PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **15,** 010108 (2019)

Assessing German professors' views of nature of science

David Woitkowski[*]
Paderborn University, Physics Education, Department of Physics,
Warburger Straße 100, 33098 Paderborn, Germany

Nora Leonie Wurmbach[†]
Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104, USA

(Received 15 August 2018; published 28 January 2019)

University professors' views of nature of science play an important role when guiding undergraduate students into the culture of practicing physicists. While these topics have a long-standing tradition in U.S. curricula they are not part of German educational standards or curricula. Additionally professors' views—in contrast to those of their students—are relatively under-researched. In this contribution, we establish the possibility of testing German physics professors' views in an economically administrable survey. We first ask for their views of the nature of science, and then about how important they see these aspects for students and how intensive they address them in their own physics classes at the university. We demonstrate that an established test instrument can be reliably used with this demographic. The results indicate that the professors tend to views of naive empiricism but besides that hold mostly adequate beliefs. Learning about the nature of science in university courses, in general, is considered of much importance, which is also reflected in the professors' reported teaching habits. However, there are aspects in this area which are considered more or less important by the professors.

DOI: 10.1103/PhysRevPhysEducRes.15.010108

I. INTRODUCTION

Becoming a physicist is more than acquiring knowledge, but rather a complex process of becoming part of a specific culture. This culture can be characterized as a community of people dealing with complex and—in their complexity—interesting problems [1–4]; people who acquire, discuss, and share knowledge using a variety of empirical and theoretical methods. Entering this culture is only possible by negotiating identity and making meaning during a long-going process [5].

The place where this process might happen, should or could be the university physics education. However, the curricula often focus on learning content knowledge and acquiring experimental or theoretical skills. Nevertheless, one might identify a kind of "hidden curriculum" consisting of epistemology, ontology, and discourse [6]. These hidden parts of physics education directly convey elements of this culture of physics but (as they are not codified in any way) depend more on the personal beliefs and attitudes of those who teach physics—the professors. Their views in this area are relatively under-researched and only little known in detail. To the authors' knowledge only Schwartz *et al.* have conducted a study (with a low sample size) within this group [7] and Karakas provides evidence that such contents do now play a considerable role in actual classroom action compared to physics content knowledge [8].

The nature of science has been part of the U.S. science curriculum for some time [9]. And the educational standards agree that there is some knowledge about how science works and what characterizes scientific knowledge that every educated adult should know [10]. The German standards, on the other hand, concentrate more on the contribution of physics to "Bildung." They do contain competencies in the area of scientific investigations, i.e., using experiments and models, but they do not call for knowledge *about* the nature of science as the American tradition does [11–13].

On this basis we will deal with the nature of science [(NOS), or epistemological beliefs which is subtly different] as part of that hidden curriculum, with those professors who supposedly teach this implicitly during their courses at universities within a culture that traditionally does not take NOS as an important content. We will claim that there is no easily applicable, economically usable test to use with German university professors. Thus we will try to assess them with tests that were originally developed for students and discuss what works and what does not. In the second part of the study, we devise a new instrument to inquire

---

[*] david.woitkowski@upb.de
[†] Present address: University of Paderborn.

about what aspects of NOS professors actually consider of importance in university physics education and whether they address these topics in their classes.

II. THEORY

When talking about nature of science we have to deal with two research perspectives with different interests and traditions [14]. Those perspectives then take different testing approaches—a matter that is further complicated by the fact that most tests are only available in English but our study is conducted in Germany. We will first give a short characterization and localize our own research.

A. Nature of science and epistemological beliefs

A person's perception of how the sciences—and physics in particular—work, how they create knowledge, and what (e.g., ontological) status that knowledge has is subject to two different research fields and thus called differently.

The conceptually broader approach comes from cognitive and developmental psychology. Here a person's belief system can be identified as one factor of influence on that person's learning in a specific area [15–17]. These beliefs cover a wide range of topics among which epistemology is only one. A typical research question here will try to characterize these beliefs in a variety of ways. The interest lies more on the effect on learning outcomes than on the factual correctness of the beliefs. Hofer and Pintrich describe four aspects of epistemological beliefs as depicted in Fig. 1 [15]; like others, these do not include aspects overly specific to the sciences.

The term nature of science, on the other hand, stems from the area of teaching and curriculum development. It carries the aspect of students *knowing* what can factually be said about how sciences work and what status its knowledge has [18]. Lederman defines the area in question as "epistemology of science, science as a way of knowing, or the values and beliefs inherent to scientific knowledge or the development of scientific knowledge" [19], p. 303. For this reason there are a number of aspect catalogs [19–21] that are a lot more comprehensive than the categories by Hofer and Pintrich. A typical list includes aspects of the *nature of scientific knowledge* (NOSK) and the *nature of scientific inquiry* (NOSI). These lists are not without criticism for tending to be too strict and too focused, e.g., on the experiment [22].

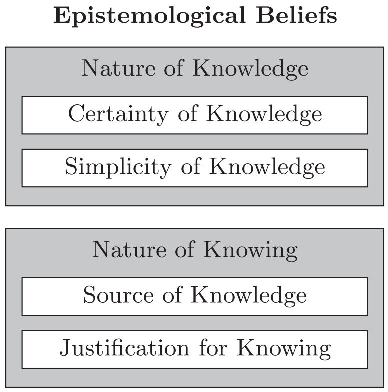


FIG. 1. Four dimensions of epistemological beliefs according to Hofer and Pintrich [15].

Later in this paper we will ask about professors' beliefs in this area. For this we need a comprehensive list of relevant fields of NOS but we do not want to inherit the *knowledge* aspect of the NOS research tradition. We rather want to characterize the professors' beliefs like the psychology discussion does. Our approach thus has to be somewhere between those two. Knowing about the different traditions we will use the terms NOS and epistemological beliefs interchangeably.

B. Testing approaches

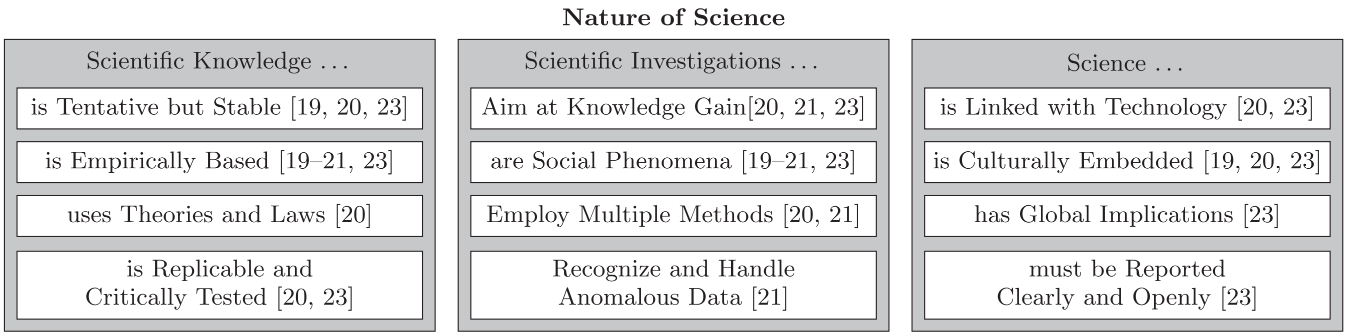
Studies in this area may be divided into two stages: The first is to assess relevant aspects of NOS as well as correct and incorrect judgments concerning these aspects. The first stage is normally conducted with several kinds of experts in the fields of science and science communication. The second phase is then to test persons from a specific demographic—often students—for their views and compare them to those that have been found correct or adequate in the first stage.

