Development of an Instrument to Assess Views on Nature of Science and Attitudes Toward Teaching Science

SUFEN CHEN
Graduate School of Technological and Vocational Education and Education Center, National Taiwan University of Science and Technology, Taipei 106, Taiwan, R.O.C

Received 28 July 2005; revised 20 December 2005; accepted 24 January 2006

DOI 10.1002/sce.20147
Published online 21 April 2006 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: This article describes the development and field test of an instrument, the Views on Science and Education Questionnaire, designed to measure participants' concepts of the nature of science (NOS) and relevant teaching attitudes. The questionnaire includes 15 questions, each followed by several items representing different philosophical positions. Participants rank each item on a five-point scale. The items were empirically based and described from the learners' perspectives, but the issues and subcategories covered were validated by a panel of experts. The latest version was administered to 302 college students. Combined conceptions and conflicting thoughts about NOS were detected. Furthermore, the instrument achieved a test–retest correlation coefficient of 0.82. The questionnaire is a valid and practical tool that can be used to determine participants' conceptions and attitudes toward teaching NOS. With this instrument, science educators and teachers can conduct comparison studies and relate views of NOS to other measurable educational outcomes. © 2006 Wiley Periodicals, Inc. *Sci Ed* **90:**803–819, 2006

INTRODUCTION

The purpose of this study was to develop a valid, meaningful, and practical instrument for creating in-depth profiles of the views of college students or adults, including pre-/ in-service teachers, about the nature of science (NOS), and NOS instruction. McComas particularly highlights the need for a valid standardized assessment tool for two reasons (Good, Lederman, Gess-Newsome, McComas, & Cummins, 2000). First, a standardized instrument is necessary for NOS to be included in large-scale assessments at the state or national level. Although for decades the understanding of NOS has been listed as one of the major objectives in science education (Shulman & Tamir, 1973), it has not yet been

Correspondence to: Sufen Chen; e-mail: sufchen@mail.ntust.edu.tw
Contract grant sponsor: National Science Council.
Contract grant number: NSC 922511S007007.
This paper was edited by former Editor Nancy W. Brickhouse.

adequately emphasized in current educational practice. A valid and reliable instrument would incorporate NOS into large-scale tests, and consequently, NOS is more likely to be explicitly taught and involved in curricula. The second reason that this assessment tool is needed is that valid and reliable instruments provide researchers with a common ground for comparing findings and a feasible tool for studying large sample sizes. Recent studies have reported both negative and positive results regarding teachers' conceptions of NOS (Abd-El-Khalick, Bell, & Lederman, 1998; Abd-El-Khalick & BouJaoude, 1997; Aguirre, Haggerty, & Linder, 1990; Cheung & Toh, 1990; Discenna & Howse, 1998). The inconsistency might have been partially due to different methodological approaches. Additionally, such an instrument enables science educators and teachers to relate NOS views to other measurable outcomes. Finally, although teachers' positive attitudes toward teaching NOS are essential for NOS instruction to succeed, a related attitude survey is not available in the literature. The instrument developed in this study is therefore designed to assess both NOS views and teaching attitudes with a flexibility that allows each part to be used independently.

CRITICISM OF THE TRADITIONAL INSTRUMENTS AND OPEN-ENDED QUESTIONNAIRES

Early research on learners' conceptions of NOS used questionnaires and survey instruments, such as the Science Process Inventory (Welch, 1966), the Test on Understanding Science (Cooley & Klopfer, 1961), the Nature of Science Scale (Kimball, 1967), and the Nature of Scientific Knowledge Scale (Rubba, 1977). These instruments were written from the perspectives of experts. As Jungwirth (1974) and Alters (1997) pointed out, the experts did not adequately represent perspectives of scientists, philosophers, or science educators. Moreover, these instruments often assumed that all scientists hold the same views and behave in the same way. They tended to oversimplify and overgeneralize views of NOS.

Researchers also found that the traditional instruments fail to detect either the subjects' perceptions and interpretations of the test items or their underlying reasons for making choices. Glen Aikenhead conducted a sequence of studies to explicate the semantic discrepancy between experts and students. In one of his studies (1979), he chose an item from the Science Process Inventory (Welch, 1966): "Scientific knowledge is tentative." The result showed that when the option "I do not understand" was available along with the agree–disagree scale, over 25% of 11th- and 12th-grade students chose that option. Hence, the outcome of SPI provides an incomplete picture of students' beliefs. After further study of more than 400 students' paragraph responses to the tentativeness of science, Aikenhead, Fleming, and Ryan (1987) concluded that almost all students believed that scientific knowledge is tentative; nevertheless, the underlying reasons were very different. If the underlying reasons cannot be assessed, the data analysis will be too superficial to reveal the students' conceptions of NOS.

Furthermore, Aikenhead (1988) found that, compared with the traditional Likert items written from experts' viewpoints, empirically derived, multiple-choice responses reduced the ambiguity by up to 20%. In other words, the way the scales were constructed is crucial, and the empirically based approach appeared to be promising. Accordingly, Aikenhead and Ryan (1992) developed an empirically based instrument entitled the Views on Science–Technology–Society (VOSTS). However, Lederman, Abd-El-Khalick, Bell, and Schwartz (2002) and Abd-El-Khalick and BouJaoude (1997) pointed out that subjects might have combinations of views that would not be reflected in the multiple-choice format. In the pilot study for the present project, the author also found that VOSTS has several problems in common with the traditional NOS instruments; i.e., it contains oversimplified, overgeneralized statements and sometimes creates different interpretations between researchers and subjects.

