IV. RESEARCH DESIGN AND METHODS

A. The physics concept and methods of assessment

The refraction of light was chosen as the target object of learning because (i) it is typically taught in introductory-level university physics courses across the world, (ii) the ray diagram is the most commonly presented visual representation used in physics textbooks, (iii) it lends itself to the design of a multimodal questionnaire that deals with the same DRA, and (iv) it is considered to be of general interest to physics educators.

Refraction is easily observable in everyday phenomena, in demonstrations, and in student-laboratory experiments. These experiences can be drawn on to generate a rich array of visual representations for physics educators and texts to illustrate the DRAs of refraction (including, “direction,” “medium,” and “speed of light”). And it is common for university physics educators to draw extensively on the representations provided in the texts that they refer their students to Ref. [58]. However, only a few of these possibilities get used in most textbooks [59]. As shown in Fig. 3(a), the most prominent representation used in introductory-level physics texts is the ray diagram, most typically with a light ray originating from the upper left-hand corner of the diagram striking a horizontal media boundary between air and water where after the light ray is bent toward the normal [59]. This one, very often used visual representation, could give the impression that it is sufficient to make the intended meaning of refraction of light (see RQ1).

Our research aims led us to prefer a questionnaire as our way of capturing students’ use of representational formulations to provide answers to tasks about aspects of refraction. First, we were interested in a large-scale study that could probe student understandings in different educational contexts. Second, the questionnaire format allowed for the construction of tasks that could (a) be described adequately with our theoretical framework and (b) relate student responses to particular visual representational formulations. To this end, a questionnaire was developed by Hüttebräuker [59] in order to assess the students’ choice-of-answer for the questions posed. These were
embedded in the refraction of light, specifically with respect to the positioning of images and objects in using different representational settings. For the representational form constitution of the questionnaire, a comprehensive review of 141 (48 high school level and 93 university level) physics textbooks was undertaken in order to establish the most commonly used range of representations drawn on to present refraction vis-à-vis the visual situating of image and object.

Based on this, a classification scheme was drawn up and the representations were sorted according to their variation and frequency of use. This resulted in ten categories being constituted [59]. This sorting of representations also led to the finding that two different and contrasting approaches were used in the textbooks—first, the more prevalent light-ray model [found in over 70% of the textbooks, see Fig. 3(a)] and second, a pictorial and schematic representation that illustrated the phenomenon of apparent depth [see Fig. 3(b)]. Both these representational formulations were used to develop the items posed in the questionnaire. The trialing of the questionnaire was undertaken with both students (as novices) and professors of physics (as experts). Following this, the questionnaire was refined prior to further use.

The final version of the questionnaire comprised 13 items grouped into 3 tasks. The four items in task 1 (T1Q1–T1Q4 in the results) involve consideration of a refracted light ray. The essential difference between tasks 2 and 3 is the perspective from which the problem is presented. In the five items of task 2 (T2a1–T2a4 and T2b), the accompanying diagrams show a side view of an object and a viewing eye—this allows the student to draw in the position of a refracted ray. In contrast, the accompanying figures to the four items in task 3 (T3a, T3b, T3c, and T3d) are given from the eye’s view, i.e., a situation is shown as it would be perceived by a viewing eye. Consequently, the viewing eye is not in the picture, and the student cannot draw a refracted ray from the object to the viewing eye.

Each question requires the same DRAs (primarily “direction of ray/rays” and “medium”) to answer and can be said to present variations of the same conceptual problem: the positions of images and objects in relation to how light is refracted at media boundaries. However, from our earlier theoretical discussion, it may be anticipated that the different formulations of representations used in the three tasks provided different collective affordances in relation to the situatedness of the images and objects arising from refraction (see Fig. 2).

Questionnaire task 1 comprises a single instance of representation formulation (the ray diagram), whereas tasks 2 and 3 include different formulations. A further difference is that in tasks 2 through 3b, the respondents were asked to choose an answer(s) from a list of options. The questionnaire was translated from English into Swedish, German, and Portuguese and written in the language of instruction prevalent at the participating university.

For the factor analysis (RQ1), two items were not included, T2a2 and T2a4. This was because they turned out to be too similar to two other items, T2a1 and T2a3, to provide new factor analysis information. On the other hand, they were included in the module analysis for multiple-choice responses (MAMCR) (RQ2). This is because the criteria for item choice inclusion and exclusion were different for our implementation of MAMCR. Here, item choices, not items, were discarded based on their frequency (see our description of MAMCR). Internal consistency of the test produced an acceptable Cronbach’s alpha value of 0.7 (see Ref. [60]) and histograms gave a good fit with a normal distribution. However, Kolmogorov-Smirnov and the Shapiro-Wilk tests were also performed and they revealed that the total test scores deviate from a normal distribution, a result that can be expected given the large sample size [60].

As with any such research instrument, the questionnaire employed in this study has its limitations. When it comes to test construction, given the large number of textbooks surveyed not all representational formulations could be included (we chose the most commonly used) and of those that were used, some may have been unfamiliar to some of the respondents. In several instances, instead of choosing one of the options provided, respondents chose to draw their own constructions. These were subsequently excluded from the coding and analysis we did for this article. In a number of the multiple-choice items, more than one answer could arguably be considered correct. In such cases, the respondents were told that they were free to select more than one answer (the selection of any of these correct answers led to coding the item as correctly answered, denoted, for example, as “no wrong choice”). However, it is then possible that this instruction on the questionnaire could have been misinterpreted by some participants. Finally, whereas clear instructions were provided, it is possible that minor variations in questionnaire administration could have occurred.

B. Participant sample

The questionnaire was administered to 1368 students at 12 universities located in 7 countries: Australia, Brazil, Germany, the Philippines, South Africa, Sweden, and the United States. The participating students were enrolled in a range of bachelor programs in science-related fields, including physics, biology, biotechnology, civil engineering, geology, life science, and mathematics. All participants were enrolled in a physics course in which refraction had been a topic of instruction prior to their answering the questionnaire. In general, the respondents took around 30 min to complete the questionnaire.
C. Analyzing the data

Factor analysis—our first approach—is a well-known technique to identify clusters of variables. Given a large sample, a successive combination of, first, exploring half of a data set with extracting factors and, second, testing the hypotheses about the explored structure with the other half of the dataset can be used, for example, see Ref. [61]. We used factor analysis to address RQ1.