Instruments in the first stage take very open approaches as they should generate lists of relevant aspects that can further be used. The most referenced works might be the textual analysis by McComas and Olson that dealt with curricula and standards of five different countries [23]; further, the Delphi study conducted by Osborne *et al.* that generated ten statements about scientific methods, the nature of scientific knowledge, and the sciences' social aspects where consent could be found [20]; lastly, the studies of Lederman *et al.* [19,24,25], which mainly dealt with the usefulness of that knowledge.

Neumann and Kremer have compiled a juxtaposition of typical aspects of NOS (Fig. 2). In comparison to aspects from the psychological tradition (cf. Fig. 1) these lists are by far more comprehensive and more specific to the sciences.

Based on (one of) these catalogs more economically usable instruments can be constructed which can be used in the second phase. Two well-known instruments are *views of nature of science* (VNOS) [24], which consists of open-ended and rather comprehensive questions (e.g., "What, in your view, is science?," "What makes science […] different from other disciplines of inquiry […]?") and *views about science survey* (VASS) [27], which uses contrasting statements among which test persons have to choose [e.g., "The laws of physics portray the real world: (a) exactly the way it is. (b) by approximation."].

When investigating professors' views of NOS, VNOS might be a suitable instrument but as an open-ended



Nature of Science

Scientific Knowledge …	Scientific Investigations …	Science …
is Tentative but Stable [19, 20, 23]	Aim at Knowledge Gain[20, 21, 23]	is Linked with Technology [20, 23]
is Empirically Based [19–21, 23]	are Social Phenomena [19–21, 23]	is Culturally Embedded [19, 20, 23]
uses Theories and Laws [20]	Employ Multiple Methods [20, 21]	has Global Implications [23]
is Replicable and Critically Tested [20, 23]	Recognize and Handle Anomalous Data [21]	must be Reported Clearly and Openly [23]

FIG. 2. Aspect of NOS (based on the juxtaposition by Neumann and Kremer [26]), employing the works by McComas and Olson [23]; Osborne *et al.* [20]; Lederman [19], p. 304; and Schwartz *et al.* [21], p. 4ff.

questionnaire it is rather complex and laborious to interpret. Furthermore, there does not seem to be a German translation of this instrument. VASS does not have this same limitation. As a closed questionnaire it should be easy to interpret and there is a German translation by Priemer [28]. Unfortunately this instrument is clearly aimed at students (e.g., Priemer targets 12th to 13th graders in advanced physics courses) and might thus be too easy for our target audience.

The German physics didactics community has only recently started to produce test instruments in this context. Two instruments covering most of the aspects in Fig. 2 come from Neumann and Kremer. Neumann clearly uses a NOS approach. She constructs items where a historical situation is present in an item stem, and then several questions ask for the students' understanding of that situation [29]. There exists both a German and an English version of her instrument. Kremer's instrument, on the other hand, fits more within the framework of epistemological beliefs [30]. The students are presented with statements and have to rate them on five-point Likert scales. The tested aspects are however oriented at the works in the context of NOS.

C. The problem of reliability and validity

When devising new test instruments or using established instruments for an audience other than the original target audience there are several concerns to be addressed. Two important ones in our context are reliability and validity of the test used.

When using several items to form a scale within a test we first need to establish that these items indeed measure the same construct, the same psychological trait, and, in our case, the same aspect of one's belief system. We say the measurement has to be reliable.

Since a measurement is always subject to uncertainties, and two items might never target exactly the same belief, we need information on how close to each other the items of our scale are [31]. In practical terms Cronbach's α is often used [32]. This index is a measure of the mean correlation of the scales' items. If that correlation is low (often $\alpha < 0.7$ is said to be too low) the items seem not to measure the same trait and thus might not form a coherent scale. Another measure is to correlate each item with the rest of the scale. If the correlation coefficient is too low (here we will use $r < 0.3$) the item in question might measure something considerably different than the rest of the scale.

While reliability can be discussed on a merely statistical basis, validity is of a more complex nature. In short, the question is "Does the scale measure the thing we think it should measure?" A modern concept of validity is presented by Kane [33]: In his words one has to give an *interpretation/use argument* which establishes that the specific interpretation or further use of the acquired test scores is justified. The problem here is that validity can never be proven but might be disproved. Thus it takes considerable work to establish a validity argument.

As our goal is not to develop a completely new NOS test for our target audience (physics professors), we want to examine the usability of already established and validated instruments. We will argue along the following lines: (a) The used scales can be considered valid for their respective original target audience (i.e., German students) due to the original authors' validation studies. (b) There are several possible sources for a lack of validity when used with our target audience. One could be language barriers or difference in sociocultural or teaching backgrounds. Those can be minimized by using scales from tests that were developed and validated within a similar cultural background. Another source for invalidity might be a different understanding of several test items because of the different professional and educational background of professors vs students. In our situation we consider this to be the main threat to validity. (c) A single misunderstood item would lower a scales' internal consistency indicated by a low α. We thus can regard a low reliability as indicating low validity. However, it is hardly conceivable that *all* items in one scale are misunderstood in the same way such that they now *again* form a consistent scale. (d) We thus first check for the scales' reliability. Scales with low reliability cannot be used for further measurement. (e) Because of (a) and (c) we do not see any threats to validity which will leave a scale with a high reliability. We thus will preliminarily consider the reliable scales as valid because of the original authors' work.

To additionally aid the reader's judgment of validity we present a translation of all used test items in the Appendix. We also present participants' comments on the used test items as further indicators for their understanding of the items.

III. AIM AND IMPLEMENTATION OF THE STUDY

Professors' views of nature of science—although they play a vital role in the processes introducing young physicists into the culture of physics—are relatively under-researched. Especially in Germany where there is no long-standing tradition of NOS as curriculum content and university professors receive only limited training in teaching. There is little to no evidence concerning their views and teaching habits in university classes.

A. Research questions

Our four research questions can be divided into two groups as follows.

In the first part we will use scales from established closed-form test instruments to test German physics professors' views of NOS. The research questions here are the following:

QI Do these scales allow for a reliable testing within the group of physics professors?

QII Are their views adequate in the light of what can be considered correct NOS knowledge?

For the second part we have developed a closed-form test instrument asking for aspects of NOS the professors implicitly or explicitly teach in their physics classes.

QIII How important are the aspects of NOS from the first part in the professors' eyes; should students learn about them at university?

QIV To what extent do professors report to teach aspects of NOS in their classes at German universities?

Most of our scales use scientific and/or epistemological concepts of which professors might have a more differentiated understanding. Further, their use of the used terms might be shaped by experience or practical work in physics research in a different way than is to be expected among students. For this reason for Q1 we would expect the typical scales not to be very reliable (i.e., show low Cronbach's α) as discussed in Sec. II C.