Consequently, in order to avoid the ambiguity of language, NOS studies after the late 1980s shifted from being more quantitative to more qualitative in nature, utilizing more flexible tools, such as the Images of Science Probe (Driver, Leach, Millar, & Scott, 1996), small-group discussion (Solomon, 1992), situated-inquiry interviews (Ryder, Leach & Driver 1999; Welzel & Roth, 1998), reviews of lesson plans and documents, field observations of classrooms and teachers, concept maps, and case studies. Presently the most popular pencil and paper tool is the Views of the Nature of Science Questionnaire (VNOS), developed by Lederman et al. (2002). This widely used questionnaire is composed of 10 open-ended questions. However, it is challenging to participants to fully articulate their views in 40–60 minutes, and it is difficult for researchers to gain the intended information from every participant. For example, a common misconception about the scientific method might be evaluated through three items of VNOS:

> Item 1: What, in your view, is science? What makes science (or a scientific discipline such as physics, biology, etc.) different from other disciplines of inquiry (e.g., religion, philosophy)?
> Item 2: What is an experiment?
> Item 10: Scientists perform experiments/investigations when trying to find answers to the questions they put forth. Do scientists use their creativity and imagination during their investigations?

It is uncertain what percentage of participants would address scientific methods as expected, especially as these items do not directly deal with the scientific method issue. It is too optimistic to assume that participants will articulate the details in these items. Other research methods such as follow-up interviews become necessary to clarify the participants' beliefs. In the present project, all efforts have been made to develop a pencil and paper assessment tool Views on Science and Education Questionnaire (VOSE) aimed at increasing validity and minimizing interpretation biases.

DEVELOPMENT OF VIEWS ON SCIENCE AND EDUCATION QUESTIONNAIRE

The procedure for developing VOSE included three stages. In the first stage, NOS issues were selected based on a thorough review of the literature. Then, a pilot study was conducted to collect empirical data about college students' NOS views and teaching attitudes. The results of the pilot study and previous research were used to determine the content and format of the questionnaire. In the second stage, the items were developed and tested. Meanwhile, two panels of experts reviewed the items for content validity and examined the philosophical meaning of each item. Seven student interviews provided information for content clarity. The third stage consisted of a final test, a retest, interviews, and data analysis to establish validity and reliability.

NOS Issues

NOS involves a wide variety of topics related to the history, philosophy, and sociology of science. VOSE focuses on seven aspects of NOS that are particularly relevant to K-12 science education and widely discussed in previous research (e.g., Kourany, 1998; Good et al., 2000; Schwartz & Lederman, 2002):

1. *Tentativeness of scientific knowledge*. On the one hand, scientific knowledge is durable and not easily changed. On the other hand, all scientific knowledge is subject to change. The change could take at least two forms, evolutionary (Popper, 1975/1998) or revolutionary (Kuhn, 1970). New knowledge may arise by refining the old knowledge according to new evidence or interpreting data from a new standard and worldview.
2. *Nature of observation*. Observations may be affected by the observers' anticipation and preconceptions, i.e., observations are theory laden.
3. *Scientific methods*. There is no universal scientific method. Scientists apply various methods in doing research.
4. *Hypotheses, laws, and theories*. A hypothesis is generally used to represent an immature theory, a speculative law, or a prediction of experimental results (McComas, 1996). A law is used to express what has been observed and to predict what has not yet been observed (Carnap, 1966/1998). A theory is defined in many ways by philosophers of science (Carnap, 1966/1998; Hacking, 1983; Radder, 2003; Suppe, 1977). In this text, theory is defined as an explanation of phenomena and associated laws according to *Benchmarks for Science Literacy* (American Association for the Advancement of Science, 1993). Furthermore, scientists create theories and laws to interpret and describe phenomena. Theories and laws are two different types of knowledge. They both have substantial supporting evidence, and one does not become the other.
5. *Imagination*. Imagination is a source of innovation. Scientists use imagination, along with logic and prior knowledge, to generate new scientific knowledge. Imagination and creativity are often presented together in documents of science education reforms. However, the pilot study found that students who object to imagination and creativity as aspects of NOS have more doubts about imagination but fewer problems with creativity. Therefore, VOSE focuses on imagination to avoid the mixed results created by these two terms.
6. *Validation of scientific knowledge*. This issue focuses on how a theory is accepted by the science community. In principle, the merit of a theory is evaluated based on empirical results. Nevertheless, the science community may also choose a theory by conventions like simplicity and the reputation of the theory's proposers. Furthermore, the norm of the paradigm such as a particular way of practicing science, a worldview, and core theories may influence the science community's judgment of competing theories.
7. *Objectivity and subjectivity in science*. Scientific knowledge is empirically based. Scientists try to be open-minded and apply mechanisms such as peer review and data triangulation to improve objectivity. On the other hand, personal beliefs, values, intuition, judgment, creativity, opportunity, and psychology all play a role in scientific activities. Additionally, science and scientists are influenced by the society, culture, and discipline in which they are embedded or educated. This subjectivity may be reflected in their observations, interpretations, use of imagination, and theory choice. In this text, subjectivity is used to represent factors other than objectivity and rationality.

In addition to questions about these seven aspects of NOS, VOSE includes five questions to examine the teaching attitudes corresponding to five of the NOS topics: teaching about the tentativeness of scientific knowledge, the nature of observation, the scientific method, the relationship between theories and laws, and the subjectivity embedded in science.