While both exploratory and confirmatory factor analysis are standard calculations in quantitative empirical work, module analysis for multiple choice responses (MAMCR) [62]—our second approach—is not, and consequently, its use here may not be familiar to many readers. The analysis generates clusters (referred to as modules in MAMCR) of related answers, where, in this instance, the open-ended tasks of the questionnaire were coded into disjunct item choices. MAMCR was used to address RQ2.

MAMCR was developed to capture the complexity of student selection patterns in multiple-choice instruments. It was applied to student answers to the Force Concept Inventory (FCI) and [62] showed residual non-Newtonian structures postteaching—even if student FCI scores were generally high. Not all structures in the analysis had straightforward interpretations in line with the original intentions of the instrument, and complex relations between structures were found. A further development, called modified module analysis (MMA), aimed to reduce the complexity of the analyzed student answering patterns and was able to identify simple components of significantly correlated answers [63–65].

Although a detailed description and illustration of the application of MAMCR are provided in Ref. [62], because the methodology is relatively new to PER, the basic principles as well as our extensions to the method are outlined below.

1. Procedure for using module analysis for multiple-choice responses

Whereas the methodology was originally developed for multiple-choice items, it can be used for virtually all items that have well-defined categories of “response” [62]. The analysis generates clusters (referred to as modules in MAMCR) of related answers, where, in this instance, the open-ended tasks of the questionnaire were coded into disjunct item choices.

A network is a collection of entities—often referred to as nodes—that are connected via links. In this work, a link always connects exactly two different nodes. The first step in MAMCR is to prepare data for network analysis. An $M \times N$ matrix $A$ is constructed where the $M$ participants are listed in rows and the $N$ item choices in columns. If student $i$ chose item choice $j$, then the corresponding matrix element would be $A_{ij} = 1$, otherwise $A_{ij} = 0$.

While it is beyond the scope of this paper to describe MAMCR in detail (for further details, see Brewe et al. [62]), it is worth noting that the analyses were done using R [66] and the igraph™ package [67], while visualizations were done using a combination of Gephi™ [68] and a vector graphics editor.

The important steps of the analysis are described below.

2. Linking students to item choices: The bipartite network

In this step, all participants and a subset of the item choices are used to create a network with two kinds of nodes: student nodes and item choice nodes. No two student nodes and no two-item choice nodes are directly connected in this network (bipartite network). In Brewe et al.’s [62] application of MAMCR, they reported that it was necessary to remove all correct item choices to get a subset of modules that could be interpreted. In their case, the correct item choices were the most chosen answers and thus linked to a large number of participants. They become a central tightly knit module that ends up obscuring the more fine-grained network relations that exist between the item choices. So, a subset of answers must be found which limits the number of item choices that are used in the analysis. For our analysis, the methodology was extended by including a procedure that finds an optimal threshold value for the prevalence of item choices that are included in the analysis. More details of this procedure are given below.

3. Focusing on the answers: The item choice network

Next, the bipartite network is collapsed into an item choice network. In this network, two item choices are connected if at least one student has chosen both of them. In this way, all item choices selected by student $i$ are connected. The weight of the link between two item choices increases by one with every student found to have chosen these two answers. This results in a network that reveals the many connections between item choices. However, the high number of connections is very likely to lead to only one module, and thus such resultant networks are unlikely to be fruitful for finding modules.

4. Selecting relevant links: The backbone network

Since the kind of network outcome just described is unlikely to be fruitful for finding modules, the next step is to remove links that can be considered insignificant or random. To do this, the local adaptive network sparsification (LANS) criterion of Foti et al. [69] is used. In this procedure, the links to remove are determined based on an analysis of the weight of links for each node. The procedure is as follows (for mathematical details, see Ref. [69]): For node $A$, the weight of each link is compared with the weight of all links of node $A$. The weight of each link is higher or equal to the weight of $n$ of $A$’s links. If $A$ has $N_A$ links, this corresponds to a fraction, $\hat{F}(w) = n/N_A$ of $A$’s links for a given weight, $w$. For example, assume that node $A$ has 20 links with one link with weight 3, one with weight 2, and
the rest weight 1. Then, \( \hat{F}(3) = 1, \hat{F}(2) = 19/20 = 0.95, \)
and \( \hat{F}(1) = 18/20 = 0.9. \) The level at which a link is taken to be significant is defined as \( \alpha(w) = 1 - \hat{F}(w). \) Links with weights that satisfy \( \alpha(w) \leq \alpha, \) where \( \alpha \) is a predetermined significance level, are retained in the network. Links for which \( \alpha(w) > \alpha \) are discarded. For example, in the backbone network with \( \alpha = 0.05, \) the links that are retained from the perspective of A have weights that are larger than or equal to 95% of the links of A in the item choice network. In the example above, \( \alpha = 0.05 \) would remove the links with weights 1, because \( \alpha(1) = 0.1 > 0.05 = \alpha. \) Links with weights that does not satisfy the condition \( \alpha(w) \leq \alpha \) are considered unlikely to “express” any meaningful connection between item choices. As with Brewe et al. [62], it was decided to keep all links that were significant for at least one node; but unlike them, we do not claim that this “guarantees a connected network” [62] (pp. 020131–020135). While it is true that by construction, a node will always have at least one link that is locally significant, two nodes could have significant links to each other and to no other nodes. Thus, they represent an island of two nodes.

5. Finding modules

The next step of MAMCR is to let the Infomap algorithm [70] partition the network into modules. Here, modules are subsets of nodes with internal structure. Brewe et al.’s original work explains how the algorithm connects to an analysis of multiple-choice questions, see Ref. [62] (Sec. II. G). Particularly, they argue that using Infomap “can be seen as a simulation of how the cohort answers” a questionnaire [62] (p. 6).

Infomap relies on the concept of a random walker. This is an entity that can travel from node to node in a network using links. When arriving at a node, the walker will choose which link to use at random with probabilities proportional to link weights. After a long time, the random walker will have visited each node with a particular frequency, called the node visit frequency. The fundamental idea behind Infomap is to convey information about the random walker’s walk in the most efficient way. If the intention is not to try to find modular structure, the most efficient way is through a procedure called Huffman coding [71]. Here, a node is assigned a code word based on the corresponding node visit frequency. A high node visit frequency would result in a short code word. Now, any part of a random walker’s walk can be conveyed by making a text message listing the code words of the nodes in the order they were visited.