For those scales that can be used reliably we expect the professors to show mostly adequate views (QII) as they are professionals in their field and conduct scientific investigations on a day-to-day basis.

As NOS does not play a significant role in German science education as a compulsory curriculum content, we would expect these aspects to be even more "hidden" than identified by Redish [6]. Thus, for QIV we expect only a few considerations of these topics when teaching physics. However, they might find them more relevant in general (QIII).

B. Test instrument

For the first part, our goal is not to develop a new instrument but to try to use scales from existing ones. Here we can choose between the three German instruments by Priemer [28] (which is a German translation of VASS), by Neumann [29] (which is available in German *and* English), and by Kremer [30] (which was developed solely in German). Additionally, some scales by Riese [34] (part of a German competency test for prospective physics teachers) test some of the relevant NOS aspects. The used instrument should be used economically (i.e., as an online survey, see below) and fit the specified demography.

The open-form instrument by Priemer would be too laborious to use and interpret. From the other two, Kremer's instrument has the advantages of operationalizing more aspects of NOS and at the same time not employing the "knowledge" approach which is typical for NOS research—we do not want to discourage the professors by *testing their knowledge* but rather we want to *learn about their views*. Furthermore, a translated test (like Primer's translation of VASS) might or might not work as expected in the new language due to sociocultural differences. To reduce the risk of misunderstood items we only use scales originally developed in German as the test will be administered in German as well.

Kremer describes five aspects of the NOSK and four aspects of the NOSI [35] and has developed test scales for most of them (Table I) [36]. For NOSK-TAL (theories and laws) an NOSI-MET (scientific methods) Kremer does not present test items so we use short scales developed by Riese [34] instead. The scales NOSK-CRT (certainty of knowledge) and NOSK-SRC (source of knowledge) in Kremer's test contain several items that seem not suitable for the target demographic. Thus these scales are also replaced: NOSK-CRT by a matching scale from Riese, NOSK-SRC by a newly constructed scale based on the description in Kremer's work [35], p. 78. To keep the original scales intact we did not drop any items from the scales before administering the test nor did we change individual items.

All items require the participants to rate a statement on a 4-point Likert scale ranging from 1 = "totally incorrect" to 4 = "totally correct." For analysis we employed means of classical test theory. The results are reported in a way that higher scores should indicate more adequate views.

For the second part of the study we want to investigate whether the professors regard the mentioned NOS aspects as important content. For this we derived 2 content items for each aspect as shown in Table II based on the descriptions by Kremer [30].

TABLE I. Overview of the test scales. Example items marked (–) are inverted. Scales from Kremer's tests [36] are marked *K*, from Riese [34] *R* and scales newly constructed on the basis of Kremer's description [35] are marked *N*. All test items can be found in the Appendix.

Id	Scale	Items	Example item	Source
NOSK-CRT	Certainty of knowledge	6	Even physical knowledge is not clearly provable and can change over time.	*R*
NOSK-DEV	Development of knowledge	8	New discoveries can change what scientists think is right.	*K*
NOSK-SMP	Simplicity of knowledge	5	The more complicated a scientific theory is, the higher its reputation is among scientists. (–)	*K*
NOSK-JST	Justification of knowledge	9	Good theories rely on the results of many different experiments.	*K*
NOSK-SRC	Source of knowledge	5	Scientific statements can only be verified by scientists. (–)	*N*
NOSI-PRP	Purpose of the sciences	5	The goal of scientific theory is to give order to part of the human experience.	*K*
NOSI-TAL	Theories and laws	3	Theories are not yet proven, laws are fact. (–)	*R*
NOSI-MET	Scientific methods	3	In order to gain new physical insights one has to proceed according to the following method: generation of hypothesis—development of appropriate experiments—observation and evaluation—derivation of laws. (–)	*R*
NOSI-CRE	Creativity and imagination	5	Creative thinking is incompatible with logic-based science. (–)	*K*

The participants were given these content items with the prompt to rate how important it is for physics students to learn these things during their physics courses at university in general. Here we used a 4-point scale ranging from 1 = "not important at all" to 4 = "very important."

After this block the professors were asked about an own class they teach on a regular basis which is at least somehow representative of their typical teaching. Now they got the same items (in different order) with the prompt to rate how intensively they address these topics *in their own class*. Here we used a 4-point scale with 1 = "is not addressed," 2 = "is addressed implicitly," 3 = "is addressed explicitly," and 4 = "is core learning objective."

An English translation of all test items used is given in the Appendix.

The test as a whole was administered as an online survey with randomized item order (within item blocks). First came the NOS scales from Table I, then demographic data, and then the teaching scales from Table II in two versions as described.

C. Sample

The online test was conducted with 50 physics professors at German universities. The participants were reached by contacting the KFP [37] who advertised the study among its members. Additionally the physics department's dean at each German university was contacted and asked to distribute the call among their colleagues. The participants remain anonymous and did not receive any compensation.

The sample demographics are shown in Table III. The participants report 20 different universities as their affiliation, 6 choose not to state their affiliation. At a maximum 5 participants come from the same location.

In Germany, physicists are typically divided into theoretical working physicists (including computational physics) and experimental working physicists (including applied physics), furthermore people working in physics didactics took part.

The participants were asked about their perceived background in NOS by rating their own experience from 1 = "much less than average" to 5 = "much more than

TABLE II. Overview of content items for part 2 of the test. All items can be found in the Appendix.

Aspect	Example content (1 of 2 in total per aspect)
NOSK-CRT	Decided under what conditions physical statements are to be classified as secure knowledge.
NOSK-DEV	To correctly estimate the temporal stability of physical knowledge.
NOSK-SMP	Can specify criteria for the decision between different possible theories.
NOSK-JST	Understand reasons for repeated measurements.
NOSK-SRC	Discuss to what extent all people, not only educated scientists, are or can be involved in research.
NOSI-PRP	The description, explanation and prediction of phenomena is an essential goal of scientific research.
NOSI-TAL	Explain the difference between laws and theories in science.
NOSI-MET	To know different possible processes from the question to the answer in the physical research process.
NOSI-CRE	Reflect on the role of creativity in theory formation in physics.

TABLE III. Sample demographics. The perceived background in NOS was measured on a scale with 1 = much less than average to 5 = much more than average, compared to their peer physicists.

		Background in NOS	
Area of work	N	Mean	SD
Experimental physics	27	3.42	0.93
Theoretical physics	17	3.47	1.01
Physics didactics	4	4.25	0.50
Sum	50	3.50	0.94

average" compared to their peer physicists. On average the participants perceive their experience as slightly higher than the average physicist, where the professors in the area of didactics are clearly ahead of the rest (Table III).

IV. RESULTS

We present the results in the order of the research questions. The used statistical methods are mentioned alongside the results.

A. Participants' comments on the test items

At the end of the online test we asked the participants for general comments on the survey or the research project in general. Out of the 50 participants, 15 left comments at all. Of these, especially comments concerning the participants' understanding of the used items are of interest.

