

Naïve and informed views on the nature of scientific inquiry in large-scale assessments: Two sides of the same coin or different currencies?

Andreas Nehring 

Institute for Science Education, Leibniz Universität Hannover, Hannover, Germany

Correspondence

Andreas Nehring, Leibniz Universität Hannover, Hannover, Germany.
Email: nehring@idn.uni-hannover.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: NE 2105/1-1

Abstract

Many models in the field of epistemic cognition conceptualize students' views as being on a continuum between the poles of naïve and informed views. Against this background, the aim of the present study was to find out whether views on the nature of scientific inquiry (NOSI views) should be conceptualized and quantitatively assessed in a more multiplistic manner, considering naïve and informed views in their own, separate dimensions. Based on a competence model defining three inquiry methods, we developed a Likert-scaled questionnaire containing 10 scales, each assessing one NOSI view. We administered the questionnaire to a sample of 802 students in the lower and upper levels of secondary school. Based on structural equation modeling, the analyses confirmed a 10-dimensional model, distinguishing between each naïve and informed views as the only adequate representation of the data. Latent class analysis and interview data revealed four profiles of NOSI views in the data, which differed with regard to their agreement or disagreement with different naïve and informed views. We interpret these findings as evidence that supports more multiplistic models, with relevance to conceptualizing, measuring, and fostering NOSI views. We derive future directions of nature of

[Correction added on October 11, 2019, after first online publication: The article title was updated to include a question mark.]

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Author. *Journal of Research in Science Teaching* published by Wiley Periodicals, Inc.

science and NOSI research linking basic and applied research using experimental studies.

KEYWORDS

large-scale assessment, latent class analysis, nature of scientific inquiry, questionnaire, views in science education

1 | INTRODUCTION

Helping students to develop informed views on the nature of science (NOS) and of scientific inquiry (NOSI) has been the subject of science education efforts for over 50 years (Abd-El-Khalick, 2012; Lederman, 2007; Wilson, 1954). However, NOS refers to the characteristics of scientific knowledge and is derived from how the knowledge is developed, NOSI and scientific inquiry (SI) refer to the concrete processes during scientific research activities. SI is concerned about the activities that students are engaged in while conducting scientific investigations, and NOSI refers to their understanding of “how scientists do their work and how the resulting scientific knowledge is generated and accepted” (Lederman et al., 2014, p. 66). NOSI views, the focus of this study, are, thus, oriented toward the practices, tools, and strategies that scientists use to generate or evaluate scientific knowledge. Together with NOS, NOSI views are part of students' epistemological cognition. A well-received conceptualization of NOSI has been suggested by Schwartz, Lederman, and Crawford (2004, p. 68) and Lederman et al. (2014). It differentiates the understanding of NOSI using a multi-dimensional approach including eight aspects of SI such as “inquiry procedures are guided by the question asked”, “all scientists performing the same procedures may not get the same results”, or scientific data are not the same as scientific evidence. As the present study was part of a larger project of research on students' inquiry competences, we conceptualize NOSI using an internationally published competence model (Nehring, Nowak, Upmeier zu Belzen & Tiemann, 2015; Nowak, Nehring, Tiemann, & Upmeier zu Belzen, 2013) and classify 10 naïve and informed NOSI views reported in science education literature (see Section 3).

In order to inform educators, researchers and policymakers on how students understand NOSI or how students react to interventions, research also focusses on creating reliable data collection tools. Although profound developments in interviews and open-ended questionnaires were implemented with a high validity on students' thinking processes (Lederman et al., 2014, 2019), the use of closed-response instruments is still a promising field. Although Likert-scaled tests have been criticized for not taking into account the particular understanding of an individual respondent, as is possible in interview situations, for example, these instruments can be employed to give an overview of NOSI understanding in larger samples or to test hypotheses on correlations to different science education outcomes (Neumann, Neumann, & Nehm, 2011). They may be the method of choice in large-scale situations where the enormous time, personnel, and financial resources would make open-ended instruments too expensive for researchers and particular research questions (Dogan & Abd-El-Khalick, 2008). Which role naïve and informed views play within the context of assessing NOSI views is subject of this study.

1.1 | Multidimensionality of students' cognition and challenges for instrument development in the field of epistemic cognition

As research devoted much attention on the assessment and dimensionality in the field of epistemic cognition (NOSI, NOS, and epistemological beliefs), researchers in these domains were confronted with similar methodological challenges while carrying out studies. Thus, hereinafter, we take methodological issues (based on multidimensional conceptualizations) from these domains also into consideration. This helps us to derive the problem that this study focusses. The readers should note, however, that our approach has a clear focus on NOSI views.

Due to the multidimensional and interlocking nature of understanding in the field of epistemic cognition, the development of instruments is particularly challenging (Harrison, Duncan Seraphin, Philippoff, Vallin, & Brandon, 2015). In this context, some studies treat an understanding as an unidimensional construct (Bell & Linn, 2000; Brunner et al., 2006; Tsai, 1999b; Wenning, 2006). These approaches distinguish between informed or adequate views, corresponding to a rather constructivist perspective on science, and alternative, inadequate or naïve views, corresponding to a rather empiricist perspective (Tsai, 1999a). Understanding is often seen as shifting on a continuum from alternative to informed views. According to the meta-analysis carried out by Greene, Cartiff, and Duke (2018) and a review carried out by Deng, Chen, Tsai, and Chai (2011), studies focusing on multiplicity of views rely on “dimensional models” or correspond to a “multidimension framework.” These studies are based upon the assumption that students' views “may not necessarily develop in a coherent manner, and that the correlations among the dimensions are not precluded” (Deng, Chen, Tsai, & Chai, 2011, p. 970).

The multidimensionality of students' cognition is well documented in research on students' conceptual development indicating that informed and alternative conceptions may coexist contemporaneously (Shtulman & Valcarcel, 2012). Reaction time studies indicate that evaluating unintuitive but scientifically correct statements lead to higher latencies and negative priming (Potvin, Masson, Lafortune, & Cyr, 2015; Shtulman & Valcarcel, 2012). Also, latent class analysis (LCA) on students' response behavior in different science contexts such as floating and sinking (Schneider & Hardy, 2013) or earth science (Straatemeier, van der Maas, & Jansen, 2008) reveals that certain students may hold both informed and alternative views at the same time. This leads to fragmented or heterogeneous reasoning profiles of simultaneously containing adequate and inadequate conceptions (DiSessa, 1993; Thorn, Bissinger, Thorn, & Bogner, 2016).

At first glance, these findings seem to be in line with the multidimensional conceptions of understanding in the field as epistemic cognition. Taking a closer look, however, a number of instruments uses Likert-scaled items being composed of statements that go back to alternative or informed views. Students, holding informed views, are expected to agree, in a larger manner, with informed statements, and to disagree with alternative statements. The items of Chen et al. (2013), Chen, She, Chou, Tsai, & Chiu, 2013, for example, relied on statements such as “there are many possible ways to solve a science problem” (informed view, positively pooled item) or “there is only one method and one set of steps to do an experiment” (naïve view, negatively pooled item) in order to assess views on the theory-ladenness and the coherence and the objectivity of science (p. 424). Similarly, Kampa, Neumann, Heitmann, and Kremer (2016) used statements such as “it is good to try experiments more than once to make sure of your findings” (informed view, positively pooled item) or “only scientists can observe natural phenomena” (negatively pooled) for creating items on the justification and the source of scientific knowledge (p. 85). The Likert-scaled items used by Harrison, Duncan Seraphin, Philippoff, Vallin, and Brandon (2015) contained statements like “scientific knowledge can change over time” (informed view, positively pooled item) or “scientific knowledge is only useful to

scientists” (naïve view, negatively pooled item) (p. 1342). Based on the works of Pomeroy (1993), Tsai (2000) administered items like “legitimate scientific ideas sometimes come from dreams and hunches” (empiricist view, negatively pooled item) or “scientists rigorously attempt to eliminate the human perspective from observations.” (constructivist view, positively pooled item) (p. 199). Using a similar instrument, Huang, Tsai, and Chang (2005) expressed that “Some items in the PNSS stated from an empiricist-oriented (or non-constructivist-oriented) perspective were scored in a reverse manner” (p. 648). These studies implicitly assume that students, holding informed views, would agree on the positively pooled items and disagree on the latter pooled items, and vice versa. In any case, students would have to decide on a particular statement in either a particular naïve, informed sense or a medium category. The items do not allow students to evaluate both views independently.

In multiple-choice instruments, rather naïve and rather informed views are used to create answer options that students have to decide on. In a modified “Views on Science–Technology–Society” version, Dogan and Abd-El-Khalick (2008) asked students whether scientific “observations made by competent scientists will usually be different if the scientists believe different theories.” Students could, *inter alia*, choose between a naïve answer option, “no, because observations are as exact as possible. This is how science has been able to advance”, or an informed option, “yes, because scientists will experiment in different ways and will notice different things” (p. 1108). Ibrahim, Buffler, and Lubben (2009), to give another example, used the Views About Scientific Measurement instrument and asked students to choose an answer option between “nature follows exact laws and scientists discover these laws” and “scientists construct theories to explain what they observe in nature” (p. 251). Again, students would have to decide between different options leading to instruments that are not able to depict the possible multiplistic nature of NOSI views.

