# APPENDIX A
## Sampling, Data Collection and Coding, Reliability, and Statistical Analyses

Information on the technical aspects of the TIMSS 1999 Video Study of eighth-grade science teaching is provided below. More detailed information can be found in the *TIMSS 1999 Video Study Technical Report, Volume 2: Science* (Garnier et al. forthcoming).

## Sampling

The sampling objective for the TIMSS 1999 Video Study was to obtain a random, nationally representative sample of eighth-grade science lessons in each participating country.[1] Meeting this objective would enable inferences to be made about the national populations of lessons for the participating countries. In general, the sampling plan for the TIMSS 1999 Video Study followed the standards and procedures agreed to and implemented for the TIMSS 1999 assessments (Martin, Gregory and Stemler 2000). The target population for the study consisted of science lessons for students in the eighth year of formal schooling, which corresponds to eighth grade in the five participating countries. All science courses in which eighth-grade students were enrolled were eligible for selection within the sampled schools.

The national research coordinators were responsible for selecting or reviewing the selection of schools and lessons in their country.[2] Identical instructions for sample selection were provided to all of the national research coordinators. For each country, a sample of at least 100 eighth-grade science classrooms was selected for videotaping. In all cases, countries provided the relevant sampling variables to Westat, so that the school samples could be appropriately weighted.

Most of the participating countries drew separate samples for the Video Study and the TIMSS 1999 assessments. For this and other reasons, the TIMSS 1999 assessment data cannot be directly linked to the video database.[3] Complete details about the sampling process in each country can be found in the technical report (Garnier and Rust forthcoming).

### Sample Design

The study made use of a two-stage stratified cluster sampling design. The first stage made use of a systematic probability-proportionate-to-size (PPS) technique to select schools. A PPS sample assigns probabilities of selection to each school proportional to the number of eligible students in the eighth-grade in schools countrywide. Although countries were strongly encouraged to secure the participation of schools selected in the first stage, it was anticipated that a 100 percent participation rate for schools would not be possible in all the countries. Therefore, replacement schools were identified for each originally sampled school, a priori. As each school was selected, the next school in the sampling frame was designated as a replacement school should the originally sampled school choose not to participate in the study.

The second stage consisted of selecting science classes within schools, and finally lesson selection. One eighth-grade science class per school was sampled. The classes were randomly selected from a list of eligible classes in each participating school. The classroom sampling design was to be an equal probability design with no subsampling of students in the classroom. One lesson from each selected science classroom was videotaped. The videotaping date was determined by a scheduler in each country, and was based on scheduling and operational convenience.

[1] Australia, the Czech Republic, the Netherlands, and the United States also collected data on eighth-grade mathematics lessons.
[2] In the United States, Westat selected the school sample and LessonLab, Inc., selected the classroom sample.
[3] Australia conducted a separate study that involved testing the science achievement of the videotaped students

Within the guidelines specified above, each country developed its own sampling strategy. Although countries had to obtain a PPS sample, they were allowed to define strata appropriate for the country.

## Exclusions in the TIMSS Video Sample

Countries were not permitted to substitute schools or classrooms in the study. If a school or teacher declined participation, the next school in the sampling frame was designated as a replacement school. Once a school agreed to participate, the science class to be videotaped was randomly selected from a list of all science classes that enrolled eighth-grade students. Schools were not allowed to select alternative classes or teachers to be videotaped. The teacher and all students in the selected class were videotaped after all legal permissions were obtained (if necessary). Students whose parents or legal guardians requested that they not be included in the study were provided alternative instruction during the videotaped class period and did not participate in the lesson.

## Response Rates

All of the TIMSS 1999 Video Study countries were required to include at least 100 schools in their initial selection of schools; however, some countries chose to include more for various reasons. The TIMSS 1999 Video Study final sample included 439 eighth-grade science lessons across the five countries. Table A.1 indicates the sample size and participation rate for each country.

**TABLE A.1. Sample size and participation rate for each country in the TIMSS 1999 Video Study**

| Country | Number of schools in initial sample | Number of eligible schools that participated | Percentage of eligible schools that participated including replacements[1] – unweighted[2] | Percentage of eligible schools that participated including replacements[1] – weighted[3] |
|---|---|---|---|---|
| Australia | 100 | 87 | 87 | 85 |
| Czech Republic | 100 | 884 | 100 | 100 |
| Japan | 100 | 95 | 95 | 95 |
| Netherlands | 98 | 81 | 83 | 81 |
| United States | 108 | 88 | 82 | 81 |

[1] The participation rates including replacement schools are the percentage of all schools (i.e., original and replacements) that participated.
[2] Unweighted participation rates are computed using the actual numbers of schools and reflect the success of the operational aspects of the study (i.e., getting schools to participate).
[3] Weighted participation rates reflect the probability of being selected into the sample and describe the success of the study in terms of the population of schools to be represented.
[4] Twelve of the lessons selected from the initial sample of 100 schools in the Czech Republic included only economic and political geography content and were excluded from the sample of eligible science lessons.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