Those 9 comments can be divided into several categories as shown in Table IV. Here the categories 1 and 2 contain comments specifically on the questions in NOSI-TAL. The comments in category 1 argue for stating this relationship more nuanced and present a rather specific (and mostly adequate) view. Those in category 2 state the opinion that there is no good reason for making a difference between theories and laws in the context of physics research [38]. Comments from both categories give rise to serious doubts concerning the validity of NOSI-TAL.

The comments in category 3 did not name any specific item or topic but stated that "some questions" should better be stated in more nuance. The general stance here could be that the reality of research in physics is more complex than can be put into a closed-form test. This can be seen as a threat to validity if our goal is to consider *all details and nuances* of science. Considering the literature review, we get the impression that research in the field of NOS merely tries to concentrate on the *important* parts.

TABLE IV. Participants' comments concerning their understanding of the used items, grouped by category.

Cat.	N	Comment paraphrase
1	2	Relationship between theory and law should be considered in a more nuanced way.
2	2	The difference between theory and law is not (practically) relevant.
3	3	Some questions should be more nuanced.
4	2	In several cases questions where unclear, then "no answer" was given.

Two further participants commented that they did not understand some questions at all and in those cases declined to answer (which was possible with the survey tool used). This behavior is expected. Scales with too many missing answers will (for statistical reasons) probably show low internal consistency.

In summary, the participants' comments mainly target the scale NOSI-TAL. As argued in Sec. II C we expect this to lead to a low reliability—which is indeed the case as we will see in the following section.

B. Scale analysis

As a measure of internal consistency, Cronbach's α can be used. A high α indicates higher internal consistency, i.e., a high correlation between the scale's items [32]. Thus the items in scales with higher internal consistency seem to measure the same (or closely related) beliefs. If this is the case the scale in question can be considered a reliable measurement of beliefs within the target demographic. Here $\alpha > 0.7$ is considered sufficiently reliable.

Additional quality control was done by examining the discriminatory power of each item. We calculated each item's correlation with the rest of the scale. Items with $r < 0.3$ were dropped from the analysis. This process typically enhances the scale's α.

The reported α's in Table V show the scales after item elimination. Most scales are sufficiently reliable and can thus be used for further analysis. However, from most of the useful scales, one or two items had to be removed due to low discriminatory power.

The other three scales NOSK-SMP, NOSK-JST, and NOSI-TAL could not be enhanced to an acceptable α by removing items. They seem not to be usable with this

TABLE V. Scale analysis: For each scale we show the scale length N and Cronbach's α (standardized) after elimination of items as well as the scale's mean and standard deviation (SD).

Id	Scale	N	α	Mean	SD
NOSK-CRT	Certainty of knowledge	~~6~~5	0.71	2.6	0.70
NOSK-DEV	Development of knowledge	~~8~~7	0.85	3.5	0.44
NOSK-SMP	Simplicity of knowledge	5	0.51	3.4	0.40
NOSK-JST	Justification of knowledge	9	0.59	3.6	0.28
NOSK-SRC	Source of knowledge	~~5~~3	0.77	3.3	0.70
NOSI-PRP	Purpose of the sciences	5	0.71	3.4	0.46
NOSI-TAL	Theories and laws	3	0.66	3.0	0.71
NOSI-MET	Scientific method	3	0.74	2.5	0.77
NOSI-CRE	Creativity and imagination	~~5~~4	0.85	3.6	0.46

TABLE VI. Correlation matrix: For each pair of usable scales ($\alpha > 0.7$) a Pearson correlation coefficient and a p value is reported (***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$).

	NOSK-			NOSI-		
	CRT	DEV	SRC	PRP	MET	CRE
NOSK-CRT	1	0.17	0.26	−0.21	0.22	0.09
NOSK-DEV		1	0.06	0.24	−0.13	0.54***
NOSK-SRC			1	0.03	0.15	−0.01
NOSI-PRP				1	0.57***	0.33*
NOSI-MET					1	−0.21
NOSI-CRE						1

demographic and are thus excluded from further analysis. Incidentally, the latter scale was commented on by the participants in a way indicating a problem with the scales' validity.

On the one hand the correlation matrix (Table VI) indicates a significant correlation between NOSK-DEV and NOSI-CRE and, on the other hand, one between NOSI-PRP and NOSI-MET. This could indicate that at least the other scales measure pairwise different concepts.

C. Adequate views

The used scales are oriented in a way that more adequate views correspond to higher scores on a 4-point scale. Considering the mean scores in Fig. 3 the professors' views on the development of knowledge (NOSK-DEV), its source (DEV-SRC), the purpose of sciences (NOSI-PRP), and on the role of creativity and imagination (NOSI-CRE) lie on average above 3 points on the scale and can thus be considered adequate.

However, there are two interesting exceptions. The mean score for the certainty of knowledge (NOSK-CRT) is only at 2.6 with a rather big variance of 0.7. This indicates that several professors think of physical knowledge as purely objective and clearly demonstrable [39]. This finding is independent of the participants' area of physics (theoretical, experimental, didactics) as indicated by a one-way ANOVA [$F(2, 45) = 2.164$, $p = 0.127$] and also independent of the perceived background in NOS (Pearson correlation with $r = 0.15$, $p = 0.32$).

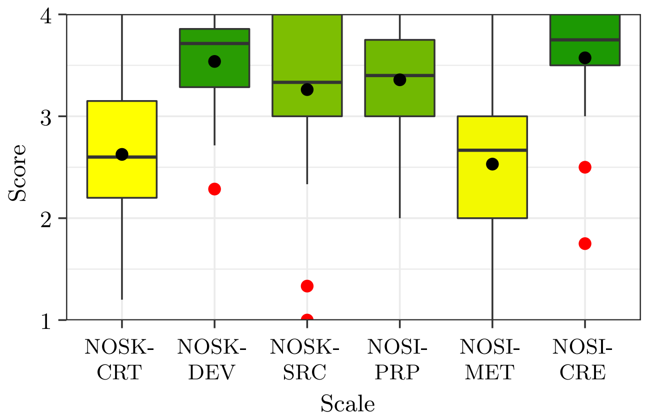


FIG. 3. Adequateness of beliefs. Higher scores indicate more appropriate beliefs in the respective areas. The horizontal bars indicate the scales' median. The black dot indicates the scales' mean as in Table V.

Similarly, their views on the scientific method (NOSI-MET) are only 2.5 on average with a variance of 0.77. According to the scale's author, this might indicate that several professors regard experiments as the sole center and origin of scientific insights [40]. Again this is independent of the participants' area of physics [$F(2, 45) = 0.019$, $p = 0.981$] and of the perceived background in NOS ($r = 0.15$, $p = 0.33$).

These two findings portray the respective participants as tending towards a naive empiricism. However, these two scales do not correlate significantly (cf. Table VI).

D. Importance of NOS when studying physics

We first investigate how important the professors consider learning about NOS during university physics courses *in general*. For a first overview we build a single scale of all content items. Here one item was removed due to low discriminatory power. The resulting scale is sufficiently reliable (Table VII). For some more details we additionally generate two scales from the items about the relevance of NOSK and NOSI aspects, respectively (Table VII).

