Chapter 5
Toward a Questionnaire to Assess Biology Student Teachers' Knowledge of the Nature of Scientific Inquiry (NOSI)

Corinne Charlotte Wacker, Marius Barth, Christoph Stahl, and Kirsten Schlüter

5.1 Introduction

The nature of scientific inquiry (NOSI) is one critical component of science literacy and is also becoming increasingly important in view of the current corona pandemic. Teaching and learning the characteristics of scientific inquiry processes, through which scientific knowledge is generated and justified, is not only emphasized in the German educational standards set by the standing conference of the ministers of education and cultural affairs (KMK, 2005), but also worldwide as part of science standards (NGSS Lead States, 2013). Thus, students as well as teachers must be able to understand, conduct and critically assess scientific investigations. However, despite the continued science education efforts, research indicates that teachers and student teachers, as well as students of varying ages, typically hold naïve NOSI views (Lederman et al., 2019; Mesci et al., 2020; Zion et al., 2018). Because teachers need to have an elaborated understanding of NOSI in order to be able to discuss it adequately in their lessons, it is important to assess whether (and how well) this educational goal is actually achieved throughout their university education to provide the necessary resources for them.

To appraise the NOSI proficiency of a person, one must first define what it means to be competent in it. This is particularly important because NOSI and nature of science (NOS) are often used as synonymous terms and are frequently combined and

C. C. Wacker (✉) · K. Schlüter
Institute of Biology Education, University of Cologne, Cologne, Germany
e-mail: cwacker1@uni-koeln.de; kirsten.schlueter@uni-koeln.de

M. Barth · C. Stahl
Research Methods and Experimental Psychology, Department of Psychology, University of Cologne, Cologne, Germany
e-mail: marius.barth@uni-koeln.de; christoph.stahl@uni-koeln.de

overlap, because they are interdependent (Mayer, 2007). Nevertheless, Schwartz et al. (2008) distinguish between them by stating that NOSI is about "how" the scientific knowledge is generated and validated, i.e. the nature of the *practices* that are most closely related to the *processes* of inquiry; whereas NOS embodies what distinguishes science from other disciplines. Thus, NOS refers to the characteristics of the scientific knowledge, i.e. the *product* of inquiry processes (Schwartz et al., 2012). Schwartz et al. (2008) and Lederman et al. (2014) identified aspects of NOSI *via* literature reviews and science practices studies. According to Schwartz et al. (2008) these NOSI sub-competences include: (*sc*1) scientific investigations all begin with a question and do not necessarily test a hypothesis; (*sc*2) there is no single set of steps followed in all investigations; (*sc*3) scientific questions scientists choose to pursue stem from many sources and can serve many purposes; (*sc*4) scientific data can be interpreted differently; (*sc*5) scientists recognize anomalous data and handle them in a reflective manner; (*sc*6) scientific data are not the same as scientific evidence; and (*sc*7) scientific inquiry is embedded within a researcher's community.

Various NOSI instruments were developed especially in the last 30 years for various stages of education using a variety of response formats (Temiz et al., 2006). However, they primarily focus on pupils up to 10th grade and little is known about natural science student teachers' NOSI views during their university education (Mesci et al., 2020). Furthermore, many of these testing instruments make use of: either a multiple-choice format, which is considered to be time- and administration-economic, but is susceptible to test-wiseness; or an open-ended format with follow-up interviews, which is considered to be time-consuming and vulnerable to discrepancies in interpretation (Temiz et al., 2006; Thoma & Köller, 2018). Another factor that needs to be considered is that the focus of most instruments lies upon experimentation and by this on causal relationships, due to the fact that the experiment is considered to be "the gold standard" of science. Other research methods, such as observations and comparisons, are often seen as preliminary stages or partial aspects of experimentation (Ayyavoo et al., 2002; Wellnitz & Mayer, 2013). It was therefore decided to develop a curriculum-independent NOSI questionnaire that can be used at any point in the academic education of biology student teachers.

The purpose of our study is: (1) to develop a closed-ended questionnaire to assess biology student teachers' NOSI views; and (2) then to validate this instrument's functioning in order to discuss its potential for research and teaching. There is a need to gain insight into the NOSI competence of future science teachers throughout their university education in order to further improve it.