Pilot Study

The pilot study was a data collection step following the literature review. Open-ended data were collected, and problems arising from the pilot study were considered in the development of items for VOSE. VOSE was based on VOSTS. Selected VOSTS items and two follow-up interviews were administered to 14 American and 10 Taiwanese preservice secondary science teachers in 2000 and 2001. Several problems were found with the use of VOSTS. First, some statements in the questionnaire are overgeneralized; therefore, VOSE avoids this type of statement. Second, some VOSTS items have ambiguous positions. For example, the most frequent answer out of eight choices to the question about scientists' observations was "scientific observations will not differ very much even though scientists believe different theories. If the scientists are indeed competent their observations will be similar." This statement seemed to imply a positivist view. However, the interview data showed that the subjects had distinct reasons behind the choice, which may be inconsistent with positivism. Some of them claimed that it is because the scientists are very objective and not affected by preconceptions (positivism), while the others argued that it is because the scientists work in the same theoretical framework and expect the same outcomes (historicism/postpositivism). Failure to acknowledge the reasons behind a choice could lead to false analyses. Ambiguous statements are clarified in VOSE to minimize diversity of interpretation and to obtain data similar to that from interviews.

The third problem with VOSTS, concerning the objectivity and subjectivity of science, was that the subjects often stated different opinions about what science or scientists are and what science or scientists ought to be. Therefore, VOSE has two sets of questions to reveal learners' views on both issues. Fourth, the subjects also pointed out that some answers on the VOSTS questionnaire had overlapping meanings and were redundant. Likewise, Botton and Brown (1998) found the "considerable difficulty students [postgraduate trainee science teachers] had when expected to differentiate between some of the statements" (p. 57). This problem significantly lowered the reliability of some VOSTS items. Hence, the overlapping answers were combined in VOSE. Fifth, VOSTS forces a participant to choose only one position. In follow-up interviews, many of them expressed combined views or informed views, which were seldom recorded on the answer sheet. For issues involving multiple phases, the subjects might therefore appear to have inadequate knowledge.

In addition to these five fundamental problems, the pilot study found that prospective science teachers avoided teaching the informed NOS views that they possessed and, ironically, proposed to teach topics about which they had naïve concepts. Evaluating both pre-/in-service teachers' conceptions about NOS and their corresponding teaching attitudes is essential for NOS instruction to succeed. The VOSE questionnaire is therefore designed to assess both. In sum, VOSE contains three parts: views on what NOS is (actual), views on what NOS ought to be (ought), and views on science education closely linked to NOS.

Item Development

The content and format of VOSE were designed to improve on the weaknesses mentioned above. All items were revised from VOSTS or generated according to statements that emerged from the pilot study and the recent literature, such as Khishfe and Abd-El-Khalick (2002), Lederman et al. (2002), and Lin and Chen (2002). Moreover, each statement was modified to reduce indirect inferences about its philosophical stance. Concerning the format, Vazquez-Alonso and Manassero-Mas (1999) have developed a Likert scale for the VOSTS items and a sophisticated scoring scheme for item analysis. Their response and scoring models make the maximum use of the VOSTS data and produce data that can be used for inferential statistics. Likewise, VOSE adapts a similar format which allows participants to rank each item on a five-point scale.

The draft contains 102 items headed by 19 questions or statements such as the following example about the validation of scientific knowledge and subjectivity in science (Question 1):

> When two different theories arise to explain the same phenomenon (e.g., fossils of dinosaurs), will scientists accept the two theories at the same time?

A. Yes, because scientists still cannot objectively tell which one is better; therefore, they will accept both tentatively.
B. Yes, because the two theories may provide explanations from different perspectives, there is no right or wrong.
C. No, because scientists tend to accept the theory they are more familiar with.
D. No, because scientists tend to accept the simpler theories and avoid complex theories.
E. No, the academic status of each theory proposer will influence scientists' acceptance of the theory.
F. No, scientists tend to accept new theories which deviate less from the contemporary core scientific theory.
G. No, scientists use intuition to make judgments.
H. No, because there is only one truth, scientists will not accept any theory before distinguishing which is best.

Most of the items begin with yes or no, followed by a reason that represents a philosophical position. Questions and statements related to teaching NOS are formulated in the same way and are focused on the high school level, as in Question 13:

> The science course in high school should investigate the definitions of and the relationships between hypothesis, theory, and law.

A. Yes, because they represent the structure of scientific knowledge.
B. Yes, because they are the fundamentals of scientific inquiry.
C. No, knowing the definition of and relationships between these terms does not help much in learning scientific knowledge.
D. No, because hypothesis, theory, and law lack definite meaning.

Participants are instructed to read the whole set of items before rating their opinion of each item on a five-point scale, i.e., strongly disagree, disagree, uncertain/no comment, agree, and strongly agree.

Finally, the last two questions are contextualized in a story about two scientists, Scientist B standing for an objective image of science and Scientist A influenced by sociocultural and religious factors. Question 14 is

> From the perspective of science education, what can junior/senior high school students learn from these two scientists?

A. A—scientists should have a conscience when doing research.
B. A—consider both scientific research and social values simultaneously.
C. A—scientific research cannot be totally divorced from socio-cultural values.
D. A—respect the diversity of people.
E. B—scientific research should be completely detached from personal beliefs.
F. B—scientific research should be completely detached from social subjective values.

G. Neither of them provides a good example to learn from because science courses should not involve value-choices.

The story provides participants with a shared context to think about what science ought to be and, for secondary school students, which image of science is worth learning—objective, subjective, both, or neither.

Pilot Test

The draft of VOSE was tested on 120 biology students at junior level or higher at a research university in Taiwan. The results were used to reduce the number of items based on two criteria: variation across respondents and rate of responses to uncertain/no comment. The first criterion is referred to as the item difficulty. If almost every subject took the same position, i.e., the variance is near zero, the item was discarded. Second, items for which over half of the subjects answered uncertain/no comment were revised if there were no other items representing the same philosophical stance; otherwise they were abandoned. Finally, when a question had fewer than three items remaining, the question and its items were deleted.