On this basis we can say that the professors consider learning about NOS at university as highly important. The difference between NOSK and NOSI is not significant [two-sample t test: $t(92.408) = 1.967$, $p = 0.052$] but might become significant with a larger sample size.

Next we consider each NOS aspect on its own. Figure 4 shows that several aspects are perceived as more important than others. An analysis of variance (ANOVA) shows a highly significant effect of the aspect on the perceived importance [$F(8, 431) = 26.63$, $p < 0.001$].

The least important aspect seems to be learning about the influence of cultural differences and the possibility of

TABLE VII. Results from the scales about teaching NOS in general and in the participants' own classes. For each scale we show the scale length N and Cronbach's α (standardized) after elimination of items as well as the scale's mean and standard deviation (SD).

Aspects		N	α	Mean	SD
NOS	in general	1̸817	0.91	3.1	0.44
NOSK	in general	10	0.84	3.0	0.42
NOSI	in general	8̸7	0.81	3.2	0.51
NOS	in own class	18	0.90	2.4	0.53
NOSK	in own class	10	0.85	2.3	0.58
NOSI	in own class	8̸7	0.76	2.7	0.55

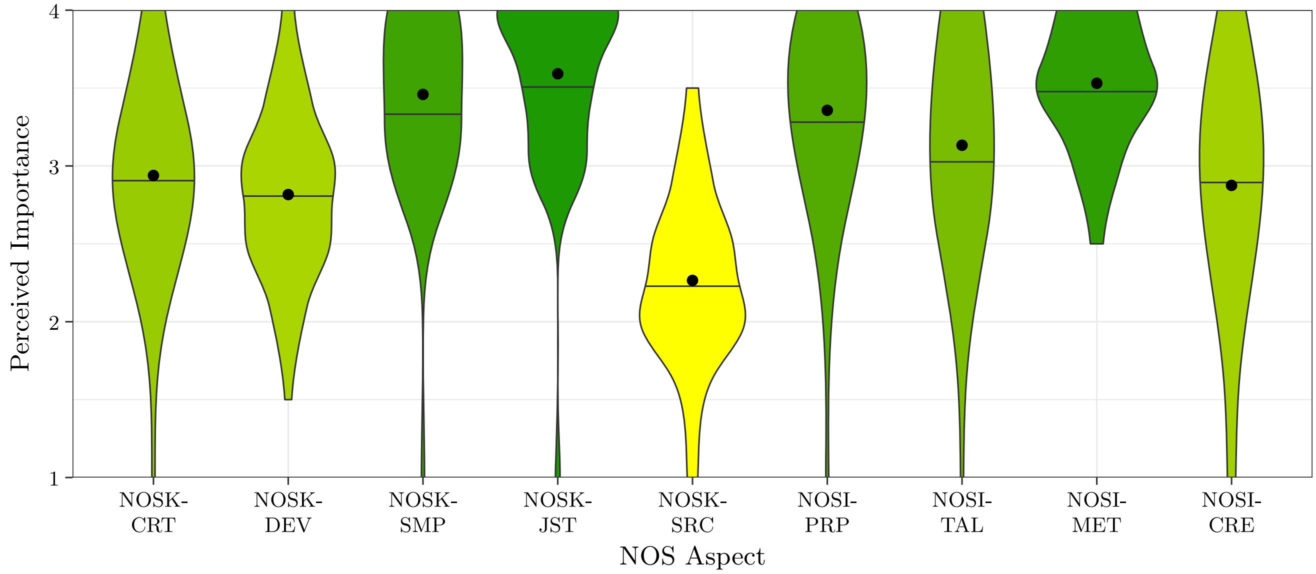


FIG. 4. Perceived importance of learning about the NOS aspects as reported by the professors. The diagram shows the average of the two content items per area on a scale from 1 = not important at all to 4 = very important. The horizontal bar indicates the median. The black dot indicates the arithmetic mean as in Table VIII.

nonscientists contributing to scientific research (NOSK-SRC). Here the pairwise difference to all other aspects is highly significant (all $p < 0.001$).

The aspects perceived as most important to learn about are (in this order) the justification of knowledge (NOSK-JST), scientific methods (NOSI-MET), the simplicity of knowledge (NOSK-SMP), and the purpose of scientific investigations (NOSI-PRP). There are no statistically significant differences between these aspects.

E. Addressing NOS in their own classes

We first construct scales the same way as in Sec. IV D (Table VII). With the removal of one item from the scale the respective reliability is good.

One might observe that here, the average score is almost one unit lower than with the scales for importance, in general. This might be an artifact of the different answer categories—only a few professors regard this content as a core learning objective in their class (56 mentions or

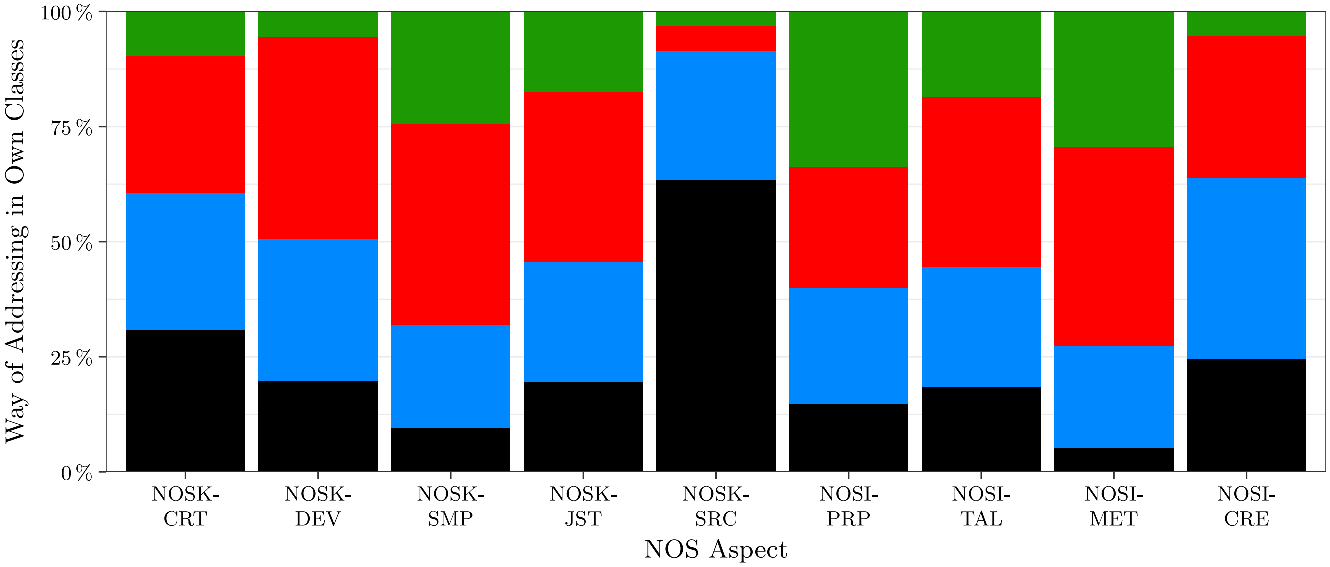


FIG. 5. Reported ways of addressing the given NOS aspects in class. For each aspect the combined percentage for both given content items per NOS aspect is given. Black = not addressed at all; blue = addressed implicitly; red = addressed explicitly; green square = is core learning objective.