5.2   Method

5.2.1   *Participants*

The NOSI questionnaire was administered to undergraduate biology student teachers in the introductory course "Basics of biology" at the University of Cologne. This was done during the winter semesters of 2018/19 and 2019/20. The sample of the

148 freshman student teachers of biology was comprised of 108 women (73%) and 40 men (27%), and the average age was 20.7 years (SD = 2.9). Data were collected in an online survey using LimeSurvey 3.21; however, during the first survey phase a paper-pencil version of the instrument was also employed due to server issues. Before participants answered the questionnaire, they were briefed, i.e. a brief introduction concerning the voluntary and anonymous participation in this pilot study was given. The response rate amounted to 77.5%.

5.2.2   Questionnaire Design

A closed-ended NOSI instrument was designed by a group of experts from the field of biology education and psychology. The developed questionnaire was constructed in reference to the seven previously mentioned NOSI aspects of Schwartz et al. (2008) and Lederman et al. (2014) (see Table 5.1). It is important to note that these authors explicitly state that these seven sub-competences are not the only ones, but nevertheless they are indispensable for students (Lederman et al., 2014; Schwartz et al., 2008; Zion et al., 2018). It was decided to extend the testing instrument to include the additional aspect of Questionable Research Practices (QRPs), i.e. either fabrication or falsification of scientific data or results (Fiedler & Schwarz, 2016). Krishna and Peter (2018) stated that approximately 10% of psychology students, who were writing a Bachelor's or Master's thesis on a psychology course at a German public University, practiced some QRPs and that lecturers have an important function in shaping the students' attitude towards them. Consequently, it was

Table 5.1 Overview and distribution of items in the NOSI testing instrument. Adapted from Schwartz et al. (2008) and Lederman et al. (2014)

Sub-competence (*sc*)	Items context	Σ
*sc*1. Scientific questions guide investigations	Investigations do not necessarily need a hypothesis, but a research question is mandatory.	5
*sc*2. Multiple methods of scientific investigations	Diverse research methods exist (e.g. experiment, correlation study).	11
*sc*3. Multiple purposes of scientific investigations	(Theoretical versus practical) goals of scientific research.	9
*sc*4. Justification of scientific knowledge	Data do not stand alone, but can be interpreted differently.	4
*sc*5. Recognition and handling of anomalous data	Dealing with anomalous information (e.g. looking for information why these data occur).	2
*sc*6. Distinctions between data and evidence	Evidence is the result of data analysis and interpretation.	8
*sc*7. Community of practice (CoP)	Scientific inquiry processes and data are checked by a researcher's community.	2
*sc*8. Questionable Research Practices (QRPs)	Transparency in data collection and reporting.	5

felt that this area, which also touches on the fifth aspect of recognizing and handling of abnormal data, is also essential for the processes of scientific inquiry, i.e. NOSI. With regards to the focus and utility of this NOSI testing instrument, it is important to acknowledge that some areas of NOSI are difficult to distinguish from each other and these sub-competences also tend to overlap with some areas of NOS. In the preliminary phase of the instrument development, content validity was therefore assessed by repeated discussion sessions by the authors, who examined the wording of the statements as well as whether each item fits in its allocated NOSI aspect.

Adjustments were made to the preliminary item pool of 85 items based on a first test survey with university students, so that items that were too easy to answer or too similar to other items, or that could not be clearly assigned to an aspect, were either modified accordingly or eliminated. Finally, the developed NOSI questionnaire consisted of 46 closed items, whereby respondents had to agree/disagree first with statements (true or false), and then subsequently rate their answer in a confidence rating (How confident are you that your answer is correct (as a percentage)? Answers: 0 = guessing, 20, 40, 60, 80, and 100 = absolutely certain). The development of this testing instrument with two response formats could be an adequate trade-off between economic reasons, i.e. in a time-effective manner, and detailed participants' NOSI views that can also take into account the participants' test intelligence. Moreover, in accordance with the second sub-competence that there is no single scientific method that all (biology) scientists follow, the testing instrument neither concentrated on a specific research method nor on a specific curriculum. The complete 46-item NOSI questionnaire is available at https://osf.io/u9gdz/.

5.2.3   Data Analysis

In order to create a combined multi-response index for each item, a multiplicative weight for each item based on both response formats (dichotomous: true/false and the post-decision confidence rating: 0%, 20%, 40%, 60%, 80%, and 100%) was calculated. The dichotomous responses were coded either −1 (incorrect) or +1 (correct) and the confidence rating responses were expressed as relative probability, i.e. coded as 0, .2, .4, .6, .8 or 1, respectively. The next step was to multiply both values to calculate a multiplicative weight for each item and each case, i.e. $x_{dichotomous} \cdot x_{confidence\ rating}$ (see Table 5.2). For example, if a test person answered one item incorrectly and was 40% sure about his/her answer, the result would be $-1 \times .4 = -.4$.