McFarland (1981) and Schuman and Presser (1981) pointed out that respondents might be affected by the order in which response options are presented. The order effect may depend on the knowledge level of respondents. Although respondents to VOSE are instructed to read all items of a question before ranking on the scale in order to eliminate the order effect, they might not follow the instruction. Therefore, different orders of items were used in the pilot test. The results showed that order effects were present in the questions about the epistemological status of theories and teaching the scientific method. This suggests that the subjects might not have had a firm position in the beginning and were swayed by the first item. In the final test, two forms with different orders were administered to eliminate the order effect. For studies in different contexts, the order effect may exist for other issues.

Expert Review

VOSE was reviewed twice, each time by six Taiwanese experts. These experts were science educators who had published in the field of NOS and were ranked associate professor or above. The literature indicates a rough consensus about the sufficient number of experts for content validity. Yaghmaie (2003) suggests three. Rubio, Berg-Weger, Tebb, Lee, and Rauch (2003) recommend including three to ten experts. A quick survey of the articles published in peer-reviewed journals from EBSCOhost ScoINDEX in the past 2 years found that nine studies used expert panels. Two of the articles did not specify the size of the panels. The other seven studies had five to ten experts. The researcher of the present study therefore concluded that six experts in each panel was appropriate.

The six experts in the first panel examined the items to establish face and content validity. They were requested to evaluate whether (1) the yes/no position matches the reason that follows, (2) the reason is meaningful to science educators and teachers, and (3) the domain of the instrument is relevant to and represents the NOS topics being measured. They were also invited to comment on the format and on how the phrasing of the statements could be improved. Some items were modified or removed accordingly. One of the experts suggested adding ontological questions. Although ontological aspects of NOS are important, they are less highlighted in NOS research and may not be of general interest to science educators. Therefore, in consideration of the length of the questionnaire, they are not included in VOSE.

Next, six experts (including two who did not participate in the first review) were presented with an interpretation of each item that would be used to categorize data. They marked “agree,” “agree with revision,” or “disagree” with comments for each interpretation. A few changes were made according to their comments. Some descriptions were clarified. For example, the term knowledge is too vague and is replaced by theory for the question about tentativeness. Items that had controversial philosophical meanings were revised or eliminated. Only those items that were agreed upon by at least five experts were preserved in the final version. The categories and interpretations of NOS issues are listed in Table 1. The items corresponding to attitudes toward teaching the NOS issues are summarized in Table 2.

Interview

The draft was presented to six college students and a high school student for clarity checking. The students were constantly reminded to make a note of any word, phrase, or sentence that they had to read more than once, that they felt uncomfortable about answering, or about which they were unsure of the meaning. They were also given opportunities to express alternative positions.

Final Test

The items were improved several times according to the students’ and experts’ comments and based on the results of the pilot test. The final version contains 15 questions and 85 items. It was originally written and tested in Chinese, then translated into English. Both the Chinese and English versions can be downloaded from http://homepage.ntust.edu.tw/SUFCHEN/. In the final test, VOSE was administered to 302 junior and senior students majoring in biology, life science, chemistry, physics, foreign languages and literature, and Chinese literature at two research universities in Taiwan. The administration time was about 15 minutes. Twenty-four subjects completed the questionnaire again within 1 to 3 months for test–retest reliability. They were also interviewed for 2 hours following the retest. Their views and attitudes toward teaching each NOS topic were collected. Moreover, each subject was asked to interpret the items and was encouraged to provide examples to justify his/her position. A set of interview questions was designed by the author, and two graduate assistants were trained to conduct and analyze the interviews.

Data Analysis

The five ranking items, from “strongly disagree” to “strongly agree,” are assigned with the numbers 0 to 4. Since most issues of NOS are multifaceted, as shown by the position categories in Table 1, the score for each issue is an average score of several items. The averaged scores are then applied to statistical analyses. For example, the issue of the tentativeness of scientific knowledge could incorporate three phases in the development of scientific knowledge: revolutionary (Kuhn, 1970), cumulative, and evolutionary (Popper, 1975/1998), corresponding to item A, B, and C of the fourth question, respectively:

> Even if the scientific investigations are carried out correctly, the theory proposed can still be disproved in the future.

A. Scientific research will face revolutionary change, and the old theory will be replaced.
B. Scientific advances cannot be made in a short time. It is through a cumulative process; therefore, the old theory is preserved.
C. With the accumulation of research data and information, the theory will evolve more accurately and completely, not being disproved.

TABLE 1
NOS Issues, Reliability, Philosophical Positions, and Item Number Tested by VOSE

Issue	α	Position	Item[a]
Tentativeness	0.34	Revolutionary	4A
		Cumulative[b]	4B
		Evolutionary[b]	4C
Nature of observations	0.47	Theory laden	8A, 8B, 8E
		Theory independent	8C, 8D
Scientific methods	0.48	The universal scientific method[b]	9A, 9B, 9F
		Diverse methods	9C, 9D, 9E
Theories and laws	0.80	Epistemology	
		Discovered[b]	5A, 5B (Theory) 6A, 6B (Law)
		Invented	5D, 5E, 5F (Theory) 6D, 6E (Law)
		Discovered or invented	5C (Theory) 6C (Law)
	0.70	Comparison	
		Laws being more certain[b]	7A, 7B
		Different types of ideas	7C, 7D
Use of imagination	0.71	Yes	3A, 3B
		No[b]	3C, 3D, 3E
Validation of scientific knowledge	0.44	Empirical evidence	1A, 1H
		Paradigm	1C, 1F
		Parsimony	1D
		Authority	1E
		Intuition	1G
Subjectivity and objectivity	0.69	Subjectivity	
		Parsimony	1D (Actual)
		Authority	1E (Actual)
		Paradigm	1C, 1F, 8A[d], 8B (Actual)
		Personal factors	1G,8A[d] (Actual) 15A, 15D, 15H (Ought)
		Sociocultural influence	2A, 2B (Actual) 15B, 15C (Ought)
		Imagination	3A, 3B (Actual)
		Methodology	9D (Actual)