TABLE VIII. Comparison between the importance of learning about the NOS areas in general and the reported teaching in the professors' own class. For each aspect mean and SD as well as a Pearson correlation coefficient r and a p value is reported (***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$).

	In general		In own class		Correlation
Aspect	Mean	SD	Mean	SD	r (p)
NOSK-CRT	2.94	0.58	2.21	0.75	0.25
NOSK-DEV	2.82	0.53	2.40	0.76	0.34*
NOSK-SMP	3.46	0.56	2.84	0.76	0.48***
NOSK-JST	3.59	0.56	2.55	0.92	0.42**
NOSK-SRC	2.27	0.53	1.48	0.64	0.48***
NOSI-PRP	3.36	0.61	2.78	0.80	0.31*
NOSI-TAL	3.13	0.68	2.56	0.68	0.41**
NOSI-MET	3.53	0.40	2.98	0.61	0.43**
NOSI-CRE	2.88	0.74	2.18	0.75	0.60***

11.2% in teaching NOSK and 79 mentions or 22.6% in teaching NOSI).

Figure 5 depicts which NOS aspects professors report to teach in their own class. For most aspects there is a lot of variance among the professors. One remarkable aspect is the source of knowledge (NOSK-SRC): 63% do not address whether there is a cultural influence or a contribution of nonscientists in science.

The aspects most named as core learning objectives are the purpose of scientific investigation (NOSI-PRP, 33%), the scientific method (NOSI-MET, 29%), and the simplicity of scientific knowledge (NOSK-SMP, 24%).

For most aspects, the importance in general and the intensity of addressing them in their own class are highly correlated (Table VIII). One notable exception is the certainty of knowledge (NOSK-CRT). These items ask (probably in a more philosophical or epistemological than physical way) whether the students should learn under what conditions physical statements can be classified as secure knowledge or should be enabled to recognize the influence of the researcher on physical findings.

V. DISCUSSION

In this contribution we used scales of well-established test instruments with 50 German university professors in the field of physics. As the scales were not originally devised for this demographic we cannot say for sure whether they are validly usable here. The participants might understand several items differently, e.g., due to a different understanding of the used terms like "experiment," "theory," or "evidence" thus answering differently compared to students.

To gain insight into this problem, we first checked the scales' reliability. In cases where a whole scale has a low reliability (i.e., the items do not correlate and thus Cronbach's α appears low) it cannot be used to reliably measure any construct at all. In these cases—as well as in cases where items were dropped due to low discriminatory power—one might suspect that professors understand them differently than the original target audience.

For the other scales we suspect that the professors understand the items similarly as the original target audience did and (considering the lack of participants' comments stating otherwise and also considering the work that went into validating the scales in the first place [35]) that they can thus be preliminarily interpreted validly (cf. Sec. II C). However, this cannot be seen as a rigorous proof of validity in any way [33].

On this basis we got reliable scales representing 6 aspects of NOS. These indicate mostly high scores that correspond to rather adequate views of nature of science. Two important exceptions are the scales NOSK-CRT and NOSI-MET, which indicate that a bigger portion of the sample hold scientific knowledge as demonstrable by evidence and might possibly adhere to some kind of naive empiricism.

Some of the comments, however, give rise to the impression that the surveyed professors have a more nuanced view of NOS in general than can be contained in a closed-form instrument. This can be interpreted as one (of possibly more) indicator for a belief system that is shaped differently by actual research as a physicist rather than by merely being taught about NOS at school.

The perceived importance of NOS when teaching physics was inquired about in two ways: When asked about the importance of several content items professors ranked them as generally very important. Differentiating between the considered NOS aspects we can identify three groups: The (probably perceived as somehow "soft") source of knowledge is of least importance. Of most importance are aspects that are very closely related to physics content knowledge (simplicity of knowledge, justification of knowledge) or conducting experiments (scientific method, purpose of scientific investigations). The other aspects are rated somewhere in between.

When asked whether these same content items are addressed in their own physics classes they often are not taught at all (like questions of cultural influence or contribution of nonscientists as an extreme example). Only in a few cases such content is seen as a core learning objective by the professors. In general, professors who see a NOS aspect as important to learn in general are also more likely to report teaching that same aspect in their own class.

These findings fit with Redish's notion of epistemology being part of the "hidden curriculum" [6], which is rarely taught explicitly and is far from being a central part of any written curriculum. Instead, it is taught because professors regard it as an important part of physics as a science.

On the basis of this data, we can, however, not state how intensive NOS is *actually* taught or *how* it is taught. It might be the case that the perceived importance was stimulated by the test items or by socially desirable response patterns. Similarly, we only asked the professors if they teach these aspects in their own class. The participants might tell what they would teach in ideal cases (e.g., if they had more time to cover those aspects they regard as important), cf. Ref. [8].

For valid statements about their actual teaching behavior or the effect on their students' views we would have to conduct a more detailed survey. One could examine their lecture notes or videotape their classes to get insights into *what* and *how* NOS is taught. Examining the students' views might be done with the same scales. They could even be asked how often and explicit these aspects are taught in class. On that basis more reliable statements about actual teaching and learning NOS at German universities could be made.

VI. PRACTICAL IMPLICATIONS

As studying physics at university is intertwined with a complex process of becoming part of a culture, we cannot easily regard the professors' views of NOS as irrelevant. We can give implications of the presented results on four levels:

First, we might have ideas of physicists' views of NOS, but there is only little evidence around. With the exception of Schwartz' study [7] the available studies use their views as a baseline to compare, e.g., students' beliefs with Ref. [27]. We might suspect that those studies mainly consider scientists who are somehow experts in this area. There is hardly any evidence concerning the broader masses of physicists. In this study we have established that at least some aspects of NOS can be measured reliably and economically. We should use them to further gain insights into our fellow physicists' beliefs.

Second, we see (and might have suspected) that the surveyed professors mainly hold views that can be considered adequate except that a major portion of the sample shows views that could be interpreted as some kind of naive empiricism—i.e., they put much value in the experiment and they tend to hold scientific knowledge as verifiable by some (probably experimental) means. Although these kinds of views are not considered adequate by us [41] they might very well shape how physicists perceive themselves and their discipline.

We further see that the participants put different emphasis into each aspect of NOS when considering its respective importance. Aspects closely linked to actual research, conducting experiments, or explaining phenomena are considered more important than aspects which are of a "softer" nature like the contribution of different kinds of people to scientific inquiry.

Third, we might reflect on how the professors get their views in the first place. In the U.S., school students are taught the scientific method starting in elementary school. In the curriculum standards, learning about how scientific investigation works is always emphasized. Students use the scientific method for creating a science fair project or for writing a lab report for an in-class experiment. The focus and repetition of the scientific method might lead to an understanding that all science follows the same method. When professors consider the statements of NOSI-MET, we could expect this emphasized idea to resurface. The German educational system, on the other hand, is only starting to be evaluated in this field. But here again the described views of NOS could be seen as a misconception that is developed from a focus on the scientific method at school. Although physics professors are in a field where this view could be challenged, it might also prove quite consistent in their day-to-day work. As they are not science philosophers by profession, this might merely be seen as a rather pragmatic view.