It is possible to reduce the length of this imaginary text message if subsets of nodes are more tightly linked with each other than with other subsets of nodes. This property can be called modular structure [72], and for the purposes of Infomap, subsets of nodes are therefore called modules. For networks with modular structure, the random walker is likely to spend more time in each module of the network before leaving and spending time in another module. The text message now uses a code word to signify when the random walker is visiting a particular module and then uses a set of code words for the nodes in that module. When the walker shifts to another module, the imaginary text message conveys that the walker has left one module and is now traversing a different module. However, when this happens, code words can be reused allowing for shorter code words in each module. For instance, in module X, one node may have the code word “01.” If the imaginary text message relays the information that the walker has shifted to module Y, then a different node can have the code word “01.” To relay information about shifting between modules, modules X and Y also need to have code words. Without the division into modules X and Y, the two nodes would need to have different code words. For networks with a modular structure, the strategy reduces the length of the imaginary text message [70], because code words within modules would be used much more often than module code names. Infomap’s purpose is to find the division of the network into a set of modules (the set could be just one module, which would be the case if Infomap cannot find a modular structure), that minimizes the length of the imaginary message. In other words, Infomap finds the partition of the network into modules that minimize the theoretical minimum length of code needed to describe a finite random walk on the network.

The result is a set of modules, a module partitioning, in which nodes share more links internally than they do with nodes in other modules. In the special case of only one module, Infomap was not able to detect modular structure. The modular structure is most often gauged with the modularity measure \( Q \) which ranges from \(-0.5 \) to \( 1, \) with \( 1 \) being perfect modularity. In the network literature, a network is said to have a significant modular structure if \( Q > 0.3 \) [73].

6. Consistency of module partitioning—Module co-occurrence matrix

Like many other contemporary algorithms for finding modular structure in networks, Infomap is probabilistic by nature. Even if Infomap is one of the most robust algorithms [74], each run of the algorithm may produce slightly different results. To visualize the robustness Infomap results, for each network to be analyzed, Infomap was run \( n_{iter} = 1000 \) times, each run resulting in a module partitioning. Then, an \( N \times N \) module co-occurrence matrix \( M \) was constructed where \( N \) is the number of item choices analyzed [62]. Each row and column represent one item choice. Each matrix element \( M_{kj} \) was calculated as the fraction of times item choice \( k \) was assigned to the same module as item choice \( l. \) Since \( M_{kl} = M_{lk}, \) the module co-occurrence matrix \( M \) is symmetric and captures the consistency of Infomap module partitionings.
To visualize the consistency of this partitioning, $M$ was plotted as a heat map (again, following Ref. [62]): In this visualization, the matrix rows and columns were ordered with respect to the most frequent module partitioning: The first $m_1$ rows and columns (starting from the left downmost corner of $M$) contained the elements calculated for item choices in the first module found by Infomap. The next $m_2$ rows contained the entries for the second module found by Infomap and so on. To complete the heat-map visualization, each entry was converted into a color on a gradient from blue ($M_{kl} = 0$) to red ($M_{kl} = 1$). For item choices grouped in the same module 100% of the time, the heat map will display red squares spanning the row and columns denoting those item choices.

In this study, the strategy for analyses was to interpret modules as they appeared in the most frequent module partitioning identified by Infomap. This strategy has merit, if Infomap finds the same partition every time or a large fraction of the time. The module co-occurrence matrix for an item choice network is a detailed visualization of the consistency of the most frequent partitioning.

7. Interpreting the modules

We follow Brewe et al. [62] and interpret the application of Infomap on networks based on student responses to multiple-choice questionnaires as a simulation of students filling out the questionnaire. The interpretation rests on an analogy between the random walker and a student; the random walker produces frequent student-answer patterns because it follows links between item choices in the network. In this interpretation of Infomap, item choices are connected with other item choices in a nonlinear manner. Hence, it is not assumed that the random walker portrays a student who starts at the beginning of the questionnaire and continues through until the end. Rather, the random walker portrays frequent answering patterns, which are then represented as modules.

An integral part of Infomap is that it uses the nonlinear connections to calculate the percentage of times that a node gets visited—a calculation which in turn is used to find modules (see Sec. IV C 5). Node visit frequencies represent the fraction of time the random walker spends at a given node. Thus, the sum of node visit frequencies for a module becomes the percentage of time that the module is visited. When interpreting a modular solution, the modules with large summed node visit frequencies represent most of the information about the network, and within a given module, item choices with the highest node visit frequencies represent most of the relevant information about the module. Furthermore, the random walker is likely to make more use of links with high weights than links with low weights. These considerations are arguments for focusing first on modules with large sums of node visit frequencies and within these modules on nodes with high node visit frequencies and on connections with high weight.

However, further analyses of modules may reveal that many nodes with smaller node visit frequencies are present and have structurally interesting roles. This analysis may shape the interpretations of modules. Thus, any analysis needs to strike a balance between considering large and smaller node visit frequencies and connections with high and low weights.

8. Extending MAMCR

For our study, we extended MAMCR in two ways. First, when creating a backbone network using LANS, we viewed the $\alpha$ level as a parameter to be optimized. A lower $\alpha$ level will tend to result in a lower amount of retained links overall in the network, and this in turn affects the modular structure of the network. Therefore, we developed a set of criteria (see below), which helped us choose modular solutions from the results of many iterations of running LANS on the item choice network. Second, we addressed an issue in the original MAMCR study, where the correct item choices were excluded, because they were much more frequently chosen than incorrect choices [62]. We extended MAMCR to generate good interpretative results even when the correct item choices are not necessarily the ones most frequently chosen. In doing so, an appreciation was needed that some item choices could be very popular and that these could effectively obscure the identification of fine-grained network relationships. The task then became one of finding a threshold frequency for which popular answers could be meaningfully excluded. This challenge was resolved by running a second iterative cycle. In other words, our extension of MAMCR consisted of two iterative cycles.