Continued

TABLE 1 (*Continued*)

Issue	α	Position	Item[a]
	—[c]	Neutral	1B (Actual)
	0.69	Objectivity	
		No influence of socioculture	2C, 2D (Actual) 15F (Ought)
		Use no imagination	3C, 3E (Actual)
		Based on experimental facts	5B, 6B, 8D (Actual)
		No influence of personal beliefs	8C (Actual) 15E, 15I (Ought)
		Methodology	8E, 9A, 9B (Actual)
		Overall	1A, 1H (Actual) 15G (Ought)

Notes. [a]The numerical number indicates the question, while the letter represents the response for that question.
[b]The corresponding items were scored in reverse to calculate alphas.
[c]This single item was not calculated for alpha.
[d]This item, although it appeared twice in the paradigm and personal factors subcategories, was input only once for reliability.

TABLE 2
Attitudes Toward Teaching the NOS Issues Tested by VOSE and Reliability

Topic	α	Attitude	Item[a]
Tentativeness	0.81	Teaching the tentativeness of scientific knowledge	12A, 12B
		Avoid teaching the tentativeness of scientific knowledge[b]	12C, 12D, 12E
Nature of observations	0.59	Training students to make objective observations[b]	11A, 11B, 11C
		Revealing the theory-laden nature of observations	11D, 11E
Scientific methods	0.78	Teaching the universal scientific method[b]	10A, 10B, 10C, 10D, 10E, 10F
		Encouraging different methods	10G, 10H, 10I
Theories and laws	0.80	Teaching the relationship between theories and laws	13A, 13B
		Avoid teaching the relationship[b]	13C, 13D
Subjectivity and objectivity	0.75	Teaching subjectivity	
		Personal factors	14A, 14D
		Sociocultural influences	14B, 14C
	0.80	Emphasizing objectivity	
		No influence of personal beliefs	14E
		No influence of socioculture	14F
	—[c]	Value free in science courses	14G

Notes. [a]The numerical number indicates the question, while the letter represents the response for that question.
[b]The corresponding items were scored in reverse to calculate alphas.
[c]The single item was not calculated for alpha.

During the revolutionary phase, knowledge and the way science is practiced are dramatically changed. During the cumulative phase, such as the period of normal science in a paradigm, scientific knowledge is accumulated. In the third phase, the change is relatively minor, and a theory may be refined to incorporate new evidence. If it is desirable for a participant to identify all these three phases, the score that the participant receives for the topic of the development of scientific knowledge is the averaged scores of these three items.

However, when the choices present oppositional stances, two approaches can be utilized to categorize the data, depending on the researcher's purpose. The first approach is to pool the data together by giving priority to the stance that most science educators have identified and reversing the scores of other stances. Each participant will have one score for that issue. For example, a subject may pay attention to the cumulative and evolutionary phases of science, which are often deemed as opposite to the view of revolutionary. The corresponding items can be scored in reverse before being averaged with the item regarding the revolutionary phase. As indicated in the footnotes of Tables 1 and 2, some items were scored in reverse to estimate the reliability of each issue. Second, the oppositional stances can be sorted into subcategories and scored separately. For example, the objective and subjective views of science are dealt with separately in Tables 1 and 2. Subcategories are listed in the tables. The scores of these subcategories may be more meaningful to science educators than a summary score for two reasons. On the one hand, the scores of subcategories convey details about the participants' thoughts. On the other hand, when the items are interpreted with a positive attitude which assumes the best of intentions, many oppositional statements are acceptable and are not necessarily treated as right or wrong. Indeed, multiple perspectives such as the subjectivity and objectivity of science are encouraged.

Furthermore, previous studies have revealed that teachers do not take a specific philosophical stance (Gallagher, 1991; Koulaidis & Ogborn, 1989). Their conceptions are not stable and are sometimes contradictory (Abd-El-Khalick & BouJaoude, 1997; Mellado, 1997). Teachers change their philosophical stance when teaching in different contexts (Hodson, 1993). Therefore, like VNOS, different contextual questions in VOSE provide information to outline a participant's conception of an issue. For example, in the final test, the second and the last two questions indicated that undergraduate students generally felt that social and cultural values influence scientists and scientific investigations, yet they seldom identified with scientists whose research activities are shaped by social and cultural values. Moreover, they believed that interactions between science and society should be made known to secondary school students. In other words, sociocultural influence on science is inevitable, but not the ought-to-be nature of science, and science education should correspond to the reality.

It should be noted that the above-mentioned analysis used raw scores which assumed a linear scale and could not objectively identify the levels of difficulty of the items. In fact, the five-point scale is not usually set at specific intervals, and some items are more difficult than others for the subjects to demonstrate an informed view. For example, concerning the epistemological status of theories and laws, it is normally more difficult for the subjects to learn that laws are invented. A scoring system that applies Rasch measurement would take into account the nonlinear effect and various difficulties of items. The original Rasch model was developed for the analysis of binary items (Rasch, 1960). The rating scale model (Wright & Masters, 1982) is an extension of Rasch for multiple-point scales and is appropriate for those VOSE items that are unidimensional. For issues in VOSE that involve multiple aspects, other models in the Rasch family, such as the multidimensional random coefficients multinomial logit model (Adams, Wilson, & Wang, 1997), can be used. θ values derived from these models should increase the fidelity in the measurement of the average score of each issue and subcategory. Researchers may consider using Rasch measurement rather than raw scores for data analysis.