Lastly, there is considerably work to be done validating these instruments. While we assumed the validity of our interpretation based on other authors' work at this point we are not able to give any empirically based assessment on how the professors' views shape their teaching in reality [42], considering that their self-reported teaching habits might as well reflect more on their wishes or perceived importance than on actual practice. We further are very interested in the effects of the professors' views on their students' views. Here we come to the core of the processes when it comes to "becoming a physicist" during one's time at university.

APPENDIX: QUESTIONNAIRE

Please note that the presented test items are translated solely for this publication. In the study they were used in their German original form. For the sources of the scales of part 1 see Table I. The items in part 2 were newly devised based on the description of Kremer [35].

1. Part 1: Individual conceptions

Items that were inverted for analysis are marked (–), dropped items (due to low discriminatory power) are italicized.

NOSK-CRT: Certainty of knowledge

- Physics, like humanities, cannot provide absolute true knowledge.
- In the sciences, valid evidence is relevant, therefore, there is no discussion about what is considered to be secure knowledge. (–)
- Physical knowledge can be clearly demonstrated (e.g., by appropriate experiments) (–)
- *Scientific findings are not purely objective, but also are influenced by the bias of the researchers.*
- Even physical knowledge is not clearly provable and can change over time.
- Physics distinguishes itself from other sciences (e.g., legal studies) in that one can clearly decide

without long discussions whether a theory is right or wrong. (–)

NOSK-DEV: Development of knowledge

- Scientific theories are changed or replaced when new evidence is available.
- Sometimes concepts change in the sciences.
- Sometimes scientists change their mind about what's true in their field.
- New discoveries can change what scientists think is right.
- *There are many questions in the sciences that even scientists cannot answer.*
- Some concepts in the sciences are different today than what scientists used to think.
- The concepts in science books sometimes change.
- Scientific theories change and evolve over time.

NOSK-SMP: Simplicity of knowledge (scale dropped)

- Scientific theories are often more complicated than they should be. (–)
- Scientific theories and laws are more complicatedly formulated than simply. (–)
- The more complicated a scientific theory is, the higher its reputation is among scientists. (–)
- Scientists strive to establish as many theories and laws as possible. (–)
- If two theories equally explain a natural phenomenon, the more complicated theory is the better one. (–)

NOSK-JST: Justification of knowledge (scale dropped)

- Good theories rely on the results of many different experiments.
- When scientists conduct experiments, they determine in advance some aspects of the exploration.
- It is important to have a concrete idea before starting an experiment.
- For scientists, experiments with unexpected results are worthless. (–)
- It is important to do experiments more than once to ensure results.
- The ideas for science experiments come from being curious and thinking about how something works.
- In the sciences, new concepts can emerge from ones own questions and experiments.
- There can be several ways in science to verify concepts.
- An experiment is a good way to find out if something is true.

NOSK-SRC: Source of knowledge

- People without scientific education cannot observe natural phenomena. (–)
- People without scientific education cannot develop scientific research questions. (–)
- Scientific statements can only be verified by scientists. (–)
- *Scientific statements always have a preliminary nature.*
- *Cultural difference are irrelevant to the sciences.*

NOSI-PRP: Purpose of the sciences

- The goal of scientific theory is to give order to part of the human experience.
- Scientists conduct experiments to make new discoveries.
- The goal of scientific theories is to explain natural processes.
- Scientists study natural phenomena and explain why they occur.
- Scientists conduct experiments to explain how certain events come about.

NOSI-TAL: Theories and laws

- Physical theories are true representations of reality. (–)
- A theory is the preliminary stage of a law. (–)
- Theories are not yet proven, laws are fact. (–)

NOSI-MET: Scientific method

- Without the results and data from appropriate experiments, no new physical theories can be established. (–)
- In order to gain new physical insights one has to proceed according to the following method: generation of hypothesis—development of appropriate experiments—observation and evaluation—derivation of laws. (–)
- New theories are always developed from the results of experiments. (–)

NOSI-CRE: Creativity and imagination

- Science theories and laws have nothing to do with creativity. (–)
- Scientific knowledge is also a result of human creativity.
- *Creative thinking is incompatible with logic-based science.* (–)
- The scientific knowledge shows the creativity of scientists.
- The creative thinking of scientists is too untrustworthy to achieve scientific advances. (–)

2. Part 2: Relevance to teaching

The following are the example content items for each NOS aspect. These were presented first under the question "For every item please indicate how relevant it is for students of physics in general." (1 = "not important at all" to 4 = "very important").

After asking for a specific and representative class taught by the test person they were again presented under the question "Please indicate how intensive these topics are addressed in your own class." (1 = "is not addressed," 2 = "is addressed implicitly," 3 = "is addressed explicitly," and 4 = "is core learning objective").

NOSK-CRT 1 Decided under what conditions physical statements are to be classified as secure knowledge.

NOSK-CRT 2 Recognize the influence of the researcher on physical findings.

NOSK-DEV 1 Provide information about the development history of the discussed physical theories.

NOSK-DEV 2 To correctly estimate the temporal stability of physical knowledge.

NOSK-SMP 1 Reasons for the standardization and abstraction of special cases to know general theories.

NOSK-SMP 2 Can specify criteria for the decision between different possible theories.

NOSK-JST 1 Be guided by the experimentation of research questions or hypotheses.

NOSK-JST 2 Understand reasons for repeated measurements.

NOSK-SRC 1 Reflect on the role of cultural influences on the research process.

NOSK-SRC 2 Discuss to what extent all people, not only educated scientists, are or can be involved in research.

NOSI-PRP 1 To know reasons for studying physics.

NOSI-PRP 2 The description, explanation, and prediction of phenomena is an essential goal of scientific research.

NOSI-TAL 1 Explain the difference between laws and theories in science.

NOSI-TAL 2 Describe the relationship of physical theories to reality.

NOSI-MET 1 To know different possible processes from the question to the answer in the physical research process.

NOSI-MET 2 Know the role of experiments as the basis of physical theories.

NOSI-CRE 1 Reflect on the role of creativity in theory formation in physics.

NOSI-CRE 2 Name examples of the interplay of creativity and logic in physics.

---

[1] N. J. Nersessian, Should physicists preach what they practice?, Sci. & Educ. **4**, 203 (1995).

[2] E. Etkina, A. van Heuvelen, S. White-Brahmia, D. T. Brookes, M. Gentile, S. Murthy, D. Rosengrant, and A. Warren, Scientific abilities and their assessment, Phys. Rev. ST Phys. Educ. Res. **2**, 020103 (2006).

[3] D. Hammer, Epistemological beliefs in introductory physics, Cognit. Instr. **12**, 151 (1994).

[4] I. Halloun and D. Hestenes, Interpreting VASS dimensions and profiles for physics students, Sci. & Educ. **7**, 553 (1998).