The first cycle helped to find an optimal $\alpha$ level by extracting backbone networks for $\alpha$ values between 0 and 0.1 in steps of 0.0001 and then running Infomap. For the lowest $\alpha$ value, this procedure resulted in a backbone network which retained links that were locally significant at the 0.01% level. Each cycle resulted in 1000 backbone networks which retained links that were locally significant at successively higher percentage levels. For each backbone network in a cycle, Infomap yielded an assignment of every node and the modularity of the backbone network. An acceptable modular structure according to a rule of thumb in network literature [73] is given for $Q > 0.3$; [62] reached $Q = 0.39$. We argue that a meaningful module solution should meet the following criteria: (i) the solution should extract different modules (and not only one or one that dominates other modules), and (ii) the solution should contain a small fraction of disconnected modules, say modules with two or three nodes, which have no further connections. The reasoning behind this is as follows: in a network where all or most of the item choices are allocated to one module, then no, or only very little, new information is gained by the partitioning into modules. On the other hand, if the solution contains a lot of isolated islands, even
if the original item choice network is heavily connected, then the solution has not retained valuable information about how responses are related to each other. For such solutions, the grain size of the analysis becomes too small to obtain a holistic picture of the relationship between item choices. Thus, a good analytic balance needs to be found between module coherency, the total number of modules, the size of the largest module, and the fraction of duos and trios.

Figure 4 provides an illustration of how we reached this balance. First (top-left panel in Fig. 4), we consider only solutions with $Q > 0.3$. This excludes $\alpha > 0.04$. Second (top-right panel in Fig. 4), we limited our search to solutions with fewer than half of the nodes grouped into isolated dyads (two-item choices) or triads (three-item choices). To do this, we calculated this fraction, $f_{dt}$, for each $\alpha$ level, and look for $\alpha$ values with $f_{dt} < 0.5$. In Fig. 4, we see that this corresponds to $\alpha$ values above roughly 0.03. Finally, we checked the number of modules was manageable ($N_{\text{modules}} \approx 10$ based on Ref. [62]; see the bottom-left panel in Fig. 4) and the largest module did not contain the majority of item choices (bottom-right panel in Fig. 4). Given these constraints, we often had a range of possible $\alpha$ values. In case of differences in parameters within this range, the solution with the highest modularity $Q$ was preferred. The results showed plateaus of solutions that are stable over a range of $\alpha$ values (Fig. 4), and around transitions, there are fluctuations. We regarded plateaus as indicative of a stable range of solutions and preferably chose the solution from that range.

The second cycle addresses the selection of item choices that are included in the network analysis. Following Brewe et al. [62], very prevalent item choices were removed. In their work, correct answers turned out to be the most prevalent; however, this cannot be expected in general. In general, item choices that have been selected numerous times are likely to be connected with many other item choices. In so doing, they become a hub thereby obscuring more fine-grained patterns. Hence, the resulting networks show little or no modular structure and add no new information. To obtain a meaningful modular structure, the strategy is to remove item choices, which have been selected above a certain threshold frequency. For example, the item choice T1Q13right in our questionnaire was selected by 81% of the participants (i.e., specified as a frequency of 81%). Hence, a threshold frequency of 80% would remove that item choice from the analysis. Thus, the crucial problem was to find a threshold frequency that would lead to an item choice set with a modular structure that could be interpreted in a meaningful way.

V. RESULTS

A. Approach 1: Factor analysis

1. Procedure of using exploratory and confirmatory factor analysis

Factor analysis was performed as part of our empirical search for related questionnaire items. The analysis was followed by an attempt to interpret the extracted factors in terms of explicit relations (all calculations were done
using R [66]). Thus, this method empirically searches for related items and clusters them into factors. To develop a factor structure, the dataset of participants (with no missing data, \(N = 1220\)) was split randomly into two halves. One half \((N = 610)\) was used to generate a factor structure using exploratory factor analysis (EFA). The other half \((N = 610)\) was used to confirm the results of the EFA with a confirmatory factor analysis (CFA). This analytic approach is widely used to explore connecting theory building to theory testing without the need for a new sample (see, e.g., Van Prooijen and Van Der Kloot [61]). An EFA was applied by running a principle component analysis with VARIMAX rotation that extracted three factors (Table I); all items were included; because according to the Kaiser-Meyer-Olkin criterion, the sample was factorable. A parallel analysis with a scree plot showed that the three-factor solution has the best fit (all loadings were >0.4 and were thus included).

### 2. Factor interpretation

We then used the hypothesized factor structure of the EFA as a factor structure to be confirmed by a CFA with our created subsample, using diagonally weighted least squares because of nominal data [75]. The chi-square test \(\chi^2(41) = 47.146, p = 0.236\) shows that there is no significant deviation between the theoretical, model-implied covariance matrix and the sample covariance-matrix. This confirms the three-factor structure as it is shown in Table I (all factor loadings were significant; the goodness-of-fit parameters are comparative fit index = 0.99; Tucker-Lewis Index = 0.99; root mean square error of approximation = 0.016) and that for this dataset, no one-to-one relation between clusters of items and the formulations of representation emerged.

Table I shows the results of the factor analysis (the assignment of the items to the factors, the factor loadings of the EFA and the CFA, item difficulties, and a brief description of the interpretation for each factor). Three factors were extracted (items are shown in the Supplemental Material [55]):

**Factor 1:** These are the four items in task 1 and involve ray diagrams similar in structure to the common textbook representations—where the change of direction of a laser beam or a focused light ray is shown (in a cross section) as it penetrates, typically, a water surface [see Fig. 3(a)]. This representation is used to make predictions about the refracted direction of a light ray and to qualitatively illustrate the application of Snell’s law; none of which requires a deeper understanding of the physics content.

**Factor 2:** Contrary to factor 1, factor 2 consists of items from tasks 2 and 3, which go beyond the standard ray diagram. In the first item, a person is shown aiming a spear at a fish in a pond; in the next item, the spear is replaced by a laser; the third item shows photographs of a coin placed in a mug filled with varying amounts of water; and the final item shows a cross section of an aquarium in which the image of a coin has to be predicted.

**Factor 3:** This factor also consists of a mix of items from tasks 2 and 3. One item is a photograph where participating students had to predict how a pencil looks when partly submerged in water. The two other items depict a cross section of an aquarium in which the image of a pencil has to be predicted. Predicting how a pencil looks when partly submerged in water was challenging, even though it is not
FIG. 5. A pencil partly submerged in water seen from the side. The lower part of the pencil appears to be shifted to the left when compared to the upper part of the pencil.

abstract and can easily be observed in experiments and demonstrations (for example, see Fig. 5). Thus, overall, the items in factor 3 presented a greater degree of semiotic difficulty than those in factor 2.

The extracted factors revealed no consistently identifiable links to the representation formulations (i.e., the ensembles of semiotic resources) used to present what is essentially the same conceptual task (RQ1). However, the factors could be interpreted with respect to the physical situation (RQ2). Factor 1 describes variations of the “typical” ray diagram. Factor 2 contains items in which the relations between pointlike objects, their images, and an observer have to be analyzed, and in factor 3, the image of an extended object has to be predicted.