Subsequently, item analysis was assessed for selecting items for the NOSI questionnaire. This is the average score of the combined multi-response index instead of item difficulty that uses scaling from −1 to +1, Measure of Sampling Adequacy (*MSA*), and Cronbach's α. In addition, further descriptive data analyses were made to better illustrate student teachers' overall NOSI understanding. This was done by calculating a NOSI total score for every biology student teacher *via* the arithmetic mean of the combined multi-response index of all items. In order to identify

Table 5.2 Combined multi-response index of both formats

| Dichotomous response was…:	… incorrect.					... guessed.	… correct.				
Confidence rating response [%]:	100	80	60	40	20	0	20	40	60	80	100
Multi-response index:	−1	−.8	−.6	−.4	−.2	0	.2	.4	.6	.8	1

potential scientific inquiry misconceptions as well as items that could be answered with test intelligence, we examined each NOSI item in detail according to the given responses by the respondents. A Maximum Likelihood (ML) exploratory factor analysis (EFA) followed by oblique rotation was also conducted in order to establish the underlying structure of factors.

5.3 Results

All analyses were performed using R 4.0.3 (R Core Team, 2020) with the package *psych* (Revelle, 2020).

5.3.1 Item Analysis

Figure 5.1a shows the distribution of the average score of the combined multi-response index for each item across all participants. It indicates that most items have an average score in the positive scale range, this means that most items were chiefly correctly answered. Nevertheless, all five items of the NOSI sub-competence "Scientific questions guide investigations" and two items (8.1 & 8.2) of the NOSI sub-competence "QRPs" were most difficult for the student teachers to answer due to their negative average scores ranging from –.046 to –.367. However, there are no items that were overall too difficult or too easy according to the histogram border-line areas of $-.8 < score_i < .8$. Measure of Sampling Adequacy (*MSA*) was employed to determine the extent to which an item was suitable for factor analysis, i.e. its discriminatory power. If the items are not at all or only weakly correlated with all other items, it is unlikely that factors can be found by which the multiplicity of the variables can be reduced on a smaller number of dimensions (Ludwig-Mayerhofer, 2017). The authors decided to use a rather low cut-off score of .4 to avoid prematurely eliminating items of the NOSI questionnaire. After excluding five items (2.5, 3.3, 3.4, 6.5, 8.3) from the instrument, which are the five points below the *MSA* limit at around .39 in Fig. 5.1b, with the two points of items 3.3 and 3.4 overlapping each other, 41 items remained in the reduced NOSI item pool within the defined boundaries (see Fig. 5.1c). For the reduced item pool, the overall *MSA* = .65 can be considered useful and the reliability can be considered acceptable with Cronbach's α = .69, Guttman's λ_6 = .84. It was also decided to report Guttman's λ_6, because Cronbach's α tends to underestimate reliability in tests with strong heterogeneity, such as comprising eight components/aspects (Osburn, 2000).
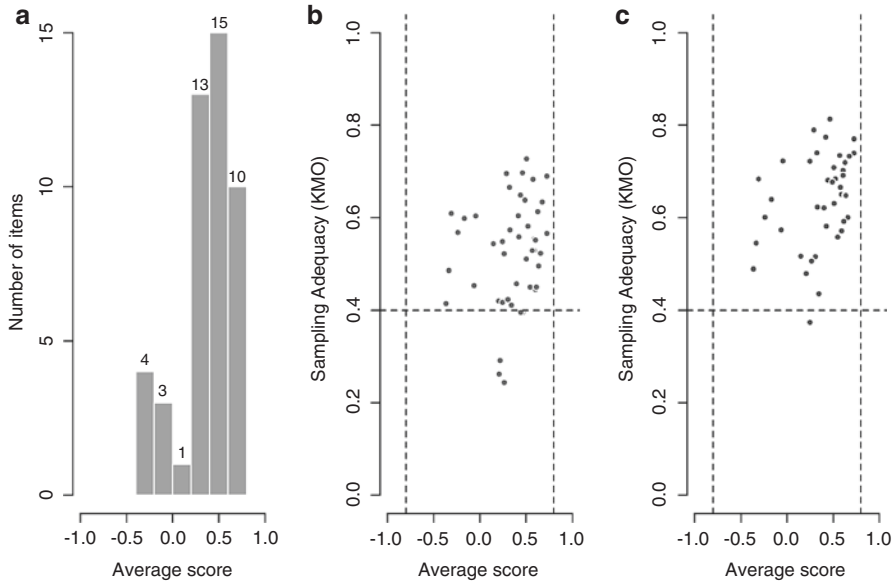
Fig. 5.1 Average score of the combined multi-response index and *MSA* for each NOSI item of the complete list of 46 items (**a**, **b**) and the reduced list of 41 items (**c**). The limits are depicted as dashed lines in (**b**) and (**c**). The numbers above the bars depict the total number of items in (**a**). Because *MSA* calculates the correlations of an item to all other items, the points in the scatterplots change from Figure (**b**) to (**c**) after eliminating five items