The weighted school response rate before replacement is given by the formula:

$$\text{weighted school response rate before replacement} = \frac{\sum_{i \in Y} W_i E_i}{\sum_{i \in (Y \cup N)} W_i E_i},$$

where Y denotes the set of responding original sample schools with grade-eligible students, N denotes the set of eligible non-responding original sample schools, $W_i$ denotes the base weight for school i, $W_i = 1/P_i$, where $P_i$ denotes the school selection probability for school i, and $E_i$ denotes the enrollment size of grade-eligible students, as indicated on the sampling frame.

## Data Collection and Coding

### Data Collection Procedures

Data for the TIMSS 1999 Video Study of eighth-grade science teaching was collected by the contractor for the study, LessonLab, following a standard set of guidelines and specifications. The designated class was videotaped once, in its entirety, without regard to the particular science topic being taught or type of activity taking place. The only exception was that teachers were not videotaped on days they planned to give a test or examination for the entire class period.

Teachers were asked to do nothing special for the videotape session, and to conduct the class as they had planned. The scheduler and videographer in each country determined on which day the lesson would be filmed.

Two cameras were used during each videotaping. One camera was placed at the back or side of the classroom with the widest angle shot of students and the teacher possible. This camera was used to capture an overall shot of the lesson as it occurred. Information from this camera can be used to verify student activities and the degree to which the entire class is focused on the same or similar activities, for example. The second camera was positioned so that it captured what an attentive student would see. For the most part, the second camera focused on the teacher. The second camera was also used to follow the teacher as s/he helped individual students during independent work periods. All videographers were trained extensively using a videographer's training manual. The training manual detailed every aspect of the videotaping procedure, from needed supplies to camera angles to checklists. Detailed information on the videographer's training manual can be found in the *TIMSS 1999 Video Study Technical Report, Volume I: Mathematics* (the data collection procedures were the same for the mathematics and science components of the video study; Jacobs et al. 2003).

The goal was to sample lessons throughout a regular school year, while accommodating how academic years are organized in each country. Most of the filming took place in 1999. In the Czech Republic filming began in 1998 and ended in 1999, and in Japan filming began in 1999 and ended in 2000. The receipt control system tracked the proportion of lessons that arrived from each country on a monthly basis, to ensure there was not a disproportionate number of tapes collected during any given month.

## Questionnaire Data

To help understand and interpret the videotaped lessons, questionnaires were collected from the eighth-grade science teachers of each lesson. The teacher questionnaire was designed to elicit information about the professional background of the teacher, the nature of the science course in which the lesson was filmed, the context and goal of the filmed lesson, and the teacher's perceptions of its typicality. Teacher questionnaire response rates are shown in table A.2.

**TABLE A.2. Teacher questionnaire response rates (unweighted)**

| Country | Number of teachers videotaped | Number of questionnaires completed | Percent returned |
|---|---|---|---|
| Australia | 87 | 87 | 100 |
| Czech Republic | 88 | 88 | 100 |
| Japan | 95 | 95 | 100 |
| Netherlands | 81 | 79 | 98 |
| United States | 88 | 84 | 95 |

SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

The questionnaire was developed in English and consisted of 27 open-ended questions and 32 closed-ended questions. Each country could modify the questionnaire items to make them culturally appropriate. In some cases, questions were deleted from the questionnaires for reasons of sensitivity or appropriateness. Country-specific versions of the questionnaire were reviewed for comparability and accuracy.

The final version of the questionnaire asked science teachers to provide additional information about the videotaped lesson, their background and experience, attitudes, and professional development. The questionnaire included the seven domains listed below:

- contextual information about the videotaped lesson (the content of the lesson, specific goals for student learning, planning for the lesson, and assessment tasks);
- description of the videotaped lesson in the context of a larger unit or sequence of lessons;
- the typicality of the videotaped lesson (teaching methods, student participation, difficulty of the lesson, and effect of the videotaping);
- ideas that guide teaching (teacher's knowledge and personal views of current science teaching);
- educational background, teaching background, and teaching load;
- school characteristics (size, type, how students are admitted, number of teachers of science, and grade levels); and
- attitudes about teaching (attitudes toward work, the students, and science).

Additional details regarding the development of the questionnaire, along with a copy of the U.S. version of the teacher questionnaire, can be found in the technical report (Garnier forthcoming).

Short questionnaires also were distributed to the students in each videotaped lesson; however student data are not presented in this report. More information about the student questionnaire, and a copy of the U.S. version of the student questionnaire, can be found in the technical report detailing the procedures used in the mathematics component of the study (Jacobs et al. 2003, appendix F).