This interpretation of the results is in line with the work of Wiesner [57] who recommends, when teaching refraction, drawing visual attention to the lifting, bending, and shifting appearances of objects’ images underwater. Factor 2 corresponds to lifting and factor 3 to bending. From this perspective, the framing for a fourth factor can be considered, which has no counterpart in our questionnaire: shifting (see Fig. 5). Shifting describes the effect when an extended object—for example, a long stick—is placed in a transparent water container in a way that its lower part is below the water level and its upper part is above the water level. When an observer looks at the object with an angle that is different from perpendicular to the water container, they observe that the lower part of the stick is shifted sideways. This effect can be explained with lifting and a rotation of the observer by 90°. Another difference to Wiesner’s classification [57] is that our factor 1 describes the phenomenon of a light ray penetrating an air-water surface, which is not included in Wiesner’s work [57].

The three factors have a different average difficulty, with items grouped in factor 1 (average difficulty: 0.70) being easier than items in factor 2 (0.45) which in turn are easier than items in factor 3 (0.37). This may be due to the different levels of complexity needed to solve the corresponding items. Items in factor 1 are fairly easy to solve because they only require the correct manipulation of light rays as per Snell’s law. In contrast, in factor 2 items, students have to identify the position of the object and the image, and in some items, relate both by light rays. Further, they have to know that an observer sees images of objects along an “invisible” straight line (i.e., “line of sight”) and hence cannot detect if light was refracted before it entered the eyes. Finally, some items can only be solved satisfactorily when more than one ray is drawn. All this makes the items in factor 2 more complex than those in factor 1. The items in factor 3 require that, in addition to what was described in factor 2, students see an extended object (such as a pencil) as an infinite set of pointlike objects strung together in a straight line and then reason accordingly.

Whereas the increasing conceptual load of the items is one possible explanation for the different average item difficulties across the factors, there are others. There are, for instance, a number of aspects of item construction that need to be taken into account when considering item difficulty. Some of these are (so to speak) external to the item, such as the student’s own ability to correctly interpret the problem situation and/or any accompanying diagrams; others, are internal to the item and relate to issues of the validity and reliability of item construction, which includes the appropriateness of the choice of distractors. The latter becomes evident when the item “T3ax (no wrong choice)” is removed from factor 2. Then factor 2 (0.36) and 3 (0.37) have about the same average difficulty. A further consideration that can be anticipated as having an impact on student performance concerns the degree of familiarity they may have with such a situation. For instance, they may have been exposed to the widely used demonstration of refraction involving a laser beam striking a water surface. They may have observed this directly in a laboratory setting or been shown, or directed toward, a video demonstration thereof. However effective this may be at demonstrating the refraction of light between air and water in this setting, it cannot be assumed that students will be able to translate their understanding of the phenomena into other settings, such as when observing a rod partially immersed in water or trying to grab a coin from the ground of a fountain base (factors 2 and 3).

There may well be other interpretations of the three factors than the one given above. For instance, two other clustering of items into sets is in contradiction to the extracted factor structure:

1. Group the items with respect to the position of the observer. This would mean that there are three item sets: ray diagrams with no observer at all, items with a view of the phenomenon from the perspective of an observer, and items with a view of the phenomenon that is detached from the perspective of an observer.
2. Grouping the items with respect to the type of representation. This would mean that there are three item sets: ray diagrams (models), pictures, photographs or realistic drawings, and technical cross sections of different settings. As can be seen from Table I, these interpretations are not supported by the extracted factor structure. Thus, the empirical results do not support that the formulation of representation (as suggested by the modeling, see Fig. 2) or “the position of the observer” can be seen as latent variables.

3. Discussion of factor analysis results

Our analysis has limitations. We rely on students’ answers to the items of the questionnaire and, thus, we do not have any empirical information as to their relevance structure vis-à-vis why and how they ended up choosing the answers that they did, for example, what kind of difficulties they experienced in the process. Further, the small number of items in the questionnaire limited the content being addressed which in turn impacts content validity. Finally, in some instances, items could be interpreted as having more than one correct answer, which increases the odds of students’ success guessing.

Interpreting the factors and the item difficulties cannot reveal students’ reasoning in relation to a given item or task. Follow-up interview studies are needed to shed light on the following questions: Is the extrapolation from a pointlike object to an extended object an actual threshold? To what extent do the students work in strict disciplinary terms and to what extent do they rely on memories of observations when they answer the items? To what extent do students confuse the apparent bending of an extended object and the bending of a light ray?

Such results suggest that the physics of refraction is being taught with limited leveraging and coordination of the disciplinary affordances of related observable phenomena. This in turn suggests that physics educators need to appreciate that there is a qualitative difference between experiencing the enactment of a pattern of relevant aspects through a set of coordinated semiotic resources and being told what the pattern is. This conclusion is based on the view that meaningful learning in introductory physics courses connects students’ experiences and prior concepts to the way physicists model and describe phenomena [76].

In conclusion, what the factor analysis and our interpretation made clear is that a conceptual understanding of refraction requires both the awareness of and making meaning of, all the disciplinary-relevant aspects (DRAs) and that these DRAs present different levels of complexity to physics students. Being able to work with and understand ray diagrams in a disciplinary way is only one, arguably the easiest, of a least three different DRAs of refraction. Further, our analysis suggests that the discernment of these DRAs cannot be directly linked to particular representations of refraction (answer to RQ1). Instead, the data seem to support that DRAs are discerned by students by making use of different formulations of representations in a particular constellation dependent on the physical situation presented (answer to RQ2). In conclusion, our factor analyses helped to cluster items with respect to students’ abilities to solve tasks correctly. It is not designed to analyze or differentiate the different incorrect answers the participants gave. To explore this aspect further, we used MAMCR as described above. The next section provides this part of our analysis.

B. Approach 2: Module analysis for multiple-choice responses

1. Procedure for using MAMCR

The objective of using MAMCR was to address our research questions by obtaining a meaningful modular structure in a network of item choices made by the participating students. To do this, we performed MAMCR, as described in Sec. IV C, with two proposed cycles for creating backbone networks that could be analyzed: First, we used threshold frequencies in the range from \( f_t = 0.1 \) to \( f_t = 0.9 \) in steps of 0.1. Then, for each of these threshold frequencies, we extracted 1000 backbone networks and modular solutions as described in Sec. IV C 8. Implementing the criteria, we described for the first cycle we selected one backbone network and its modular solution for each threshold frequency.