5.3.2   Biology Student Teachers' NOSI Competences

The arithmetic mean of the responses to all NOSI items of the test, i.e. the participant ability score, was calculated for each of the 148 freshman biology student teachers (see Fig. 5.2). The positive range of NOSI competence for all participants is between .02 and .65 with an average of $M = .36 \pm .13$.

Despite the overall positive range of NOSI understanding by the freshman biology student teachers, two interesting response patterns were identified within the combined multi-response index by looking at the items in detail. A few items received incorrect responses that were given with high confidence ranging from 80% to 100% (= absolutely certain), and some items were answered correctly, but with a low confidence rating ranging from 0% to 20% (= guessing) by the respondents. The authors focused especially on the items of the testing instrument where more than 10% of the biology student teachers showed these response patterns (see Fig. 5.3). For eight items 14–30% of the participants were certain, i.e. 80–100% confidence, that their answers were correct, although this was not the case. The top item within this group, which is called *NOSI misconceptions,* was: 'A scientific investigation always checks a hypothesis' (item 1.3). Conversely,
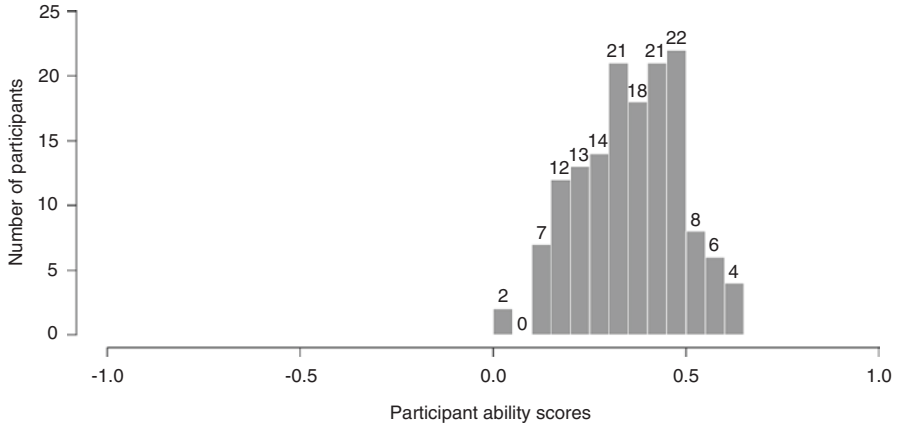
Fig. 5.2 Biology student teacher's overall NOSI score. The numbers above the bars depict the total number of participants
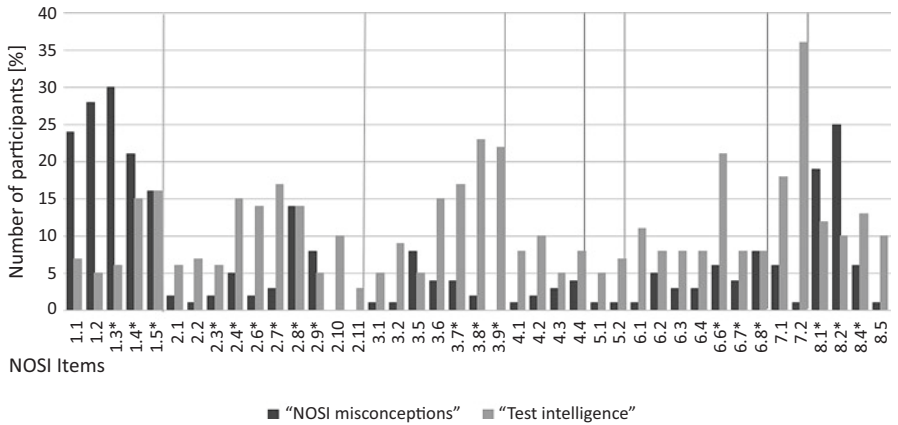
Fig. 5.3 Proportion of all "high-confidence + incorrect" (80–100% confidence and incorrectly answered) and "low-confidence + correct" (0–20% confidence and correctly answered) responses. Negatively formulated questions are marked with an asterisk. The vertical lines separate the eight NOSI-sub-competences from each other