2 | STUDY PURPOSE

On the basis of our previous work in the context of SI, the purpose of this study was to find out whether multidimensionality in NOSI assessment, using Likert-scaled items, should be addressed according to a more fragmented profile of naïve and informed views or on a continuum over different dimensions with naïve and informed views being part of this continuum. Although the latter perspective would correspond to the title metaphor of naïve and informed views being two sides of the same coin, the first perspective would correspond to different currencies.

3 | THEORETICAL BACKGROUND

3.1 | Competence-based framework incorporating naïve and informed NOSI views

In order to build upon a theoretical foundation, we used an internationally published competence model (Nehring et al., 2015; Nowak et al., 2013), defining three inquiry methods as a framework for clustering different naïve and informed NOSI views reported in science education literature. The three inquiry methods are “observing as theory-driven activity”, “experimenting as manipulation of variables”, and “using models as tools for inquiry” which have been confirmed, in a validation study, to structure the competences in the field of inquiry by functioning as partial competences.

“Observing as theory-driven activity” means selecting observation criteria based upon theory, testing a hypothesis using collected data, carrying out theory-driven observations using auxiliary equipment if necessary, or deriving classifications if more than two objects or phenomena were

observed (Gropengießer, 2009; Hofstein & Lunetta, 2004). “Experimenting as manipulation of variables,” in the sense of the competence model, corresponds to applying the control-of-variables (CVS) strategy (Chen & Klahr, 1999; Schwichow, Croker, Zimmerman, Höffler, & Härtig, 2016) in science experiments. Going beyond observations, this means to identify dependent, independent, and control variables, to actively intervene with research objects, to manipulate variables, and to keep control variables constant “Using models as tools for inquiry” corresponds to using models as research tools by carrying out an investigation with the help of models (Crawford & Cullin, 2005; Krell, Upmeier zu Belzen, & Krüger, 2014; Schwarz et al., 2009; Upmeier zu Belzen & Krüger, 2010). That means formulating hypotheses based on models and carrying out model experiments, drawing conclusions on the underlying original theory by using substitute objects, and testing the model with regard to the data of the underlying original object.

Based on these three partial competences, we identified five rather naïve and five rather informed NOSI views. We chose views that would enable or hinder students to more or less successfully implement the inquiry methods during scientific investigations (Table 1). We are aware that we could have chosen many more views from the huge body of NOSI literature. Due to practical reasons and sample requirements, however, we limited our selection to 10 views that seemed particularly relevant to enabling or hindering students in implementing the inquiry methods from a theoretical point of view.

Following the approaches in the NOS literature, naïve views correspond to a more positivist, unsystematic, or empiricist perspective on science, whereas informed views correspond to a more constructivist, systematic, and relativist perspective on science (for further details, see Abd-El-Khalick, 2012; Demir & Abell, 2010; Deng, Chen, Tsai, & Chai, 2011; Lederman et al., 2014; Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003). What is more, the criterion of choosing the views was that implementing one of the three methods according to the naïve views would decrease the chance of gaining valid results in scientific investigations, whereas following the informed views would increase them.

TABLE 1 Overview of the naïve and informed NOSI views included in this study

Inquiry method	Corresponding informed NOSI views (rather hypothetico-deductive, systematic, and constructivist)	Corresponding naïve NOSI views (rather positivist, unsystematic, or empiricist)
Observing as theory-driven activity	(1) Scientist are guided by ideas and plan observations. (2) Scientists interpret observations using theory.	(6) While observing, scientist try out and see if it works. (7) Scientists discover laws or theories within observations (randomly). (8) Observations that do not confirm a hypothesis, are useless.
Experimenting as manipulation of variables	(3) Scientists change only one variable at a time for valid experiments	(9) Scientists change many variables at a time for valid experiments
Using models as tools for inquiry	(4) Scientists carry out investigations on models. They test hypotheses about an original object using models. (5) Scientists use models as tool for communicating their investigations and results.	(10) Scientist use models as an exact copy of reality.

Abbreviation: NOSI, nature of scientific inquiry.

As shown in Table 1, we numbered the views from 1 to 10. In the following, we explain the 10 views, give examples for questionnaire items, and rationalize connections between views and competences. However, views, as operationalized in our study, tend to be more on a meta-level, competences relate to being able to solve concrete problems that occur during scientific investigations. We suppose that views and competences are both interrelated on different levels. On an intra-individual level, the interplay of views and competences might be complex including several processes. When it comes to solve problems that occur during scientific investigations views might contribute to a meta-knowledge about the inquiry methods (Künsting, Wirth, & Paas, 2011). Students holding adequate views might be able to activate and to apply certain these views in order to use them as research strategies making them more successful during scientific investigations. This might help them to implement certain steps that contribute to more expedient problem solving. Moreover, acting according to one's views contributes to a coherent perspective on the own actions, whereas acting in disaccord to views might induce a feeling of cognitive dissonance. Finally, research suggests that students develop views from inquiry activities—especially under the condition of explicit reflection (Khishfe & Abd-El-Khalick, 2002; Vorholzer, von Aufschnaiter, & Boone, 2018). More adequate and less naïve views might, of course, be the result of inquiry-based learning activities in school science and therefore relate on the another. In each case, literature reviews articulate relating views and inquiry competences as one of the major gap in the field of research on inquiry (Rönnebeck, Bernholt, & Ropohl, 2016).

The readers should note, however, that this study is on conceptualizing views on an inter-individual sample level. The question of how views and competences play together on an intra-individual (and not only interindividual) level or the question of how to model views and competences qualitatively or quantitatively may be subject to future studies (see Section 6). These are the views included in this study:

1. *Scientist are guided by ideas and plan observations*: This informed view corresponds to a rather systematic and knowledge-based perspective on science with investigations not just following a random procedure. Although, there is no single scientific method, observations are guided by questions (Lederman et al., 2014) and scientist have an idea that guides them during the planning process (Carey, Evans, Honda, Jay, & Unger, 1989). Observing corresponds to a planned behavior and not just to “looking.” Example item in our Likert-based questionnaire: When scientists carry out an investigation, they want to find out if their assumption matches the observations or measurements. *Connection to the partial competence “observing as theory-driven activity”*: This view describes the process of the inquiry method of “observing” almost directly. As observing is not just “looking” but a planned and theory-based activity, it is likely that students holding this view are rather able to carry out investigations that require a planned and theory-based approach. They may use this view to solve problems based on a planed behavior and not apply less successful strategies as try-and-error procedures, for example.
2. *Scientists interpret observations using theory*: This informed view refers to the theory-laden NOS indicating that existing knowledge guides observations and helps to interpret data derived from observations. Observations and inference are distinct (Lederman, 2007) and observations can have meaning for theory. Example item in our Likert-based questionnaire: When scientists carry out an observation or a measurement, a theory helps them to understand the results of an observation or measurement. *Connection to the partial competence “observing as theory-driven activity”*: This view also refers to a particular property of the inquiry method “observing.” Students holding this view have understood that scientific knowledge should be used in order to

make sense of what can be observed on a phenomenological level. If they distinguish between observation and inference, they are more successful in interpreting data from observations or measurement. Holding this view could consequently be a factor supporting students in interpreting the behavior of a phenomenon with regard to scientific theories during investigations.

3. *Scientists change only one variable at a time for valid experiments*: This view corresponds to the widely researched CVS (Chen & Klahr, 1999; Schwichow, Croker, Zimmerman, Höffler, & Härtig, 2016) where isolating and controlling variables leads to valid results within experiments. Example item in our Likert-based questionnaire: When scientists carry out experiments to find out whether a certain variable affects a property, they change only one variable at a time. *Connection to the partial competence “Experimenting as manipulation of variables”*: This view is very close to a concrete strategy that can be applied for carrying out controlled experiments. It is evident that an understanding of the CVS should help students to identify, vary, and control variables purposefully.
4. *Scientists carry out investigations on models*: This informed view on modeling procedures corresponds to the research-based aspects of modeling practice (Grosslight, Unger, Jay, & Smith, 1991; Upmeier zu Belzen & Krüger, 2010). Students holding this view would agree that scientists carry out observations and experiments on models in order to explain and to predict properties of an original phenomenon. Example item in our Likert-based questionnaire: When scientists work with a model, they test an assumption with the help of the model. *Connection to the partial competence “Using models as tools for inquiry”*: An understanding that models are used as tools for inquiry is a prerequisite for deriving new knowledge based on models. Students should understand that models are theoretical reconstructions that can be used auxiliary tools for research. Within the model competence as conceptualized by Upmeier zu Belzen and Krüger (2010) or Krell, Upmeier zu Belzen, and Krüger (2014), p. 114, this view refers to “to the ‘predictive nature of models’ which reflects the scientific perspective of models as research tools”. It goes hand in hand with a higher model competence indicating that students might be more able to perform investigations based on models.
5. *Scientists use models as tools for communicating their investigations and results*: This informed view corresponds to the communicative aspects of modeling practice (Schwarz et al., 2009). Scientists use models as tool for presenting ideas or results. Example item in our Likert-based questionnaire: When scientists work with a model, they explain an original to other scientists. *Connection to the partial competence “Using models as tools for inquiry”*: This view corresponds to the second important perspective of models and model competence (Krell, Upmeier zu Belzen, & Krüger, 2014; Upmeier zu Belzen & Krüger, 2010). Besides the perspective of models as tools for inquiry, scientists use models as tools for communicating. Although this view is not conceptualized as being part of higher levels of model competence, it is part of a more holistic understanding of the role models play in science. This view might go hand in hand with a more holistic understanding of the role models play in science so that it might be easier for students holding this view to work with models in the context of inquiry.
6. *While observing, scientist try out and see if it works*: This naïve view goes back to the study of Carey, Evans, Honda, Jay, and Unger (1989). It includes views on scientists who “tries something to see if it ‘works’ or ‘reacts’ or to ‘find out about the thing they’re experimenting on’” (p. 523). Carrying out observations is seen as a random procedure. Example item in our Likert-based questionnaire: When scientists carry out an investigation, they start the investigation without preparation. *Connection to the partial competence “observing as theory-driven activity”*: Students holding this view might not proceed in a very planned manner. If they apply a try-out-

and-see strategy, it is likely that they will not be able to solve problems during investigations as successfully as students that would be aware of a planned approach.