This procedure results in nine backbone networks each with a different modular solution for grouping item choices from the questionnaire. For the purposes of this study, we wanted to focus on a single modular solution. Therefore, as part of our analysis, we compared and contrasted the nine modular solutions to choose one for the final analysis and interpretation. Below, we describe this process.

2. The optimal network to analyze

The procedure described in Sec. IV C 1 produced nine backbone networks with corresponding modular solutions for grouping item choices; one for each threshold frequency, \( f_t = 0.1, 0.2, \ldots, 0.9 \). We refer to them as \( f_{10}, f_{20}, \ldots, f_{90} \). For the purposes of our study, we focussed on one modular solution and we therefore searched for the threshold frequency, which we argue produced the optimal solution on which to base our interpretations. This required additional interpretations and analyses, which we describe here.

Table S1 provided in the Supplemental Material [55] displays the results for the optimal modular solutions (i.e., item choices in modules) for each investigated threshold frequency. Five modular solutions were excluded from consideration for further analysis for the following reasons. First, the three solutions with threshold frequencies of \( f_t > 0.6 \) were excluded because they have a large fraction of small modules (consisting of two or three nodes) and a
large fraction of item choices that fall into the largest cluster. f40 was excluded because half of the item choices were in one module, thereby making it difficult to interpret due to the large number of answers, and finally, f10 was excluded because it contains only half of the item choices.

Figure 6 provides another way to view the different modular solutions, see Refs. [77–79]. Solution f50 contains two modules of roughly equal accumulated node visit frequencies (bottom and middle modules) and a third module with a smaller sum of node visit frequencies (top module). These three modules contain 21 item choices (accumulating 38% of the node visits), 23 item choices (accumulating 34% of the node visits), and 9 item choices (accumulating 18% of the node visits), colored blue, yellow, and red, respectively. When comparing the solution f50 to f60, the upper module of f60 gains the item choice T3aii (which shows a picture of a coin in a cup becoming visible after adding water). The inclusion of this response does not significantly change the modular structure, but it adds a considerable amount of node visits to the upper module (the added gray area to the red module). However, from a content perspective, we found that this item choice added little new knowledge (it can conceptually be placed in module 3 discussed below), whereas the node representing T3aii became dominant in that cluster thus potentially obscuring the identification of other relationships (see the Supplemental Material [55]). Apart from this, a splitting of modules from f50 to f60 was observed. In contrast

FIG. 6. Graphical representation of how modules are related between optimal modular solutions for threshold frequencies of 20% (f20), 30% (f30), 50% (f50), and 60% (f60). The height of the bars depicts the accumulated prevalence of item choices in terms of node visit frequencies. The codes on the left and right indicate the names of the item choices with the highest node visit frequency in each module in the modular solution f20 (f60). The colored streamlines between modular solutions indicate the accumulated shifts of item choices relative to f50. For f50, we have indicated the three modules that were eventually selected for in-depth analysis (see Sec. V B 3).

FIG. 7. A map of the most frequent module solution found with Infomap for the f50 backbone network. The size of the circles around the item choices is proportional to their accumulated node visit frequencies. The thickness of the links between the item choices indicates that there are strong connections between the three largest modules: module 1, module 2, and module 3.
to the relatively small changes in the transition from f50 to f60, there are large changes in the transition from f20 to f30 and from f30 to f50. The change from f30 to f50 is very large: the addition of eight new high-frequency item choices splits the very large (bottom) module of f30 into the two largest modules of f50. This makes f50 our preferred modular solution. The change from f20 to f30 shows that the three largest modules of f30 contain item choices from smaller modules of f20, so f30 is preferable to f20. As a result, f50 was kept as the solution with the most information. This solution retained 76 out of 81 possible item choices.

Figures 7 and 8 depict the modular solution f50 in two different ways: (i) by showing the connections between the modules in a map (with sizes indicating the sum of node visit frequencies for each module) and (ii) by plotting the co-occurrence matrix for the chosen optimal network.

For f50, the three largest modules dominate the solution, and the modules share connections (visible as the links in Fig. 7). The sharing of connections may signify that if modules represent strategies when answering the questionnaire, then these strategies can overlap (see Sec. V B 3).

The module co-occurrence matrix (Fig. 8) shows that Infomap consistently finds the same solution as evidenced by the item choices being assigned to the same module in most runs (see Supplemental Material [55] for the same diagrams for all modular solutions).
when we ran Infomap multiple ($n_{iter} = 1000$) times. Specifically, we consistently find the three large modules that are shown in Figs. 9–11.

Figures 9–11 show the internal structure of the three largest modules in modular solution f50. Following Brewe et al. [62] and to aid further interpretation, we modified the diagrammatic schematic pictorial representations that were used in the questionnaire to represent each item choice and inserted them as nodes in these illustrations.

3. Interpretation of the modules that emerged from the network analysis

The network analysis extracted ten modules from the data. We used the following criteria to choose modules for interpretation and thus to select item choices that were most important for this:

- We excluded three modules that contained only “no-answer” item choices. These item choices represent items that participants chose not to answer.
We excluded one module with three-item choices because these represent different nonstandard solutions (e.g., light-ray drawings with rays pointing in a lot of different directions).

We excluded one module with three-item choices because these item choices belong to one item only and the solutions do not differ (participants drew in straight light rays in all three-item choices indicating that they saw no refraction at a media-boundary in this representation formulation).

We excluded two modules that contained only two-item choices each because these modules do not explain much of the relations between the item choices.
For every module, the item choices with the highest node visit frequencies and the strongest links between the item choices were given priority for the interpretation. This applies to five- or six-item choices per module.

This procedure led to three modules (Table II).

Module 1 contains almost only incorrect answers that seem to follow a pattern that can be expressed by the rule: “objects-are-moved-downward-and-extended-objects-get-bent-away-from-observer.” In contrast, module 2 includes correct item choices with a high node visit frequency and may be expressed by “extended-objects-are-moved-upward-and-get-bent.” However, the item choices with smaller node visit frequency in this module indicate that there is little consistency in how pointlike objects are perceived to shift. Module 3 contains item choices that indicate that pointlike objects are lifted upwards but the bending of extended objects is neglected: they shorten. It may be expressed by the rule: “objects-are-moved-upward,” which includes both pointlike objects and extended objects. The item choices of module 3 with high-node visit frequencies are correct. Thus, the three modules can be interpreted to represent structures that are related to the conceptual understanding of the students (RQ2).