11–36% of the biology student teachers answered 17 questions correctly, although they felt rather uncertain, i.e. 0–20% confidence, that their responses were right. The top item within this group, which is labelled *test intelligence*, was: 'Scientists organize themselves in professional societies to set standards for scientific work' (item 7.2). In addition, four items showed even both these answer patterns by more than 10% of the participants (1.4, 1.5, 2.8, and 8.1).

5.3.3 Factor Extraction Results

We conducted a parallel analysis of the remaining 41 items, which suggested keeping five factors for an exploratory factor analysis. Only 17 items with loadings greater than ±.40 were used to characterize the 5-factor construct in Table 5.3. The other 24 items aren't shown here because of their very low communalities. Furthermore, only one or two items loaded on the factors 1, 3, and 4, whereas eight items loaded on factor 2, with loadings ranging from .59 to .40. Moreover, there were four items that loaded on factor 5 with loadings ranging from .59 to .44.

Table 5.3 Correlation matrix (EFA with oblique rotation). Negatively formulated questions are marked with an asterisk. The communalities are depicted in the last column as h^2 indicating each item's variance that can be explained by the corresponding model

Item	Factor					h^2
	1	2	3	4	5	
5.2 Unexpected results can make it necessary to change a scientific theory.	**.99**					1.00
2.2 Depending on the research question, different methods are applied.		**.43**				.25
2.10 Different methods can lead to contradictory results.		**.56**				.37
2.11 When interpreting scientific data, the methods used should always be considered.		**.51**				.30
4.1 Two scientists interpreting the same data can justifiably come to different conclusions.		**.59**				.33
4.2 Several investigations using the same procedures may have different results.		**.40**		.21		.27
4.4 Scientific research data can be interpreted differently.		**.58**				.39
5.1 Unexpected results may cause the data to be reinterpreted.		**.58**				.39
6.4 Scientific evidence is based on the newly collected data of an investigation, but also on the previous state of research.		**.49**				.26
2.4* For a given research question there is also a given method to answer it.			.46	.24	.26	.41
6.6* The evidence of a scientific study is not directly related to the research question.			**.73**			.55
1.3* A scientific investigation always checks a hypothesis.				**.60**		.38
1.4* Before a research question is formulated, hypotheses about possible results of an investigation must have been derived.			.23	**.61**		.46
3.1 Scientific investigations can serve to develop technologies.					**.44**	.29
3.7* Mechanisms that cannot be directly observed cannot be scientifically investigated.				.24	**.49**	.43
6.3 Data are the observations made during a scientific investigation.				−.22	**.59**	.40
7.2 Scientists organize themselves in professional societies to set standards for scientific work					**.54**	.35

5.4 Discussion

The present study found that the NOSI questionnaire with a combined multi-response index has an acceptable instrument reliability before and after the elimination of five of its items. The average scores of the combined multi-response index (each referring to a single item) in Fig. 5.1a indicate a wide answering range between –.37 and .73, and the overall mean for all 41 items of $M = .36 \pm .13$ indicates an already existing moderate NOSI understanding of all biology student teachers. Moreover, there is no single item with a negative average combined multi-response index score below –.37, which might indicate that all not participants have internalized the same naïve NOSI views in their diverse school biology education. It is important to note that the arithmetic mean of the participant ability scores of all 41 items is based on an unequal number of items in each of the eight sub-competences, which was a result of the item selection process (originally starting from 85 items).