7. *Scientists discover laws or theories within observations (randomly)*: This naïve view is contrast with the theory-laden nature of observations (Chen, She, Chou, Tsai, & Chiu, 2013) and corresponds to an empiricist view on SI. Students holding this view would think that scientists carry out their investigations without a theory or an idea behind it so that scientists would simply discover theories within observations. Example item in our Likert-based questionnaire: When scientists carry out an investigation, they discover the result without a theory. *Connection to the partial competence “observing as theory-driven activity”*: Students holding this view might not proceed in a very planned manner. If they apply a try-out-and-see strategy, it is likely that they will not be able to solve problems during investigations as successfully as students that would be aware of a planned approach.
8. *Observations that do not confirm a hypothesis are useless*: This view corresponds to the well-documented confirmation bias where students seek to confirm a hypothesis even under conflicting findings or think that the purpose of investigations is to seek for confirmation (Eberbach & Crowley, 2009; Poletiek, 1996). Example item in our Likert-based questionnaire: When scientists carry out an investigation, the investigation is worthless if it contradicts their assumptions. *Connection to the partial competence “observing as theory-driven activity”*: Also, this view might hinder students to carry out purposeful investigations. If students proceed accordingly, they apply the confirmation bias and only seek for observations that confirm their existing knowledge. This might hinder them to test hypothesis or to generate further hypothesis that go beyond their existing knowledge (Chinn & Brewer, 1993).
9. *Scientists change many variables at a time for valid experiments*: This naïve view corresponds to experimenting as an activity in which the scientist, for example, changes many variables at once to create a certain effect, for example. Often, this goes hand in hand with a rather engineering mode of experimentation, rather than a science mode where the interplay between observations and ideas is central (Schauble, Klopfer, & Raghavan, 1991). Example item in our Likert-based questionnaire: When scientists carry out experiments to find out whether a certain variable affects a property, they change all the variables at once that could have an effect at once. *Connection to the partial competence “Experimenting as manipulation of variables”*: This view corresponds to approaches of confounding experiments and changing too many variables at a time. Students that would proceed accordingly would not be able to design valid experiments (Kuhn & Dean, 2005).
10. *Scientist use models as an exact copy of reality*: This naïve view corresponds to the well-documented conception of models being an enlarged or a reduced copy of an original. Accordingly, models should match the original best in all properties (Grosslight, Unger, Jay, & Smith, 1991; Grünkorn, Upmeier zu Belzen, & Krüger, 2014). Example item in our Likert-based questionnaire: When scientists work with a model, the model displays all the characteristics of the original. *Connection to the partial competence “Using models as tools for inquiry”*: Thinking of models as an exact copy the reality is associated with the lowest levels of model competence (Krell, Upmeier zu Belzen, & Krüger, 2014; Upmeier zu Belzen & Krüger, 2010). Although this view might not be associated with a lower performance while working on one concrete model, students, holding this view, might be less successful when it comes to change a model with regard to underlying theoretical assumptions. Also, choosing a suitable model out of several candidate models on the same phenomenon will not be that easy for these students. This might reduce competences as a situation that requires procedures that contradict this view.

3.2 | Deriving unidimensional and multidimensional candidate models

In the following, we stick to the two-step approach supposed by Burnham and Anderson (2002) of formulating of a set of candidate models and selecting a model to be used in making inferences. This approach was also applied by Neumann, Neumann, and Nehm (2011) and Harrison, Duncan Seraphin, Philippoff, Vallin, and Brandon (2015) in the context of NOS and NOSI views. In the present study, each model represents different implicit assumptions about the interrelatedness of naïve and informed NOSI views, the understanding of NOSI, and to question whether it is appropriate to picture a simultaneous presence of naïve and informed views in assessment. We derived five models (for an overview, see Table 2). Note that the metaphorical description (finance metaphor) of the model assumptions refers to a Likert-scaled assessment assigning a “more or less” to students. We are aware that understanding is a much more complex process than shifting between two poles or having more or less money, for example. We think, however, that this is a helpful tool for communicating the essence of the model. The models

TABLE 2 Overview on the candidate models being tested in this study

Model	Dimensionality	Underlying assumption about the interrelatedness of views and the understanding of NOSI	Description within the metaphor
1	Unidimensional	<ul style="list-style-type: none"> • No particular structure between naïve and informed views. • Understanding as shifting on a continuum between naïve and informed views. 	Understanding as having more or less money
2	Two dimensional (naïve and informed views)	<ul style="list-style-type: none"> • Distinguishing between naïve and informed views. • Understanding as promoted informed and decreased naïve views. 	Naïve and informed views as two sides of the same coin
3	Three dimensional (inquiry methods)	<ul style="list-style-type: none"> • Distinguishing between naïve and informed views. • Understanding as promoted informed and decreased naïve views. 	Understanding as more or less money in three currencies
4	Six dimensional (naïve and informed views within inquiry methods)	<ul style="list-style-type: none"> • Distinguishing between naïve and informed views within inquiry methods. • Understanding as promoted informed and decreased naïve views within inquiry methods. 	Naïve and informed views as different sides of a coin in three currencies
5	Ten dimensional (each view corresponds to one dimension)	<ul style="list-style-type: none"> • Distinguishing between 10 naïve and informed views. • Understanding as promoted informed and decreased naïve views within 10 dimensions. • Allowing fragmented view profiles in a detailed manner. 	Naïve and informed views as different sides of a coin in 10 different currencies

Abbreviation: NOSI, nature of scientific inquiry.

derived and tested in this study are shown in Table 2. This table shows the underlying assumption about the interrelatedness of views and the understanding of NOSI as well as the implication on the metaphor appearing in the title.

4 | METHODS

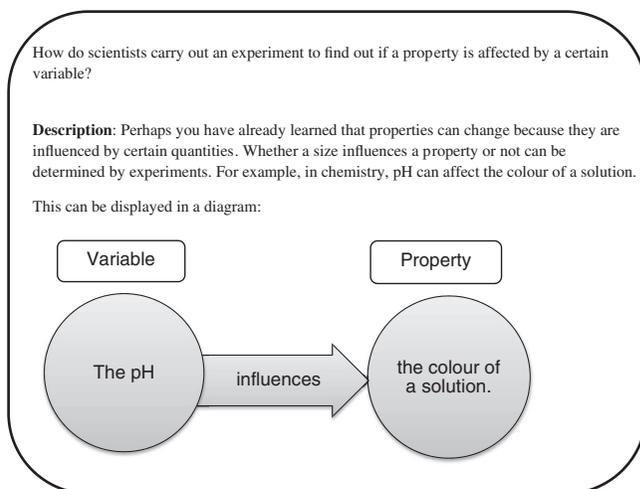
4.1 | Developing the questionnaire

In order to gain data on students' views, we developed a 4-point Likert-scaled questionnaire (4 = "strongly agree," 3 = "agree," 2 = "disagree," 1 = "strongly disagree") containing 47 items in total. By using a 4-point scale, the students are forced to decide on agreeing or disagreeing, there is no neutral position that is harder to interpret. Nine views were assessed using five-item scales each. Due to a very good performance in the pre-studies (see below), one view ("scientists change only one variable at a time for valid experiments") was assessed using a three-item short scale. Each scale was laid out within a common table, leading to a block of five items or three items. The students could express their agreement or disagreement on a 4-point scale. Short texts explaining the characteristics of the inquiry methods to the students were presented before three- or four-item blocks each. In order to contextualize this information, a school example from chemistry or biology was given. This was due to fact that we applied and validated the competence model in the contexts of these two disciplines. The questionnaires contained either the chemistry or the biology examples and were assigned randomly to the students. Figure 1 shows this example for the inquiry method "experimenting." In order to distribute possible sequence effects across the scales, we created eight booklets containing different sequences of inquiry methods, and different sequences of scales on naïve views, following scales on adequate views and the other way around.