Specifically, we argue that the three modules

![Diagram of Module 3]

**TABLE II.** Summary of extracted modules of the network analysis and their characteristics (the questionnaire with all the tasks, items, and their codes can be found in the Supplementary Material [55]). Correct item choices are highlighted using *italics*.

<table>
<thead>
<tr>
<th>Module</th>
<th>N_items</th>
<th>Top node visit frequency item choices</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>T3cleft, T2a2e, T2a1e, T2b1P2, T2a3g</td>
<td>Pointlike objects appear to be shifted downwards, and extended objects appear bent away from observer, at the boundary between two different media.</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>T2a2d, T2a1d, T1Q12right, T3cright, T2a4c, T2a3a</td>
<td>Extended objects appear to be both shifted upwards and bent at the boundary between two different media, and pointlike objects get shifted in different directions.</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>T2b1P1P2P3, T3blaserB, T3bspearC, T2a3b, T2a4b</td>
<td>Pointlike objects appear to get shifted upwards, and the length of extended objects shorten when they travel through two different media.</td>
</tr>
</tbody>
</table>
represent relevance structures; connected expressions of what students perceived necessary in the situation for accessing the concept of refraction.

These results point to the same type of answer to RQ1 that was obtained using the factor analysis approach. All modules include different formulations of representation and no one-to-one relation was found between modules of correct and different incorrect answers to the questionnaire items and the formulations of representation that these items were embedded in.

4. Discussion of the analytical procedure followed for the network analysis

Our MAMCR suffers from many of the same limitations as our factor analyses (see Sec. VA 3), as this part of the analysis relies on the same data. Moreover, MAMCR is still a developing methodology, and we have made several choices in our analyses, which could have been made differently. For example, for each threshold frequency, we opted to select 1 out of 1000 networks based on rules of thumb (looking for high modularity solutions and fractions of duos and trios) and manageability (number of modules and number of item choices in the largest cluster). In addition, out of backbone networks for nine different threshold frequencies, we chose one threshold frequency on which to focus our interpretative analysis. To further contribute to addressing the aim of our article, future work should include expanding on how these selections are made in a reproducible manner. Another apparent limitation of the current version of MAMCR is that it does not include all item choices. However, since the goal of using MAMCR for our study was to tease out fine-grained answering patterns, we argue that excluding item choices that do not contribute to this goal is warranted.

When using MAMCR, setting an appropriate value for the threshold frequency is crucial. The smaller the threshold frequency (i.e., the fraction of participants who select the item choice) the more fine-grained the structures are that the analysis detects. On the other hand, if the threshold frequency is set too low, the detected patterns may only reflect noise. Here, noise means that the patterns only reflect random connections between item choices because a few students happened to select both. A higher threshold frequency allows for more frequent item choices to be included, thus representing more of the student population’s choices. The price for this inclusion is that high-frequency item choices tend to dominate the network hiding fine-grained structures.

As an example, consider T1Q13 where a light ray must be drawn similar to the item in Fig. 12. Here, 81% of the participants gave a correct response. If this answer is included in the analysis, it dominates the entire network because it is linked to so many other items. None of the other item choices for this item reached above the 10% threshold and were included in the f10 backbone network (see Supplemental Material’s Fig. 10 [55]). However, none of the patterns that the remaining item choices for T1Q13 were part of in this network, represented anything conceptually meaningful. Thus, we saw these patterns as noise. This example shows that setting a frequency makes a fair balance possible between omitting item choices and avoiding the detected patterns representing only noise. Similarly, a balance needs to be struck when finding a manageable number of modules.

We found three modules (see Table II) with sets of closely related item choices in terms of how they were answered. All three modules can be interpreted as characterizing a relevance structure [13]. In this study, relevance structures appeared to be built on two aspects—refracted objects will appear bent and appear shifted, each aspect having its own dimensions. The first is a relatively smaller shift toward an observer, and the second is a relatively larger shift upwards, which taken together generates a heuristic shift perspective that evokes a conceptual and visual expectation: an object’s image can be expected to be seen shifted upwards and toward the observer. The application of such a heuristic also has two application frames: pointlike objects and extended objects. Seeing a relevance structure “at work” in the analysis is educationally valuable. For example, when using a “point-like-objects-are-removed-upward,” heuristic may give rise to an adequate conclusion for a fish and spear problem—a fish appears to the spear thrower to be lifted upwards. However, this heuristic becomes inadequate for dealing with extended objects that are placed across two different media (for example, a pencil in a glass of water). In this respect, the network analysis showed clearly that many incorrect item choices are related to each other (module 1) and that a mix of correct and incorrect item choices falls into one of two modules (either module 2 or 3). Thus, leading to the

![Figure 12](image-url)
VI. GENERAL DISCUSSION

The two analytic approaches that we used to explore the aim of our article tackled this task through the framing of two research questions. These research questions were written in ways that could be specifically and appropriately addressed by FA and MAMCR. Together the results provide a detailed exploration into relational aspects that theoretically could be found between formulations of representation and students’ performance on conceptual tasks that are located in the same conceptual domain (refraction). We found that the students’ performances did not reveal any direct one-to-one relations with the representational formulation used to present the task. This is probably because of the unaccounted-for complexity that the role of the students’ relevance structures brought to the data (which must always be the case when the situational coordination of multiple DRAs is an essential part of the task under study). What brings further complexity to the research situation is that most physics representations are multimodal and consideration cannot be given solely to the affordances of the individual modes nor to a simple addition of modes. The modes do semiotic work together, which in Lemke’s words, “multiple meaning.” [80].

On the contrary, the extracted clusters of both the FA and the MAMCR do contain distinct collections of formulations of representation that are complexly related. Even though the FA and the MAMCR cluster items and item choices differ, in the clusters of both methods, the visual representations are “mixed”: clusters contain pictures and sketches as well as views from an observer’s perspective and views detached from an observer. To our knowledge, this is the first empirical study with a large sample size in PER which shows that clusters of related items and item choices—indicating a particular competence in relation to a particular object of learning—are related to a group of representational formulations. Even though it may seem obvious, we now have empirical evidence that a set of representations—a critical constellation—is needed to fully facilitate the awareness needed for an intended object of learning (i.e., a certain learning objective). The review of physics textbooks done by Hüttebräuker [59], and the teaching resources that have emerged from PER work, indicates that such research-informed use of representations has only very recently started to influence physics educational practices.