5.4.1 NOSI Misconceptions and Test Intelligence

Two thought-provoking response patterns could be identified by looking at the combined multi-response index of the NOSI questionnaire in more detail (see Fig. 5.3). One group of respondents featured a "false certainty" because they answered some items incorrectly, but were nevertheless sure that their answers were correct. These items could point towards NOSI misunderstandings, which may have been acquired through schooling. In particular, the top negatively coded item in this group, 'A scientific investigation always checks a hypothesis' (item 1.3), hints on the supposed 'general procedure' of 'the Scientific Method'. Almost any inquiry assignment in school science curricula seems to start with generating a hypothesis for an experiment and it seems that a study is only deemed a success if the results serve to confirm this hypothesis (Bencze, 1996). This circumstance could therefore easily lead to a widespread NOSI misconception. In addition, the other four items of the same sub-competence 'Scientific questions guide investigations' (sc1), were also answered by more than 10% of student teachers in a similar pattern. Thus, this sub-competence seems to include popular misconceptions about the role of hypothesis *versus* research question. Another sub-competence with two "false certainty" response items refers to 'Questionable Research Practices (QRPs)' (sc8). The items are 'On the basis of the data collected, the hypotheses of the study should be adapted.' (8.1) and 'If an expected effect is not yet statistically significant, data collection should be continued so that the effect can become significant.' (8.2). They both indicate that student teachers' NOSI understanding in the areas of HARKing ("Hypothesizing After the Results are Known") and "optional stopping" need to be improved.

There is also a larger group of 17 NOSI items within six sub-competences of the questionnaire, where more than 10% of participants felt rather uncertain that their correct response was right. For instance, the most often "truly guessed" item was: 'Scientists organize themselves in professional societies to set standards for

scientific work' (item 7.2). One could argue that 36% of the participants could have deduced from own as well as from heard experiences of others that the existence of CoPs seems highly likely, maybe because meeting with fellow students and (work) colleagues for exchanging experiences is part of everyday life.

Four of the 41 items in the NOSI questionnaire even show both response patterns ("false certainty" and "truly guessed"). In some cases, this could be a hint that there are difficulties in interpreting an item, e.g. item 2.7: 'Chance should not play a role in research'. Both ratings (true as well as false) regarding the correctness of the item have their legitimacy, depending on a student teachers' way of thinking. On the one hand the item is wrong, because chance sometimes plays an important role in science (e.g. discovery of the antibiotic Penicillin by Alexander Fleming in 1928) (Copeland, 2019). On the other hand, the item is correct, because when experimenting, all variables except for the one under investigation must be kept constant. Chance should be excluded in this case as far as possible, because otherwise one would not get any reliable and interpretable results. Therefore, this item cannot be interpreted on its own and must either be reformulated or interpreted in the context of other items. In general, there is a need to continue to improve student teachers' NOSI views so that they have a better understanding about scientific inquiry processes, and by this can fulfill their important role as future teachers in shaping students' NOSI views.

5.4.2 Exploratory Factor Analysis (EFA)

The correlation matrix of the EFA of the NOSI questionnaire with 41 items shows no obviously recognizable matrix structure. Although a 5-factor construct can be identified in Table 5.3, the factors 1, 3, and 4 each contain less than three item loadings. Moreover, items of the NOSI aspects 'Justification of scientific knowledge' (sc4) and 'Recognition and handling of anomalous data' (sc5) are not included in the pattern matrix. A possible explanation for the inconclusive EFA matrix structure – besides the two reliability limitations concerning the number of items and sample size, which are discussed in detail in the next chapter – could be that the respondents' NOSI abilities differ from the authors' theoretical construct. This could be indicated by the fact that the factors 2 and 5 include items from more than three different NOSI sub-competences. Thus, at this point the factor structure of the questionnaire is unclear and future research is required to better understand the underlying constructs and their relations.

5.4.3 Limitations

Despite the fact that the NOSI test instrument provides a first insight into the NOSI understanding of biology student teachers and their potential scientific inquiry misconceptions at the start of their academic education, there are obvious limitations to