4.2 | Qualitative and quantitative prestudy

The questionnaire was developed in two draft versions and evaluated in two preliminary studies, and was also judged by linguists and science educators. The first draft version contained 57 Likert items in total with 31 items going back to two informed views (Views 2 and 4) and 26 items going back to

FIGURE 1 Example text for communicating the characteristics of the inquiry method experimenting containing the chemistry example



two naïve views (Views 6 and 4). The wording of the items was kept as close to the students' wording as possible. Teachers and science educators were asked to evaluate the wording. The basic structure was the same as described above.

For the quantitative prestudy, the first draft questionnaire was administered in a sample of $n = 135$ students from the 9th and 10th grades in two secondary schools in the federal states of Berlin and Brandenburg in 2015. It was evaluated with regard to internal consistency, to item discrimination, and factor loadings in an exploratory factor analysis (EFA). The results showed 43 items having an item discrimination of $r_{it} > .20$. The internal consistency of a scale composed of these items was at Cronbach's $\alpha = .85$. EFA using the method of parallel analysis (Field, 2009; O'Connor, 2000) indicated a four-factor solution with loadings that ranged, after promax rotation, between $.33 < \lambda_i < .82$. Based on these results, we used these 43 items for further item development in a second draft version and constructed, on this basis, new items for further views.

In order to ensure that the wording in the items met the understanding of the students in our sample, the second draft version was evaluated in a qualitative study using think aloud protocols and semistructured interviews. Four students from the 8th grade, two students from the 10th and two students from the 11th grades (eight students in total) were asked to fill out the questionnaire in a one-to-one situation with the interviewer. While filling out the questionnaire students were asked to think aloud in order to gain data about their reasoning while choosing answer options. The interview sessions were carried out in September 2016. All students were interviewed in an individual setting. After a block of 5 or 10 items each, the students were explicitly asked about their understanding of central notions such as “model,” “original,” “property,” “theory,” “assumption,” “hypotheses,” or “variable” and justified the choice of their decisions. The students were also asked to reword the introductory texts. The interviews were analyzed with special regard to cognitive validity (Thelk & Hoole, 2006). Therefore, we tried to gain data on whether the reasoning processes correspond to the construct the items pretend to assess—which in our case were NOSI views. This means that for each item, we analyzed whether the agreement or disagreement on the Likert-scale was in line with the reasoning processes or the justifications made by the students. In order to perform the analysis, we used the software ELAN that is freely available and supplied by the Max Planck Institute for Psycholinguistics (Hellwig, 2018). Using a simple dichotomous two-step classification system, we classified whether a student agreed or disagreed on a certain item and whether the agreement or disagreement involved reasoning that was associated with the view the item is supposed to assess (process validity given) or whether the reasoning showed aspects that had nothing to do with the view, such as misinterpretation of wording, for example (process validity not given). Notions that were not understood by students or that showed potential to bias the response behavior were replaced. Items that showed hints for lowering validity were reworded after an interview so that each student read a slightly adapted version of the questionnaire. The interview questions concerning central notions and introductory texts, however, stayed the same. After a final iteration, the questionnaire was evaluated by a team of linguists of the Leibniz Universität Hannover that are specialized in German as second language (see Acknowledgement section). Their analysis focused on barriers for understanding for students with a migration background. In particular, recommendations for simplifying syntax were made and adopted. Both pre-studies suggested the main study could be carried out with an instrument having a comparatively high degree of reliability, objectivity and validity. It should however be noted that this validated version is in the German language and the translations in this study, that readers find in the Supporting Information Appendix, would have to be revalidated if a reader would want to use the questionnaire.

4.3 | Sample

During the main study, data were collected in secondary schools with the highest educational track (“Gymnasien”) in Hannover (Germany) and in Berlin (Germany) in 2017. In total, 802 students coming from the lower and upper secondary levels participated in the study. The students were aged 13 and 18. 55% of the students were girls. Table 3 gives an overview of the sample.

The sample is a convenience sample that comprises students in schools. These students were involved in school science learning during the period of data collection. Consequently, they were part of a population of interest. As for medical or psychological experiments, convenience samples allow testing hypothesis on correlations or identifying types of persons, which is the purpose of the study, but they do not allow conclusions on a population level, which was not the purpose of this study. Each student participated voluntarily and anonymously in the study. The study was carried out according to the legal requirements of the federal states of state of authors and Berlin. The questionnaire was administered in chemistry or biology lessons with the consent and the presence of the chemistry or biology teachers.

4.4 | Data analysis

The analysis was carried out in R package lavaan (Rosseel, 2012) in Version 0.6–1.1141). A robust maximum likelihood estimator was used to deal with nonnormal distributions. Missing data were handled with the help of the full maximum likelihood algorithm (Collins, Schafer, & Kam, 2001). According to the recommendations of Little (2013) and Urban and Mayerl (2014), we used measurement models of three items per view in order to avoid biased evaluations of models fit. Three items build up a just identified measurement model and help to avoid arbitrarily improved model fit. We choose the items which had the highest factor loadings in measurement evaluations for every single view. Three reasons justify our three-item approach. First, the items within each scale are quite homogeneously formulated (like if scientists carry out an experiment to find out if a variable affects a property, then, Item no. 1: ...they change all the variables that could have an impact once; Item no. 2: ...they change all the variables that could have an impact once; Item no. 3: ...they change all the variables that could have an impact once). We sure to maintain content validity concerning each view also with three items. Second, the analysis of item discriminations (corrected item–total correlations) for each of the 10 scales revealed an average (calculated as median) discrimination $r_{\text{mdn}}(\text{items}) = .51$, confirming a comparatively homogenous functioning of the scale. Third, reliability analysis showed that the McDonald's ω ranges between $0.64 < \omega < 0.73$ indicating that we have—for a relatively “fuzzy” construct like views—quite homogeneous three-item scales. As we administered Likert items, the probability of having one student marking a “I agree” (= 3) on one item and choosing “I agree” (= 3) on the other items is quite high. As we then used the scale means (and not the scale's sum) for the LCA analysis, we assume to not lose much information about the students' views using this approach. Third, as shown in Section 5, the factor loadings are comparatively in a very good range. The items remained stable in all of the five final model estimations. The readers should

TABLE 3 Sample composition

Class level	8	9	10	11	12	Total
Age, mean (SD)	13.83 (0.39)	14.58 (0.77)	15.66 (0.70)	16.74 (0.72)	17.17 (1.53)	15.42 (1.19)
<i>N</i>	24	344	229	192	13	802

note that we carried out the structural equation modeling (SEM)-based analysis also using five items. The results and implications are in line with the reported results in the following and are part of the Supporting Information Appendix.

In order to evaluate the models' fit, we used the comparative fit index (CFI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual. While a value of RMSEA <0.05 indicates a good model fit, Schreiber, Nora, Stage, Barlow, and King (2006) argue that RMSEA <0.06 and a CFI of >0.90 (or better of >0.95) should be given. We also relied on the χ^2 -difference test, on the Akaike information criterion (AIC), and the sample size-adjusted Bayesian information criterion (BIC). Latent correlations were used for analyzing the correlation structure between naïve and informed views.

Based on the chosen models, we carried out LCA using MPLUS 8 and fitted mixture models (Hickendorff, Edelsbrunner, McMullen, Schneider, & Trezise, 2018) containing one to eight classes. In order to converge these models, we used 10000 random sets of starting values for the initial stage and 20 iterations for the final stage. We compared the models with regard to their theoretical implications and different empirical indices. These are the AIC and BIC values as well as the Lo–Mendel–Rubin adjusted likelihood ratio test (LRT), the parametric bootstrapped likelihood-ratio test (LMR) and the Vuong–Lo–Mendell–Rubin likelihood ratio test (VL-LRT) (Asparouhov & Muthén, 2012; Nylund, Asparouhov, & Muthén, 2007). Additionally, we used the entropy index, which ranges between 0 and 1, as an indicator for quantifying a more certain classification of single individuals with 1 showing better classifications. Based on these indices and theoretical interpretability, we decided on the number of classes. The approach of using scale means, after having checked for reliability, has been carried out in the field of epistemic cognition by Kampa, Neumann, Heitmann, and Kremer (2016) or in the field of conceptual understanding by Schneider and Hardy (2013).

5 | RESULTS

5.1 | Comparing model fit

Table 4 presents the model fit indices for each of the five models. The results show that only Model 5, including all 10 views as 10 scales, is suitable for fulfilling the mentioned SEM-based fit criteria (see Section 4). Its CFI index is >0.95 , the RMSEA is <0.05 , and it shows comparatively small BIC and AIC indices. Although, it is the most complex model, which could fit just better due to its complexity, the information criteria, which penalizes more complex models, also indicate the acceptance this model as representing the data most adequately.

5.2 | Comparing factor loadings

The standardized factor loadings for each of the five models are presented in Table 5. All loadings are significant on the $p < .001$ level.