[5] H. T. Holmegaard, L. M. Madsen, and L. Ulriksen, A journey of negotiation and belonging: Understanding students' transitions to science and engineering in higher education, Cult. Stud. Sci. Educ. **9**, 755 (2014).

[6] E. F. Redish, Introducing students to the culture of physics: Explicating elements of the hidden curriculum, AIP Conf. Proc. **1289**, 49 (2010).

[7] R. S. Schwartz, *Epistemological views in authentic science practice: A cross-discipline study of scientists' views of nature of science and scientific inquiry* (Oregon State University, Corvallis, OR, 2004).

[8] M. Karakas, Cases of science professors' use of nature of science, J. Sci. Educ. Technol. **18**, 101 (2009).

[9] NRC, National Research Council, *National Science Education Standards* (National Academy Press, Washington, DC, 1996).

[10] NRC, National Research Council, *Next Generation Science Standards: For States, by States* (National Academy Press, Washington, DC, 2013), Appendix H: Understanding the Scientific Enterprise: The Nature of Science in the Next Generation Science Standards.

[11] KMK, Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss: Beschluss vom 16.12.2004*, Beschlüsse der Kultusministerkonferenz (Luchterhand, München, 2005).

[12] KMK, Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung* (Bonn, 2008).

[13] KFP, Konferenz der Fachbereiche Physik, *Zur Konzeption von Bachelor- und Master-Studiengängen in der Physik: Handreichung* (KFP, Berlin, 2010).

[14] A good overview that tries to find similarities and differences between these traditions is done (in German) by Neumann and Kremer [26].

[15] B. K. Hofer and P. R. Pintrich, The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning, Rev. Educ. Res. **67**, 88 (1997).

[16] W. G. Perry, *Forms of Ethical and Intellectual Development in the College Years: A Scheme* (Jossey-Bass, San Francisco, 1999).

[17] L. Linsing and A. Elby, The impact of epistemology on learning: A case study from introductory physics, Am. J. Phys. **73**, 372 (2005).

[18] An introduction into these fields could be found in, e.g., A. Rosenberg, *Philosophy of Science* (Routledge, New York, 2012); or L.-G. Johansson, *Philosophy of Science for Scientists*, Springer Undergraduate Texts in Philosophy (Springer International Publishing, Cham, 2016).

[19] N. G. Lederman, in *Scientific Inquiry and Nature of Science: Implications for Teaching, Learning and Teacher Education*, edited by L. B. Flick and N. G. Lederman (Springer, Dordrecht, 2006), pp. 301–317.

[20] J. Osborne, S. Collins, M. Ratcliffe, R. Millar, and R. Duschl, What "ideas-about-science" should be taught in school science? A Delphi study of the expert community, J. Res. Sci. Teach. **40**, 692 (2003).

[21] R. S. Schwartz, N. G. Lederman, and J. Lederman, An Instrument to Assess Views of Scientific Inquiry: The VOSI Questionnaire, *Proceedings of the Annual Meeting of the National Association for Research in Science Teaching, 2008*, http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=85DF376205B50DC841F57C2C2FE756FB?doi=10.1.1.519.2693&rep=rep1&type=pdf.

[22] D. Hodson, *Towards Scientific Literacy* (Sense Publishers, Rotterdam, 2008), ISBN 978-90-8790-505-7.

[23] W. F. McComas and J. K. Olson, in *The Nature of Science in Science Education: Rationales and Strategies*, edited by W. F. McComas (Kluwer, Dordrecht, 1998), pp. 41–52.

[24] N. G. Lederman, F. Abd-El-Khalick, R. L. Bell, and R. S. Schwartz, Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science, J. Res. Sci. Teach. **39**, 497 (2002).

[25] N. G. Lederman, Nature of science: Past, present, and future, in *Handbook of research on science education*, edited by S. K. Abell and N. G. Lederman (Hrsg.) (Lawrence Erlbaum, Mahwah, NJ, 2007), pp. 831–879.

[26] I. Neumann and K. Kremer, Nature of Science und epistemologische Überzeugungen—Ähnlichkeiten und Unterschiede, Z. Naturwissenschaften **19**, 211 (2013).

[27] I. Halloun, Views About Science and physics achievement: The VASS story, in *The changing role of physics departments in modern universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by E. F. Redish and J. S. Rigden (American Institute of Physics, College Park, MD, 1997), pp. 605–614.

[28] B. Priemer, Ein diagnostischer Test zu Schüleransichten über Physik und Lernen von Physik—eine deutsche Version des Tests "Views About Science Survey", Z. Naturwissenschaften **9**, 160 (2003).

[29] I. Neumann, *Beyond Physics Content Knowledge: Modeling Competence Regarding Nature of Scientific Inquiry and Nature of Scientific Knowledge: Dissertation*, Studien zum Physik- und Chemielernen (Logos, Berlin, 2011), Vol. 117.

[30] D. Urhahne, K. Kremer, and J. Mayer, Conceptions of the nature of science—are they general or context specific?, Int. J. Sci. Math. Educ. **9**, 707 (2011).

[31] For a comprehensive treatment of the issue see your favorite book on test theory, e.g., R. P. McDonald, *Test Theory: A Unified Treatment* (Psychology Press, Mahawa, NJ, 2013).

[32] L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika **16**, 297 (1951).

[33] M. T. Kane, Validating the Interpretations and Uses of Test Scores, J. Educ. Measure. **50**, 1 (2013); and other contributions to that special issue.

[34] J. Riese, *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften: Dissertation*, Studien zum Physik- und Chemielernen (Logos, Berlin, 2009), Vol. 97.

[35] D. Urhahne, K. Kremer, and J. Mayer, Welches Verständnis haben Jugendliche von der Natur der Naturwissenschaften?: Entwicklung und erste Schritte zur Validierung eines Fragebogens, Unterrichtswissenschaft—Z. Lernforschung **36**, 72 (2008).

[36] K. H. Kremer, Die Natur der Naturwissenschaften verstehen: Untersuchungen zur Struktur und Entwicklung von Kompetenzen in der Sekundarstufe I, Ph.D. thesis, Universitätsbibliothek Kassel (2010), https://hds.hebis.de/ubks/Record/HEB228930200.

[37] KFP is short for *Konferenz der Fachbereiche Physik*, a meeting of representatives of most German physics departments.

[38] One comment: "Is Newton's law more true than the general theory of relativity? Actually we know that the first is merely an approximation of the second." Another: "The difference between theory and law as implied by the survey is not comprehensible to me."

[39] "Hier wurde erhoben, ob die Probanden physikalische Erkenntnisse als eindeutig und absolut beweisbar (1) ansehen oder nicht (4)." [34], p. 93.

[40] "Hier wurde erhoben, ob die Probanden der Vorstellung eines naiven Empirismus anhängen (1) oder nicht (4)." [34], p. 94.

[41] For logical reasons one might consider Karl Popper's classic *The Logic of Scientific Discovery.*

[42] A model of such research could be the investigation of M. Z. Hashweh, Effects of science teachers' epistemological beliefs in teaching, J. Res. Sci. Teach. **33**, 47 (1996).