Finally, the clarity of the items was checked again in the interviews after the final test. The interview data were coded in N6, software developed by QSR International for analyzing qualitative data. For each item, the graduate assistants coded whether the subjects' interpretation was consistent with the researcher's. The percentage of the subjects who presented dissimilar interpretations was calculated to examine the ambiguity issue of validity.

RESULTS AND DISCUSSION

Concerning validity and reliability, Aikenhead and Ryan (1992) and Rubba, Schoneweg Bradford, and Harkness (1996) specify that conventional concepts of validity and reliability cannot apply to empirically developed instruments such as VOSTS. An empirically based instrument is developed from a qualitative perspective, which stresses the trustworthiness and authenticity of data (Erlandson, Harris, Skipper, & Allen, 1993) rather than consistency across constructs and measurements. It yields dependable results because the items originated from the respondent's point of view instead of from the researcher's presumption of reasonable answers, and thus has a high reliability. Likewise, the process of developing the items should have ensured that the instrument measures what it purports to measure, and establishes an inherent validity for the instrument.

Therefore, the developer of VOSE focused on the quality and meaningfulness of the items instead of pursuing a high internal consistency. The reliability of each issue was expected to be within a range of reasonable values, but not a main criterion for item selection. Cronbach's alphas were used to verify the appropriateness of discarding items in the pilot test. The alphas of almost all issues increased after some items were deleted. The exception was the issue of tentativeness, in which seven items were deleted and the alpha dropped to 0.34. The three remaining items belonged to independent subcategories and therefore lacked internal consistency. The final test revealed that Cronbach's alphas of all issues ranged from 0.34 to 0.81 (see Tables 1 and 2). The alphas alter if different subcategories are chosen or pooled. For example, when both the objectivity and subjectivity of science were valued, these two aspects were measured separately, and both alphas were 0.69. However, postpositivists severely question tenets proposing truth and objectivity. Based on postpositivism, students should oppose the myth of objectivity and learn the subjective aspect of science. Thus, the items about objectivity and subjectivity in science would be grouped together with those related to objectivity scored in reverse, and the alpha for the combined issue would be changed to 0.58. The scoring system displayed in Tables 1 and 2 is merely one of many versions. Researchers may adjust the scoring system based on their philosophy of science, definition of desirable NOS views, and purposes of study. The alpha values will alter accordingly.

Furthermore, Creswell (1994) has suggested diversifying the data resources and investigators for triangulations to enhance the internal validity of qualitative research: "Discuss plans to triangulate, or find convergence among sources of information, different investigators, or different methods of data collection" (p. 158). Accordingly, this study embraced multiple sources of information such as literature, interviews, and responses on VOSTS and VNOS to construct items. In addition, the issues and subcategories covered in VOSE were examined by a panel of experts. The interpretation for each item also went through expert review. The procedure was designed to ensure high reliability and validity.

The interview data from 24 subjects were used to establish validity concerning the ambiguity issue. The results showed that five and three of the subjects interpreted, respectively, item G of question 1 (1G) and item D of question 3 (3D) differently from the researcher.

Nevertheless, for the other 83 items, more than 90% of the subjects interpreted the items consistently with the researcher. Overall, the respondents attributed similar meanings to the VOSE items as the developer and the panel of science educators. Moreover, in the final test, fewer than one-fifth of the 302 subjects chose uncertain/no comment on 48 items, whereas a higher percentage of the subjects (30.7–41.6%) selected uncertain/no comment on eight items. The respondents seemed not to have much difficulty making decisions about the vast majority of the statements.

Finally, the entire instrument achieved a test–retest correlation coefficient of 0.82. Botton and Brown (1998) evaluated the dependability of the data drawn from the epistemology section of VOSTS by administering the section to 29 preservice postgraduate science teachers in England twice within a month. On five of the 22 items, the Pearson chi-square statistics accepted the null hypothesis that test/retest responses were independent at the significance level of 0.05. They cross-tabulated the responses from the remaining 17 items, and the average of reselection values was 0.61. The average reliability was increased to 0.72 after some of the responses were grouped. Lin and Chen (2002) retested 11 items from the same section of VOSTS on 34 prospective science teachers in Taiwan within a month and obtained an average reliability of 0.77. Therefore, the stability of VOSE is considered high with the limitation of a small sample size ($n = 24$) and a relatively long duration between the tests (1–3 months).

VOSE data create an in-depth profile of a subject's NOS views and educational ideas. Using VOSE data, researchers can gain an understanding of a subject's knowledge level of an issue as well as concerns and attitudes about teaching that issue. Moreover, VOSE allows respondents to express their opinions on each item. The results showed that respondents often circled agree or strongly agree on more than one item for an issue. For example, concerning the question about how scientists choose a theory between competing theories, some subjects strongly agreed that scientists accept both theories before they can objectively distinguish the best and, on the other hand, also believed that scientists tend to choose the theory which deviates less from the contemporary core theory or the theory whose proposer has higher academic status. In other words, combined conceptions and conflicting thoughts can be detected. Particularly for issues that involve multiple phases, subjects' beliefs in each phase can be assessed.