The data show that the factor loadings are higher within the higher dimensional models. Items going back to naïve or informed views can have negative loadings. This occurs in the first and third models. These are the models that comprise naïve and informed views in one latent dimension (or in the metaphor: models with naïve and informed views as being two sides of a coin). This would suggest that naïve items can be indicators for informed views. The loadings, however, are lower on a descriptive

TABLE 4 Overview of model fit indices

No.	Model	χ^2	df	CFI ^a	RMSEA ^a	SRMR ^a	AIC	BIC ^b
1	Unidimensional	2,294.30	405	0.562	0.087	0.079	51,290.42	51,711.81
2	Two dimensional (naïve and informed views)	2,021.90	404	0.631	0.071	0.072	50,929.22	51,355.29
3	Three dimensional (inquiry methods)	1,807.09	402	0.680	0.075	0.066	50,650.72	50,790.82
4	Six dimensional (naïve and informed views within inquiry methods)	1,110.36	390	0.837	0.048	0.051	49,778.84	49,937.03
5	Ten dimensional (each view corresponds to one dimension)	564.92	360	0.954	0.030	0.040	49,134.41	49,337.80

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; SRMR, standardized root-mean-square residual.

^aRobust versions.

^bSample size-adjusted version.

level when being used as a negatively pooled indicator in lower dimensional models indicating that these items do not seem to be as adequate indicators in the models which merge naïve and informed views.

5.3 | Analyzing latent correlations

According to this analysis, we used Model 5 in order to analyze the latent correlations (Table 6).

Table 6 shows highly significant positive correlations within the informed and within the naïve views. These correlations range between from a high and to a moderate level. Confirming to our expectations, informed and naïve views are correlated negatively, to a highly significant level. In order to give a more tangible overview of these results, we calculated medians for these correlations. This reveals that informed views are intercorrelated on a higher level (Mdn = 0.53) than naïve views (Mdn = 0.31) or informed and naïve views (Mdn = -0.30). Although informed views and naïve views go hand in hand to a certain extent, and holding informed views reduces the chances of holding naïve views, the correlations are not that high that holding informed views would automatically mean to not holding naïve views.

5.4 | Comparing latent classes in search of profiles of NOSI views

In order to follow-up this interpretation, we tried to identify students' view profiles in a LCA. Table 7 gives an overview of the model fit statistics for the 8 calculated mixture models. Each model contains 10 dimensions according to the results of the SEM-based analysis.

On a descriptive level, the fit indices (log likelihood) and information criteria (AIC and BIC) show the best fit for the model containing eight classes. However, the relative group frequency is comparatively small, and the LMR and VL-LRT indicate that this model does not fit significantly better than the seven-class model does. The same goes for models containing seven, six, and five classes. The indices for Model 4 indicate that this model could fulfill the requirements of a comparatively better fit and a class frequency with a more relevant percentage of students than just 1 or 2 %. With regard to

TABLE 5 Factor loadings for the five candidate models

Item	View	Category	Standardized factor loadings				
			Model 1	Model 2	Model 3	Model 4	Model 5
1	1	Informed	0.57	0.58	-0.59	0.62	0.65
2	1	Informed	0.54	0.55	-0.56	0.61	0.66
3	1	Informed	0.60	0.60	-0.64	0.66	0.73
4	2	Informed	0.49	0.53	-0.49	0.54	0.60
5	2	Informed	0.45	0.50	-0.44	0.51	0.69
6	2	Informed	0.43	0.48	-0.43	0.49	0.64
7	3	Informed	0.45	0.45	-0.68	0.72	0.73
8	3	Informed	0.48	0.50	-0.56	0.65	0.65
9	3	Informed	0.46	0.47	-0.57	0.62	0.62
10	4	Informed	0.49	0.53	0.57	0.61	0.70
11	4	Informed	0.45	0.50	0.55	0.59	0.72
12	4	Informed	0.33	0.39	0.41	0.46	0.49
13	5	Informed	0.46	0.49	0.56	0.56	0.68
14	5	Informed	0.41	0.43	0.55	0.50	0.58
15	5	Informed	0.45	0.49	0.53	0.55	0.64
16	6	Naïve	-0.47	0.53	0.50	0.54	0.61
17	6	Naïve	-0.47	0.38	0.53	0.61	0.67
18	6	Naïve	-0.52	0.28	0.56	0.63	0.72
19	7	Naïve	-0.49	0.51	0.52	0.54	0.58
20	7	Naïve	-0.34	0.56	0.36	0.38	0.65
21	7	Naïve	-0.22	0.60	0.24	0.28	0.54
22	8	Naïve	-0.16	0.23	0.18	0.24	0.25
23	8	Naïve	-0.50	0.56	0.53	0.57	0.81
24	8	Naïve	-0.47	0.55	0.51	0.65	0.74
25	9	Naïve	-0.30	0.34	0.52	0.56	0.55
26	9	Naïve	-0.36	0.38	0.62	0.79	0.78
27	9	Naïve	-0.30	0.35	0.55	0.70	0.71
28	10	Naïve	-0.29	0.36	-0.28	0.69	0.69
29	10	Naïve	-0.32	0.39	-0.34	0.73	0.73
30	10	Naïve	-0.33	0.38	-0.35	0.64	0.63
			Medians of standardized factor loadings				
			Model 1	Model 2	Model 3	Model 4	Model 5
			0.09	0.49	0.39	0.60	0.66

these criteria, we decided to choose this model for depicting profiles between naïve and informed views. Based on mean scores for each of the 10 dimensions, Figure 2 presents these profiles.

TABLE 6 Latent correlations between the 10 naive and informed views 1–10

1	2	3	4	5	6	7	8	9	10
Planned observations	Theory-based observations	Scientists change only one thing at a time	Models as tools for investigations	Models as tools for communication	Scientists try out and see if it works	Scientists discover laws or theories	Unconfirmed hypothesis are useless	Scientists change many variables at a time	Models as copy of reality
1	–								
2	–0.59								
3	–0.52	–0.47							
4	–0.55	–0.39	–						
5	–0.51	–0.55	–0.45	–					
6	–0.59	–0.42	–0.33	–0.44	–0.37	–			
7	–0.45	–0.44	–0.30	–0.32	–0.52	–			
8	–0.55	–0.28	–0.33	–0.34	–0.54	–0.31	–		
9	–0.23	–0.23	–0.16	–0.23	–0.27	–0.39	–0.26	–	
10	–0.30	–0.16	–0.16	–0.28	–0.31	–0.24	–0.37	–0.20	–

Note: All correlations are significant on the $p < .001$ level, except the correlations between Views 2 and 10, Views 4 and 10 ($p = .002$) as well as between Views 3 and 7, 10 and 7 ($p = .001$).

TABLE 7 Comparison of fit statistics, indices, and likelihood ratio tests for LCA

No. of classes	Log likelihood	No. of free parameters	AIC	BIC	LMR	BLRT	VLRT	Entropy	Smallest group frequency (%)
1	-43,623.97	20	87,287.95	87,318.08					
2	-43,073.17	31	86,208.33	86,255.03	0.000	0.000	0.000	0.93	13.4
3	-42,850.57	42	85,785.14	85,848.42	0.001	0.000	0.001	0.79	9.0
4	-42,717.74	53	85,541.47	85,621.32	0.050	0.000	0.048	0.84	6.1
5	-42,614.81	64	85,357.63	85,454.05	0.053	0.000	0.051	0.86	1.6
6	-42,542.36	75	85,234.72	85,347.72	0.066	0.000	0.065	0.88	1.5
7	-42,488.45	86	85,148.90	85,278.47	0.240	0.000	0.240	0.83	1.5
8	-42,438.05	97	85,070.10	85,216.23	0.650	0.000	0.650	0.87	1.5

Abbreviations: AIC, Akaike information criterion; BIC, Bayesian information criterion; BLRT, parametric bootstrapped likelihood ratio test; LCA, latent class analysis; LMR, Lo-Mendell-Rubin adjusted likelihood ratio test; VL-LRT, Vuong-Lo-Mendell-Rubin likelihood ratio test.