Our results also provide further insight into how students’ relevance structures may be interpreted vis-à-vis from a theoretical viewpoint about being competent in critical constellations of representations. While we have not identified a critical constellation of representations that students need to develop competency to develop an appropriate and holistic understanding of refraction fundamentals, we did find a degree of consistency in how different representational formulations get used to answer visual image-object refraction questions. Our factor analysis used students’ correct answers and clustered items according to their difficulty. The rays, pointlike, and extended object factors that we obtained can be interpreted as expressions of three DRAs that need to be afforded access through the use of representations made up of modes that do the necessary semiotic work together to offer such disciplinary affordances. A result that is arguably vital for educational endeavors aimed at optimizing learning outcomes. Furthermore, the visual representations in the items may have evoked more complex relevance structures. For example, the ray diagram [task 1, Fig. 3(a)] shows the problem in a two-dimensional sketch and provides the incoming light ray and the boundary of the two media while item T3c—a picture of a pencil that is stuck into the water—is three dimensional and contains no “typical” tools (like light rays, a given media boundary, and the omission of irrelevant aspects) that may help the students to solve the problem.

The three main modules found with our expanded version of MAMCR can be identified as relevance structures related in different ways to the apparent shifting (up and down), bending, and shrinking of pointlike and extended objects. The three modules differ from each other in the way in which pointlike objects seem to be moved (in module 1, downward; in module 2, shifted; and in module 3, upward). This can be interpreted as a different conceptual understanding and, hence, a different relevance structure was evoked. One of the relevance structures is associated with wrong answers and is not part of a critical constellation of representation. The other two, however, may in some situations lead to correct answers and in other situations to incorrect answers. Hence, a formal answer to our RQ2 may be expressed as follows: Our results show how two different cluster analysis approaches lead to two different structures, each of which can be seen as making a unique contribution to address the aim of our article. As is the case for the RQ1 results, the RQ2 results indicate that the theory that our article set out to explore has solid empirical support. This is despite our analysis being unable to also bring out what that critical constellation is for the refraction DRAs that make up our questionnaire. Another way to appreciate the results for both RQs is that they present good empirical evidence that students’ relevance structures, which are taken to mirror students’ conceptual growth in a particular area of physics, need to, and can be better developed through the insightful educational use of representations vis-à-vis taking into account the present and appresent affordances that the modes making up these representations may be able to offer for the task at hand. Following the proposal made by Fredlund et al. [14], such
insightful educational use of representations would need to incorporate the following three essential features:

1. The identification of DRAs for the task at hand. For the refraction of light at a boundary between two media, these are “direction,” “medium,” and “speed of light.”

2. The selection of an ensemble of representations, which provides the best potential to present these DRAs. Our work suggests that in terms of visual representations, teachers would need to include representation formulations, which include rays, pointlike objects, and extended objects.

3. Within the selected representations, dimensions of variation need to be formulated against a background of sameness to optimize the possibility of discernment of the DRA and the ways that these dimensions of variation are related to each other. For our example, this would not only include variations of direction and media but also variations in objects and points of view.

Such a teaching design would take practical advantage of our result, which certain representational formulations do not correspond to certain disciplinary-relevant aspects. Even though FA and MAMCR analyzed our dataset in different ways and yielded different types of results, both can be strongly related to making learning possible through the development of relevance structures. In this way, the complexity of a certain item (FA) and the student response behavior (MAMCR) can both be seen to present cognitive aspects that are related to problem solving. Representations will always afford access to both relevant and irrelevant aspects of a given task and in this way evoke different relevant structures in terms of complexity and thus the possible range of conceptual constitution.

Thus, we argue that very different representations may contribute to a DRA. Or put in another way: When teaching a certain topic and breaking the topic down into different DRAs, each DRA is likely to need more than one representation to cover it fully. This makes the matter more complex because each visual representation can carry much more than its present aspects. Take, for example, the items (task 2) where a cross section of a pencil stuck in water is shown and students were asked where an observer would see the tip of the pencil. To correctly answer these questions, students must be aware of appresent aspects (e.g., that the tip of the pencil emits light in all directions and only one light path is relevant or that an observer sees the image of the tip in the reverse direction in which the light enters the eye). These appresent aspects make these items difficult (see Sec. VA) and prone to inconsistent or misinformed answers (see Sec. V B). Hence, discussing refraction using (i) only ray diagrams will not be sufficient for covering both the deflection of light as well as explaining refraction as a whole. (ii) When using diagrams, present aspects, as well as appresent aspects, have to be explicitly addressed by research-informed teaching and learning activities. We assume that this holds for other objects of learning as well even though this still needs to be shown.

In summary, the main outcome of our study is that with a considerably large sample, we have been able to show empirically that every cluster of students’ answers contains distinct collections of semiotic formulations of representations and that these relations are highly complex. This outcome strongly supports the theoretical proposal that achieving an intended object of learning requires a particular representational competence in a set of representational formulations made up of a critical constellation of systematically used semiotic resources, which are the semiotic modes particular to the physics community and thus its discursive practices (for further discussion, see Ref. [11]). Our study also has implications for what it means to develop representational competence. A natural assumption may be that students would need fluency in a critical constellation of representational forms before being introduced to a particular disciplinary way of knowing or doing. However, the structures identified in this study indicate that there is a complex set of relations between the access to disciplinary ways of knowing and the ability to appropriately understand and use disciplinary formulations of representation. Thus, it may be more fruitful to educationally see the development of discursive fluency being “part and parcel” of the development of representational competence as per the model presented in Fig. 1.

ACKNOWLEDGMENTS

We thank Anne Linder, Eugenia Etkina, Jayson Nilsson, Tobias Ludwig, and Jonathan Clark for the valuable comments and suggestions they made on our developing versions of this article. We would especially like to thank Anne Linder and Jonathan Clark for the extensive time they repeatedly gave to help improve grammar flow and logical clarity as the article developed. We also would like to thank our end-stage reviewers for the extremely careful and in-depth readings they did. This provided us with two sets of extremely helpful and constructive feedback. Many years of highly rewarding collaborative work have gone into the production of this article, and the supportive funding that we received from the Swedish Research Council, Grants No. 2008-5339 and No. 2016-04113, is most gratefully acknowledged. The article processing charge was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491192747 and the Open Access Publication Fund of Humboldt-Universität zu Berlin.