CONCLUSIONS AND IMPLEMENTATIONS

VOSE is an empirically based instrument that can determine subjects' conceptions and attitudes toward teaching NOS. This dual function is an innovation in assessment tools for NOS research. Moreover, the reasons underlying those conceptions and attitudes are examined. Furthermore, VOSE can be used to perform comparison studies, and can yield data useful for inferential statistics. Although only a few issues are incorporated in VOSE, researchers may apply a similar approach to develop questionnaires for other issues, particularly those emerging from national standards and Osborne, Collins, Ratcliffe, Millar, and Duschl's (2003) Delphi study.

The content and format of VOSE are an improvement on the weaknesses of the traditional questionnaires, VNOS and VOSTS. Unlike traditional questionnaires designed from experts' viewpoints, VOSE items were constructed from learners' perspectives, which reduced their ambiguity. In comparison with VNOS, VOSE reduces the amount of time needed to administer and code data. The administration time of VOSE is manageable in most contexts. Moreover, a field test of the issues covered by both VOSE and VNOS using a group of preservice teachers revealed that the domain of measurement is more focused in VOSE. Although many of them provided insightful views on VNOS, these views were not related to the NOS issues being studied. It was difficult to gain direct information from VNOS. The construct of VOSE is more explicit. Furthermore, with VNOS, participants need skills and a sharp mind to fully articulate their views within the time frame. The preservice teachers felt more frustrated when taking VNOS than when taking VOSE.

In comparison with VOSTS, VOSE has taken four steps forward. First, some items from VOSTS were revised to clarify the underlying reasons and avoid oversimplified and overgeneralized statements. Second, the items were reviewed by a panel of science educators to ensure that the items are meaningful to science educators, and interpretations of the items were confirmed by a panel of science educators and 24 subjects. Third, an analysis of the percentages of subjects selecting uncertain/no comment on the five-point scale showed that the subjects did not have much difficulty making a decision about most items. Fourth, the format of VOSE allows researchers to detect combined conceptions and conflicting thoughts. However, VOSE does not cover as wide a range of topics as VOSTS.

It should be noted that VOSE was tested in Chinese in this study. The various data sources for developing the items, including the subjects in the pilot study and the literature, revealed that subjects in the United States and Taiwan have similar views about NOS. Nevertheless, meanings may be altered in translation. The English version may need further revision and input of local context. A panel of experts may be established to review the items and categories in Tables 1 and 2. Other research methods such as interviews may be used to ensure the validity of the instrument. Furthermore, a small number of students have pointed out that they agreed with reasons for both yes and no positions on some questions, but felt uneasy about choosing inconsistent positions. Subjects like these are aware of the logic behind their choices. Researchers may remind them that there are no right or wrong answers, and other research methods are recommended to verify the results. In addition, the order effect may be a concern for subjects who are unfamiliar with NOS issues. Researchers may remind respondents to read the whole set of items for a question before answering, or researchers may take the order effect into consideration by using two forms that switch the items for yes and no positions.

The conception and attitude sections can be administered together or independently. VOSE can be administered to, but is not limited to, in- and pre-service teachers. Any group of adults may take part in discussions and policy making about science education, including parents, community members, industry representatives, educators, or content experts. Their views about NOS and attitudes toward teaching NOS may influence science curricula in various ways. Therefore, VOSE is not designed exclusively for in- and pre-service teachers. It can be used for any group of college students or adults. Finally, multiple methods are recommended to draw systematic and valid conclusions about subjects' NOS views and teaching attitudes.

I would like to thank Chin-Chung Tsai, Hai-En Yang, Ling Liang, and three reviewers for their insight and thoughtful comments that led to significant improvements in the manuscript. I am especially grateful to Glen Aikenhead for providing VOSTS and to Kai-Ping Yao for suggestions regarding the Rasch model.

REFERENCES

Abd-El-Khalick, F., Bell, R. L., & Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. Science Education, 82, 417 – 436.

Abd-El-Khalick, F., & BouJaoude, S. (1997). An exploratory study of the knowledge base for science teaching. Journal of Research in Science Teaching, 34, 673 – 699.