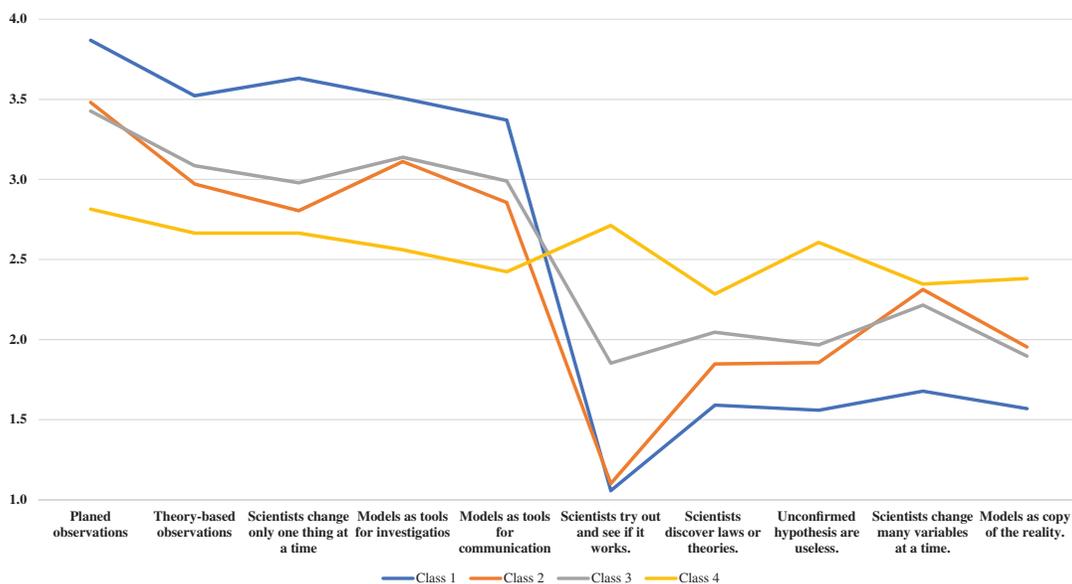


FIGURE 2 Mean scores for latent profiles of NOSI views using four classes. NOSI, nature of scientific inquiry [Color figure can be viewed at wileyonlinelibrary.com]

Figure 2 shows that the four profiles differ regarding the extent of agreement or disagreement with naïve and informed views. Although Class 1 students are able to distinguish clearly between naïve and informed NOSI views, Class 4 students seem to consider naïve as well as informed views as equal plausible when it comes to describe a scientists' inquiry. Class 3 students seem to be unsure about the informed views as well as whether scientists try something out and see if it works. In this dimension, they differ compared to Class 2 students. Figure 2 shows that the profiles overlap and do

not simply follow the same trend as different levels on a continuum which, again, confirms a certain multiplicity of naïve and informed NOSI views.

In Supplementary Information Appendix, we use interview data from the qualitative prestudy in order to provide insights to the reasoning processes that underlie students' response behavior and belonging to a certain profile.

6 | DISCUSSION AND IMPLICATIONS

6.1 | Summary

With regard to NOSI approaches conceptualizing views in this field as a continuum between naïve and informed views, we examined whether naïve and informed views should be conceptualized as separate dimensions — at least when it comes to large-scale assessment. The results indicate relatively clearly that treating naïve and informed views as single and separate dimensions it could be justified to. Spoken within the article's title metaphor, naïve and informed views rather correspond somewhat more to different coins in different currencies than to two sides of the same coin. These results of the SEM-based model testing, the correlation and factor loading analysis, the LCA as well as the interview data support this conclusion for our approach. In the following, we will discuss these results using three implications and derive directions for future research.

6.2 | Implication 1: Negatively and positively worded items may decrease construct validity in NOSI assessments and lead to a loss of information

Our results implicate that researchers may gain more relevant information on students' views by distinguishing between naïve and informed views within the conception of views and within the conception of their instrument. Studies that do not take these views into account independently of each other may lose relevant information. Using positively and negatively worded items for assigning students to one continuous NOSI scale in a particular dimension may mix up naïve and informed views. We will show this idea using three fictive students as example. One student with an average score of 2 on one single continuous scale could correspond to a student (Number 1) who rather agrees with informed views and strongly agrees with naïve views, to a student (Number 2) who strongly disagrees with informed views and rather disagrees with naïve views or to a student (Number 3) who rather disagrees with informed views and rather agrees with naïve views. Thus, a single scale, merging naïve and informed views, may not comprise different knowledge profiles without being able to assess it. Figure 3 presents this fictive example graphically.

Correspondingly, quantitative studies, having many students in the middle of a scale, might have students with different profiles between informed naïve views in their sample. They might not be able to identify these students due to putting naïve and informed views together on one single dimension. This would go for internationally published studies using Likert-scaled items like the studies on views and beliefs such as Conley, Pintrich, Vekiri, and Harrison (2004), Kampa, Neumann, Heitmann, and Kremer (2016), Harrison, Duncan Seraphin, Philippoff, Vallin, and Brandon (2015) and Neumann, Neumann, and Nehm (2011) who refer to the instrument Lombrozo, Thanukos, and Weisberg (2008). It might be fruitful to deepen the analysis in these studies or to develop these or further instruments with regard to a more distinguished analysis between negatively and positively worded items.

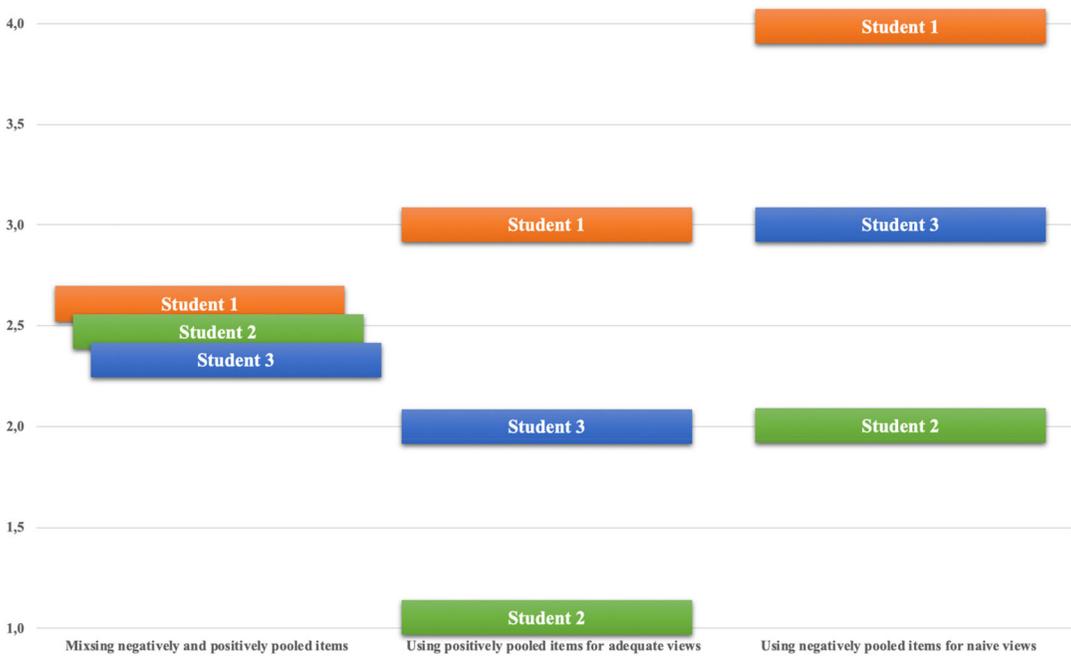


FIGURE 3 Three fictive students show how researchers might lose information and represent students views inadequately by mixing up items containing adequate and naïve views. All three students would get the same score on a common scale although they might hold different adequate and naïve views [Color figure can be viewed at wileyonlinelibrary.com]

It may be discussed or even questioned whether these studies represent the constructs of naïve and informed views adequately with regard to construct validity. Future developments of questionnaires may increase construct validity by using separate scales for naïve and informed views. This may be one important piece within the puzzle of increasing the validity of Likert-based instruments in the field of students' epistemology (Sandoval, Greene, & Bråten, 2016).

6.3 | Implication 2: Models for describing students cognition based on a “ladder metaphor” may be oversimplified

In a broader view, one may even ask how models for NOS and NOSI views that go back to a “ladder metaphor” can be justified with regard to this evidence. These models, which describe students' cognition in the field of epistemology, define ordered levels of more and more adequate understanding rather than a continuum. For example, Carey, Evans, Honda, Jay, and Unger (1989) describe Levels 1–3 that correspond to making to a distinction between ideas and experiments, distinguishing between ideas and experiments, and recognizing the cyclic, cumulative NOS. Mesci and Schwartz (2017, pp. 334) describe levels of a NOS continuum from naïve, over mixed, to increasing levels of understanding. The authors state that if “participants demonstrate inconsistent views about an aspect, they were placed in the (+) range of the continuum and considered to hold “mixed” views. Also, Lederman et al. (2014) describe, for the widely used VASI questionnaire, three levels for categorizing views. They distinguish between naïve, transitional, and informed views within an NOSI aspect. Regarding these levels, Mesci and Schwartz (2017, pp. 335) argue that the “use of a continuum

enables identification of the ‘in between’ to be represented as such. ‘In between’ are those perspectives that do not fully align with ‘naïve’ or ‘informed’. Likewise, the continuum enables relative representation of views within the ‘informed’ range (less informed/ more informed/even more informed etc.)”.

With regard to this study, it could be justified to treat the “steps of the ladder” as single dimensions leading to models that rely on a “more of the informed and less of the naïve” logic. Thus, naïve and informed views could be conceptualized independently from each other. Before questioning these models, however, we consider that more qualitative and quantitative research on the interrelatedness of naïve and advanced is necessary.

It might be that more advanced students, such as Hendrik from our example (see Supporting Information Appendix), are able to relate the contradicting and noncontradicting naïve and informed dimensions. They might realize that if they consider a view in one informed dimension as being informed, they cannot consider the views in other naïve dimensions as also being informed. Less advanced students, such as Niklas (see Supporting Information Appendix), might not be able to recognize these implicit contradictions. They may hold both naïve and informed views at the same time, as they are not able to connect and to compare both views. From a modeling views perspective, this means that higher dimensional models might be more fruitful in describing less advanced NOSI learners, whereas lower dimensional models might be sufficient to describe more advanced NOSI learners. Future research would have to take in more data on these assumptions with regard to discussing the appropriateness of models using a “ladder metaphor” for describing students' NOSI views.