this pilot study. Even though a sequential cross-section research design over more than 1 year was used and the response rate was high (77.5%), the sample of $N = 148$ for a 46-item questionnaire was more than half too small compared to the required size of $N = 400$ (Eid et al., 2017). Therefore, the validation study should be continued or even extended by the participation of, for example, student teachers of other science subjects in order to reach the required sample size. Secondly, the two NOSI sub-competences 'Recognition and handling of anomalous data' ($sc5$) and 'Community of practice' ($sc6$) had only two items instead of the recommended minimum of three items per aspect/factor, which is essential when conducting an EFA of a multidimensional construct such as NOSI (Raubenheimer, 2004). This is due to the fact that in the first NOSI questionnaire test survey, the other items of these sub-competences were eliminated because of their easy item difficulties. Because of these two limitations, the results of the EFA should only be interpreted carefully and a final selection of items for the NOSI questionnaire is therefore premature at this stage. Nevertheless, the identification of the two significant item response patterns within the combined multi-response index, i.e. potential NOSI misconceptions and test intelligence, allows the authors to have an additional decision criterion for the final item selection. One may consider eliminating NOSI items that indicate test intelligence, while retaining items that may point to NOSI misunderstandings. In the target group of freshman student teachers of biology, eight potential scientific misunderstandings could be found (see Fig. 5.3). However, their origin cannot be determined, although they are presumably due to biology or other scientific school lessons. To this end, a mixed-methods study with interviews could be conducted to learn more about the reasons and sources of misconceptions (as well as about the causes for possible test intelligence phenomena). Further studies are currently planned to explore more deeply the informative value of the newly developed NOSI test instrument by applying it to biology (and maybe even other natural sciences) student teachers in other years and phases of their academic education, such as in the Master's programme. This will on the one hand further validate the instrument's functioning and on the other hand help to identify and correct latent NOSI misconceptions that may have been created or propagated throughout former school and university education.

References

Ayyavoo, G., Bencze, J., Corry, A., & van Oostveen, R. (2002). Correlational studies in school science: Beyond experimentation. *OISE Papers in STSE Education, 3*, 221–230.

Bencze, J. L. (1996). Correlational studies in school science: Breaking the science-experiment-certainty connection. *School Science Review, 78*(282), 95–101.

Copeland, S. (2019). On serendipity in science: Discovery at the intersection of chance and wisdom. *Synthese, 196*(6), 2385–2406.

Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden*. Beltz Verlagsgruppe.

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science, 7*(1), 45–52.

KMK. (2005). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Luchterhand München.

Krishna, A., & Peter, S. M. (2018). Questionable research practices in student final theses – Prevalence, attitudes, and the role of the supervisor's perceived attitudes. *PLoS One, 13*(8), e0203470.

Lederman, J. S., 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.

Lederman, J., Lederman, N., Bartels, S., Jimenez, J., Akubo, M., Aly, S., et al. (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.

Ludwig-Mayerhofer, W. (2017). *ILMES – Internet-Lexikon der Methoden der empirischen Sozialforschung: Faktorenanalyse*. Retrieved 5th October, 2020, from http://wlm.userweb.mwn.de/Ilmes/ilm_f3.htm

Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In *Theorien in der biologiedidaktischen Forschung* (pp. 177–186). Springer.

Mesci, G., Çavuş-Güngören, S., & Yesildag-Hasancebi, F. (2020). Investigating the development of pre-service science teachers' NOSI views and related teaching practices. *International Journal of Science Education, 42*(1), 50–69.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Retrieved 1 October 2020, from https://www.nextgenscience.org/

Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological methods, 5*(3), 343.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Retrieved 30 October 2020, from http://www.r-project.org/index.html

Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology, 30*(4), 59–64.

Revelle, W. R. (2020). *psych: Procedures for personality and psychological research*.

Schwartz, R. S., Lederman, N. G., & Lederman, J. S. (2008). *An instrument to assess views of scientific inquiry: The VOSI questionnaire*. Paper presented at the National Association for Research in Science Teaching (NARST), Baltimore, MD.

Schwartz, R. S., Lederman, N. G., & Abd-el-Khalick, F. (2012). A series of misrepresentations: A response to Allchin's whole approach to assessing nature of science understandings. *Science Education, 96*(4), 685–692.

Temiz, B. K., Taşar, M. F., & Tan, M. (2006). Development and validation of a multiple format test of science process skills. *International Education Journal, 7*(7), 1007–1027.

Thoma, G.-B., & Köller, O. (2018). Test-wiseness: ein unterschätztes Konstrukt? *Zeitschrift für Bildungsforschung, 8*(1), 63–80.

Wellnitz, N., & Mayer, J. (2013). Erkenntnismethoden in der Biologie – Entwicklung und Evaluation eines Kompetenzmodells. *Zeitschrift für Didaktik der Naturwissenschaften, 19*, 315–345.

Zion, M., Schwartz, R. S., Rimerman-Shmueli, E., & Adler, I. (2018). Supporting teachers' understanding of nature of science and inquiry through personal experience and perception of inquiry as a dynamic process. *Research in Science Education, 50*(4), 1281–1304.