Adams, R. J., Wilson, M. R., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. Applied Psychological Measurement, 21, 1 – 23.
Aguirre, J. M., Haggerty, S. M., & Linder, C. J. (1990). Student–teachers' conceptions of science, teaching, and learning: A case study in preservice science education. International Journal of Science Education, 12, 381 – 390.
Aikenhead, G. S. (1979). Using qualitative data in formative evaluation. Alberta Journal of Educational Research, 25, 117 – 129.
Aikenhead, G. S., Fleming, R. W., & Ryan, A. G. (1987). High-school graduates' beliefs about Science-Technology-Society, I: Methods and issues in monitoring student views. Science Education, 71, 145 – 161.
Aikenhead, G. S. (1988). An analysis of four ways of assessing student beliefs about STS topics. Journal of Research in Science Teaching, 25, 607 – 629.
Aikenhead, G. S., & Ryan, A. (1992). The development of a new instrument: "Views on Science-Technology-Society" (VOSTS). Science Education, 76, 477 – 491.
Alters, B. J. (1997). Whose nature of science? Journal of Research in Science Teaching, 34, 39 – 55.
American Association for the Advancement of Science. (1993). Benchmarks for science literacy: Project 2061. New York: Oxford University Press.
Botton, C., & Brown, C. (1998). The reliability of some VOSTS items when used with preservice secondary science teachers in England. Journal of Research in Science Teaching, 35, 53 – 71.
Carnap, R. (1998). The confirmation of laws and theories. In J. A. Kourany (Ed.), Scientific knowledge (pp. 164–175). Wadsworth, CA: Belmont. (Reprinted from Philosophical foundations of physics, pp. 3–6, 19–22, 32–35, 225–235, by R. Carnap, 1966, New York: Basic Books.)
Cheung, K. C., & Toh, K. A. (1990). In the eyes of the beholder: Beginning teachers' conception of the nature of science and science teaching. Paper presented at the annual conference of the Educational Research Association, Singapore.
Cooley, W. W., & Klopfer, L. E. (1961). Manual for the test on understanding science. Princeton, NJ: Education Testing Service.
Creswell, J. W. (1994). Research design: Qualitative and quantitative approaches. Thousand Oaks, CA: Sage.
Discenna, J. L., & Howse, M. A. (1998). Biology and physics students' beliefs about science and science learning in non-traditional classrooms. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young people's images of science. Buckingham, UK: Open University Press.
Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). Doing naturalistic inquiry: A guide to methods. Newbury Park, CA: Sage.
Gallagher, J. J. (1991). Perspective and practicing secondary school science teachers' knowledge and beliefs about the philosophy of science. Science Education, 75, 121 – 134.
Good, R., Lederman, N., Gess-Newsome, J., McComas, W., & Cummins, C. (2000). Nature of science: Implications for research, assessment, and teacher education. A symposium and paper presented at the annual international meeting of the Association for the Education of Teachers in Science, Akron, OH.
Hacking, I. (1983). Representing and intervening. Cambridge: Cambridge University Press.
Hodson, D. (1993). Philosophical stance of secondary school science teachers, curriculum experiences, and children's understanding of science: Some preliminary findings. Interchange, 24, 41 – 52.
Jungwirth, E. (1974). Testing for understanding of the nature of science. Journal of College Science Teaching, 3, 206 – 210.
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, 551 – 578.
Kimball, M. E. (1967). Understanding the nature of science: A comparison of scientists and science teachers. Journal of Research in Science Teaching, 5, 110 – 120.
Koulaidis, V., & Ogborn, J. (1989). Philosophy of science: An empirical study of teachers' views. International Journal of Science Education, 11, 173 – 184.
Kourany, J. A. (1998). Scientific knowledge: Basic issues in the philosophy of science. Belmont, CA: Wadsworth Publishing Company.
Kuhn, T. S. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press.
Lederman, N. G., Abd-El-Khalick, F., Bell, B. L., & Schwartz, R. S. (2002). Views of Nature of Science Questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. Journal of Research in Science Teaching, 39, 497 – 521.
Lin, H., & Chen, C. (2002). Promoting preservice chemistry teachers' understanding about the nature of science through history. Journal of Research in Science Teaching, 39, 773 – 792.

McComas, W. F. (1996). Ten myths of science: Reexamining what we think we know about the nature of science. School Science and Mathematics, 96, 10 – 16.

McComas, W. F. (1998). The nature of science in science education: Rationales and strategies. Boston, MA: Kluwer Academic Publishers.

McFarland, S. G. (1981). Effects of question order on survey responses. Public Opinion Quarterly, 45, 208 – 215.

Mellado, V. (1997). Preservice teachers' classroom practice and their conceptions of the nature of science. Science and Education, 6, 331 – 354.

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, 692 – 720.

Popper, K. (1998). The rationality of science revolutions. In J. A. Kourany (Ed.), Scientific knowledge (pp. 286–300). Wadsworth, CA: Belmont. (Reprinted from Problems of scientific revolution: progress and obstacles to progress in the sciences, pp. 72–101, by R. Harre, Ed., 1975, Oxford: Clarendon Press.)

Radder, H. (2003). The philosophy of scientific experimentation. Pittsburgh, PA: University of Pittsburgh Press.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Institute of Educational Research.

Rubba, P. A. (1977). The development, field testing and validation of an instrument to assess secondary school students' understanding of the nature of scientific knowledge. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Rubba, P. A., Schoneweg Bradford, C., & Harkness, W. J. (1996). A new scoring procedure for the Views on Science–Technology–Society instrument. International Journal of Science Education, 18, 387 – 400.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. Social Work Research, 27, 94 – 104.

Ryder, J., Leach, J., & Driver, R. (1999). Undergraduate science students' images of science. Journal of Research in Science Teaching, 36, 201 – 219.

Schuman, H., & Presser, S. (1981). Questions and answers in attitude surveys: Experiments in question form, wording, and context. New York: Academic Press.

Schwartz, R. S., & Lederman, N. G. (2002). "It's the nature of the beast": The influence of knowledge and intentions on learning and teaching nature of science. Journal of Research in Science Teaching, 39, 205 – 236.

Shulman, L., & Tamir, P. (1973). Research on teaching the natural sciences. In R. M. W. Travenal (Ed.), Second handbook of research on teaching (pp. 1098–1140). Chicago: Rand McNally.

Solomon, J. (1992). The classroom discussion of science-based social issues presented on television: Knowledge, attitudes and values. International Journal of Science Education, 14, 431 – 444.

Suppe, F. (1977). The structure of scientific theories (2nd ed.). Chicago: University of Illinois Press.

Vazquez-Alonso, A., & Manassero-Mas, M.-A. (1999). Response and scoring models for the "Views on Science–Technology–Society" instrument. International Journal of Science Education, 21, 231 – 247.

Welch, W. W. (1966). Science Process Inventory, Form D. Minneapolis: University of Minnesota.

Welzel, M., & Roth, W-M. (1998). Do interviews really assess students' knowledge? International Journal of Science Education, 20, 25 – 44.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: Measurement, Evaluation, Statistics, and Assessment Press.

Yaghmaie, F. (2003). Content validity and its estimation. Journal of Medical Education, 3. Retrieved June 10, 2005, from http://www.sbmu.ac.ir/Journal/MedEdu/jme7no1/contents.htm