With regard to learning NOSI views, it would be very promising to think about how to foster informed and how to reduce naïve views in an effective manner.

6.4 | Implication 3: Adequate and naïve NOSI views should (quantitatively) be conceptualized as interrelated but distinct constructs

In the following part, we will try to describe and to systematize the four profiles we chose in the LCA. Class 1—Advanced learners holding informed NOSI views: These students are able to distinguish clearly between naïve and informed NOSI views. They are very sure about informed views, and disagree strongly with naïve views, as well as scientists only trying something out during investigating. These students strongly disagree with views about discovering theories randomly, about the conformation bias, about confounded experiments and about models being a copy of reality. Based on the most likely latent class membership, this class comprises the highest number of students (48.9%). Class 2—Slightly unsure learners holding more informed than naïve NOSI views: These students are able to distinguish between naïve and informed NOSI views. However, they tend to slightly agree or disagree with these views; except for the view of scientists trying something out and seeing if it works, where they are very sure about their disagreement. The peak of confounded experiments almost reaching the scale mean shows that they are not very sure if scientists manipulate many variables at a time, although they are quite sure about the CVS in experiments. Based on the most likely latent class membership, this class comprises 31.2% of the students. Class 3—Unsure learners being rather sure about informed naïve NOSI and tending to agree with unplanned investigations and confounding variables: These students are able to distinguish between naïve and informed NOSI views and tend to slightly agree or to disagree with these views. Contrary to Class 2, this also goes for the view of scientists trying something out and seeing if it works. They may think that scientists investigate randomly in some but not all cases. These are quite sure about the CVS in

experiments. However, they only disagree very slightly with confounded experiments. Based on the most likely latent class membership, this class comprises 13.8% of the students. Class 4—Learners holding informed and naïve NOSI at the same time: These students are not able to distinguish between informed and naïve NOSI views. Although there might be a slight trend toward informed views, their agreement and disagreement range around the scale mean, indicating that, for them, naïve and informed views are equally likely to describe how scientists carry out investigations. Based on the most likely latent class membership, this class comprises 6.1% of the students.

6.5 | Would relying on contradicting naïve and informed views have led to different results?

Critical readers may state that this may be due to the fact that we did not define naïve and informed views as direct opposites, but in a more or less complementary manner. Our criterion was whether a naïve or informed view would contribute to a purposeful implementation of scientific investigations by the students or not. Looking at the correlation table (Table 6), we observe much higher negative correlations for views being in a more conceptual opposite, such as View 3 (“scientists change only one variable at a time for valid experiments”) and View 9 (“scientists change many variables at a time for valid experiments”) with a correlation of $r = -.57$ ($p < .001$) or View 1 (“scientists are guided by ideas and plan observations”), and View 6 (“while observing, scientists try something out and see if it works”) with a correlation of $r = -.59$ ($p < .001$). These correlations are much higher than those between complementary naïve and informed views. However, it has to be stated that the negative correlations between opposite views are very far from being $r = -1.0$. Even opposite naïve and informed views may have a complementary function within the questionnaire, as students may express a more or less sure view by only rather agreeing with the one and rather disagreeing with the other, which would correspond to the conception that “sometimes scientists do it like that or sometimes they do it like this.”

6.6 | Limits and further future directions

One basic limit, and the basic strength of this study, is in its quantitative approach. Although it cannot take into account the individual understanding of a single student, it can provide a clear confirmatory approach to deriving and testing models against data from a comparatively large sample. We have invested some effort in ensuring the validity of the questionnaire. One limit, which applies to all Likert-scaled questionnaires, is that the students do not have to produce an answer, but to evaluate statements that go back to different views. The evidence gained in this study applies, therefore, strongly to large-scale assessment (as mentioned in the title). Implications on further aspects of NOSI research have the status of data-based assumptions. We see our approach strongly in this particular area, but think that our results may shed an interesting light on further NOSI research.

In addition to the possibilities mentioned above, we can derive three main directions for future research. First, reanalysis of existing instruments could test the “two-sides-of-the-same-coin-models” that distinguish between naïve and informed items, or views against existing data. This could be the first step toward replicating the findings of our study. Second, mixed-method studies may try to validate the multiple informed and naïve NOSI profiles. We consider the interview data presented in this article as one small first step in this direction. Having more data on reasoning processing within these dimensions would add a new perspective the quantitative findings, to a large extent. It would be fruitful to learn more about how students justify their views, and which

scientific examples they would use for argumentation, for example. Third, experimental studies involving students in learning processes that focus on informed views or on informed and naïve views could examine which NOSI dimensions change, to which extent which students would reduce naïve views when learning about informed views, and which profiles react to which intervention. This would link intervention studies on NOSI learning with basic research, and possibly trigger a new branch of NOS and NOSI research.

ACKNOWLEDGMENTS

The authors cordially thank Jasmin Meyer, Jesco Schrader, Kristin Tscherning, Prof. Dr. Hans Bickes, Prof. Dr. Harald Gropengießer, Prof. Dr. Rüdiger Tiemann, Prof. Dr. Annette Upmeier zu Belzen, and all participating students and teachers for the support during this study.

ORCID

Andreas Nehring  <https://orcid.org/0000-0002-8723-5552>

REFERENCES

- Abd-El-Khalick, F. (2012). Examining the sources for our understandings about science: Enduring confluences and critical issues in research on nature of science in science education. *International Journal of Science Education*, 34(3), 353–374. <https://doi.org/10.1080/09500693.2011.629013>
- Asparouhov, T., & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the web with KIE. *International Journal of Science Education*, 22(8), 797–817. <https://doi.org/10.1080/095006900412284>
- Brunner, M., Kunter, M., Krauss, S., Baumert, J., Blum, W., Dubberke, T., ... Neubrand, M. (2006). Welche Zusammenhänge bestehen zwischen dem fachspezifischen Professionswissen von Mathematiklehrkräften und ihrer Ausbildung sowie beruflichen Fortbildung? *Zeitschrift für Erziehungswissenschaft*, 9(4), 521–544. <https://doi.org/10.1007/s11618-006-0166-1>
- Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed). *Ecological Modelling* (Vol. 172). <https://doi.org/10.1016/j.ecolmodel.2003.11.004>, 96, 97
- Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, 11(5), 514–529. <https://doi.org/10.1080/0950069890110504>
- Chen, H. J., She, J. L., Chou, C. C., Tsai, Y. M., & Chiu, M. H. (2013). Development and application of a scoring rubric for evaluating students' experimental skills in organic chemistry: An instructional guide for teaching assistants. *Journal of Chemical Education*, 90(10), 1296–1302. <https://doi.org/10.1021/ed101111g>
- Chen, S., Chang, W. H., Lieu, S. C., Kao, H. L., Huang, M. T., & Lin, S. F. (2013). Development of an empirically based questionnaire to investigate young students' ideas about nature of science. *Journal of Research in Science Teaching*, 50(4), 408–430. <https://doi.org/10.1002/tea.21079>
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098–1120. <https://doi.org/10.1111/1467-8624.00081>
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1–49. <https://doi.org/10.2307/1170558>
- Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330–351.

- Conley, A. M., Pintrich, P. R., Vekiri, I., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology, 29*(2), 186–204. <https://doi.org/10.1016/j.cedpsych.2004.01.004>
- Crawford, B. A., & Cullin, M. (2005). Dynamic assessments of preservice teachers' knowledge of models and modeling. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), *Research and the quality of science education* (pp. 309–323). Springer: Dordrecht, The Netherlands.
- Demir, A., & Abell, S. K. (2010). Views of inquiry: Mismatches between views of science education faculty and students of an alternative certification program. *Journal of Research in Science Teaching, 47*(6), 716–741. <https://doi.org/10.1002/tea.20365>
- Deng, F., Chen, D.-T., Tsai, C.-C., & Chai, C. S. (2011). Students' views of the nature of science: A critical review of research. *Science Education, 95*(6), 961–999. <https://doi.org/10.1002/sce.20460>
- DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2–3), 105–225. <https://doi.org/10.1080/07370008.1985.9649008>
- Dogan, N., & Abd-El-Khalick, F. (2008). Turkish grade 10 students' and science teachers' conceptions of nature of science: A national study. *Journal of Research in Science Teaching, 45*(10), 1083–1112. <https://doi.org/10.1002/tea.20243>
- Eberbach, C., & Crowley, K. (2009). From everyday to scientific observation: How children learn to observe the Biologist's world. *Review of Educational Research, 79*(1), 39–68. <https://doi.org/10.3102/0034654308325899>
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Greene, J. A., Cartiff, B. M., & Duke, R. F. (2018). A meta-analytic review of the relationship between epistemic cognition and academic achievement. *Journal of Educational Psychology, 110*, 1084–1111. <https://doi.org/10.1037/edu0000263>
- Gropengießer, H. (2009). Beobachten. In H. Gropengießer, U. Harms, & U. Kattmann (Eds.), *Fachdidaktik Biologie* (pp. 273–277). Hallbergmoos: Auslis-Verlag.
- Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching, 28*(9), 799–822. <https://doi.org/10.1002/tea.3660280907>
- Grünkorn, J., Upmeier zu Belzen, A., & Krüger, D. (2014). Assessing students' understandings of biological models and their use in science to evaluate a theoretical framework. *International Journal of Science Education, 36*(10), 1651–1684. <https://doi.org/10.1080/09500693.2013.873155>
- Harrison, G. M., Duncan Seraphin, K., Philippoff, J., Vallin, L. M., & Brandon, P. R. (2015). Comparing models of nature of science dimensionality based on the next generation science standards. *International Journal of Science Education, 37*(8), 1321–1342. <https://doi.org/10.1080/09500693.2015.1035357>
- Hellwig, B. (2018). *ELAN*. Nijmegen. Retrieved from <https://tla.mpi.nl/tools/tla-tools/elan/>
- Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences, 66*, 4–15. <https://doi.org/https://doi.org/10.1016/j.lindif.2017.11.001>
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education, 88*(1), 28–54. <https://doi.org/10.1002/sce.10106>
- Huang, C.-M., Tsai, C.-C., & Chang, C.-Y. (2005). An investigation of Taiwanese early adolescents' views about the nature of science. *Adolescence, 40*(159), 645–654.
- Ibrahim, B., Buffler, A., & Lubben, F. (2009). Profiles of freshman physics students' views on the nature of science. *Journal of Research in Science Teaching, 46*(3), 248–264. <https://doi.org/10.1002/tea.20219>
- Kampa, N., Neumann, I., Heitmann, P., & Kremer, K. (2016). Epistemological beliefs in science—a person-centered approach to investigate high school students' profiles. *Contemporary Educational Psychology, 46*, 81–93. <https://doi.org/10.1016/j.cedpsych.2016.04.007>
- Khishfe, R., & Abd-El-Khalick, F. (2002). Influence of explicit and reflective versus implicit inquiry-oriented instruction on sixth graders' views of nature of science. *Journal of Research in Science Teaching, 39*(7), 551–578. <https://doi.org/10.1002/tea.10036>
- Krell, M., Upmeier zu Belzen, A., & Krüger, D. (2014). Students' levels of understanding models and modelling in biology: Global or aspect-dependent? *Research in Science Education, 44*(1), 109–132. <https://doi.org/10.1007/s11165-013-9365-y>

- Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*(11), 866–870. <https://doi.org/10.1111/j.1467-9280.2005.01628.x>
- Künsting, J., Wirth, J., & Paas, F. (2011). The goal specificity effect on strategy use and instructional efficiency during computer-based scientific discovery learning. *Computers & Education*, *56*(3), 668–679. <https://doi.org/10.1016/j.compedu.2010.10.009>
- Lederman, J., Lederman, N., Bartels, S., Jimenez, J., Akubo, M., Aly, S., ... Zhou, Q. (2019). An international collaborative investigation of beginning seventh grade students' understandings of scientific inquiry: Establishing a baseline. *Journal of Research in Science Teaching*, *56*(4), 486–515. <https://doi.org/10.1002/tea.21512>
- Lederman, J., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Teaching*, *51*(1), 65–83. <https://doi.org/10.1002/tea.21125>
- Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 831–879). Mahwah, NJ: Erlbaum.
- Little, T. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.
- Lombrozo, T., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for accepting evolution. *Evolution: Education and Outreach*, *1*(3), 290–298. <https://doi.org/10.1007/s12052-008-0061-8>
- Mesci, G., & Schwartz, R. S. (2017). Changing preservice science teachers' views of nature of science: Why some conceptions may be more easily altered than others. *Research in Science Education*, *47*(2), 329–351. <https://doi.org/10.1007/s11165-015-9503-9>
- Nehring, A., Nowak, K. H. K. H., Upmeyer zu Belzen, A., Tiemann, R., zu Belzen, A. U., & Tiemann, R. (2015). Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables. *International Journal of Science Education*, *37*(9), 1343–1363. <https://doi.org/10.1080/09500693.2015.1035358>
- Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. <https://doi.org/10.1080/09500693.2010.511297>
- Nowak, K. H. K. H., Nehring, A., Tiemann, R., & Upmeyer zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education*, *47*(3), 182–188. <https://doi.org/10.1080/00219266.2013.822747>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. <https://doi.org/10.1080/10705510701575396>
- O'Connor, B. (2000). SPSS and BAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, *32*(2), 396–402.
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What “ideas-about-science” should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, *40*(7), 692–720. <https://doi.org/10.1002/tea.10105>
- Poletiek, F. H. (1996). Paradoxes of falsification. *The Quarterly Journal of Experimental Psychology: Section A*, *49*(2), 447–462. <https://doi.org/10.1080/027249896392720>
- Pomeroy, D. (1993). Implications of teachers' beliefs about the nature of science: Comparison of the beliefs of scientists, secondary science teachers, and elementary teachers. *Science Education*, *77*(3), 261–278. <https://doi.org/10.1002/sce.3730770302>
- Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education*, *13*(1), 21–43. <https://doi.org/10.1007/s10763-014-9520-6>
- Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground—A literature review of empirical research on scientific inquiry activities. *Studies in Science Education*, *52*(2), 161–197. <https://doi.org/10.1080/03057267.2016.1206351>
- Rosseel, Y. (2012). Lavaan : An R package for. *Structural Equation Modeling*, *48*(2), 1–20.

- Sandoval, W. A., Greene, J. A., & Bråten, I. (2016). Understanding and promoting thinking about knowledge: Origins, issues, and future directions of research on epistemic cognition. *Review of Research in Education*, 40(1), 457–496. <https://doi.org/10.3102/0091732X16669319>
- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28(9), 859–882. <https://doi.org/10.1002/tea.3660280910>
- Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology*, 49(9), 1639–1649. <https://doi.org/10.1037/a0030976>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: An explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, 88(4), 610–645. <https://doi.org/10.1002/sc.10128>
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654. <https://doi.org/10.1002/tea.20311>
- Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review*, 39, 37–63. <https://doi.org/10.1016/j.dr.2015.12.001>
- Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124(2), 209–215. <https://doi.org/10.1016/j.cognition.2012.04.005>
- Straatemeier, M., van der Maas, H. L. J., & Jansen, B. R. J. (2008). Children's knowledge of the earth: A new methodological and statistical approach. *Journal of Experimental Child Psychology*, 100(4), 276–296. <https://doi.org/10.1016/j.jecp.2008.03.004>
- Thelk, A. D., & Hoole, E. R. (2006). What are you thinking? Postsecondary student think-alouds of scientific and quantitative reasoning items. *The Journal of General Education*, 55(1), 17–39.
- Thorn, C. J., Bissinger, K., Thorn, S., & Bogner, F. X. (2016). “Trees Live on Soil and Sunshine!”—Coexistence of scientific and alternative conception of tree assimilation. *PLoS One*, 11(1), e0147802. <https://doi.org/10.1371/journal.pone.0147802>
- Tsai, C. C. (1999a). “Laboratory exercises help me memorize the scientific truths”: A study of eighth graders' scientific epistemological views and learning in laboratory activities. *Science Education*, 83(6), 654–674. [https://doi.org/10.1002/\(SICI\)1098-237X\(199911\)83:6<654::AID-SCE2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1098-237X(199911)83:6<654::AID-SCE2>3.0.CO;2-Y)
- Tsai, C. C. (1999b). The progression toward constructivist epistemological views of science: A case study of the STS instruction of Taiwanese high school female students. *International Journal of Science Education*, 21(11), 1201–1222. <https://doi.org/10.1080/095006999290156>
- Tsai, C. C. (2000). Relationships between student scientific epistemological beliefs and perceptions of constructivist learning environments. *Educational Research*, 42(2), 193–205. <https://doi.org/10.1080/001318800363836>
- Upmeyer zu Belzen, A., & Krüger, D. (2010). Modellkompetenz im Biologieunterricht. Struktur und Entwicklung. *Zeitschrift Für Didaktik Der Naturwissenschaften*, 16, 41–57.
- Urban, D., & Mayerl, J. (2014). *Strukturgleichungsmodellierung. Ein Ratgeber für die Praxis*. Wiesbaden: Springer VS.
- Vorholzer, A., von Aufschnaiter, C., & Boone, W. J. (2018). Fostering upper secondary students' ability to engage in practices of scientific investigation: A comparative analysis of an explicit and an implicit instructional approach. *Research in Science Education*, online first, 1–22. <https://doi.org/10.1007/s11165-018-9691-1>
- Wenning, C. J. (2006). Assessing nature-of-science literacy as one component of scientific literacy. *Journal of Physics Teacher Education Online*, 3(4), 3–14.
- Wilson, L. L. (1954). A study of opinions related to the nature of science and its purpose in society. *Science Education*, 38(2), 159–164. <https://doi.org/10.1002/sc.3730380209>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Nehring A. Naïve and informed views on the nature of scientific inquiry in large-scale assessments: Two sides of the same coin or different currencies? *J Res Sci Teach.* 2020;57:510–535. <https://doi.org/10.1002/tea.21598>