### *Coding of Questionnaire Items*

The teacher questionnaire consisted of both open- and close-ended items. The open-ended items in the teacher questionnaire required development of quantitative codes, a procedure for training coders, and a procedure for calculating inter-coder reliability.

Teachers' responses to open-ended questionnaire items were translated into English by coders who were bilingual in English and the relevant language being used (e.g., Czech). Coding of the open-ended items was then carried out using the English translations.

Separate codes were developed for open-ended items. The codes captured both anticipated responses to the items as well as those not-anticipated but that were provided by teachers. The final set of codes developed for the open-ended teacher questionnaire items reflected the frequency of response, the significance of the code, and the importance of the category for understanding teachers' responses.

## Video Data

This section provides information on the development and application of codes to the video data by four project teams. More details about each of these groups and the codes they developed and applied can be found in the technical report (Lemmens, Garnier, and Roth forthcoming).

### *Science Code Development Team*

An international team was assembled to develop codes to apply to the TIMSS 1999 Video Study science data. The team consisted of country associates (bilingual representatives from each country) and was directed by a science education researcher (see appendix B for team members). The Science Code Development Team was responsible for creating and overseeing the coding process, and for managing the international video coding team. The team discussed coding ideas, created code definitions, wrote a coding manual, gathered examples and practice materials, designed a coder training program, trained coders and established reliability, organized quality control measures, consulted on difficult coding decisions, and managed the analyses and write-up of the data.

The Science Code Development Team worked closely with two advisory groups: a group of national research coordinators representing each of the countries in the study, and a steering committee consisting of five North American science education researchers (see appendix B for advisory group members).

## *International Video Coding Team*

Members of the International Video Coding Team represented all of the participating countries (see appendix B for team members). They were fluently bilingual so they could watch the lessons in their original language, and not rely heavily on the English-language transcripts. In almost all cases, coders were born and raised in the country whose lessons they coded.

Coders in the International Video Coding Team applied 174 codes in 11 coding dimensions to each of the videotaped lessons.

## *Specialist Coding Teams*

The majority of codes for which analyses were conducted for in this report were applied to the video data by members of the international video coding team, who were cultural insiders and fluent in the language of the lessons they coded. However, not all of them were experts in science or teaching. Therefore, two specialist coding teams with expertise in the area of science were employed to create and apply special codes regarding the scientific nature of the content and the discourse in the science lessons.

- ***Science Content Coding Team.*** The Science Content Coding Team was comprised of individuals with expertise in science content and science education (see appendix B for group members). They developed and applied a series of codes to all of the scientific content in the videotaped lessons.

  The Science Content Coding Team constructed a comprehensive, detailed, and structured list of the predominant scientific topics covered in eighth-grade in all participating countries. In addition to coding the nature of the scientific topics, the group also coded the types of science knowledge, the level of difficulty of the science content, and the different modes of content development (see chapter 4 for definitions of science content topics and types of science knowledge, and chapter 5 for definitions of level of content difficulty and modes of content development).

- ***Text Analysis Team.*** The Text Analysis Team used all portions of the science lesson transcripts designated as public interaction to conduct various text analyses (see appendix B for group members). The group utilized specially designed computer software for these quantitative analyses of classroom talk.

  Because of resource limitations, computer-assisted analyses were applied to English translations of lesson transcripts.[4] In the case of the Czech Republic, Japan, and the Netherlands, all lessons were translated from the respective native languages.

[4] Transcribers/translators were fluent in both English and their native language, educated at least through eighth-grade in the country whose lessons they translated, and had completed 2-weeks training in the procedures detailed in the TIMSS 1999 Video Study Transcription and Translation Manual (available in Garnier et al. forthcoming). A glossary of terms was developed to help standardize translation within each country.

## Reliability

### Questionnaire Coding Reliability

Separate codes for each open-ended item were developed using a four-phase process. First, categories of anticipated responses were developed based on current research in teaching and learning and advice from subject matter specialists. This part of the process helped the code developers (1) form a common interpretation of the question, (2) identify categories that may not be provided in the teachers' responses, and (3) address culturally specific issues, such as the meanings of phrases used in the different countries. Second, categories were further developed based on the responses from the first 10 teacher questionnaires received from each country. Third, codes were created using the categories generated in the preceding two phases considering frequencies of responses, the cultural significance of a code, and the importance of a category in understanding teachers' beliefs and goals. Fourth, the codes were checked for reliability. Using these results, the codes were further revised and then applied to the remainder of the questionnaires.

Coders initially reviewed the codes with the code developers, practiced applying codes to teacher responses from five questionnaires, and then discussed the codes with the code developers to resolve any questions. For each item, two coders independently coded 10 randomly selected lessons from each country. All codes applied to the open-ended items had to meet an 85 percent inter-coder reliability, at a minimum. If the 85 percent reliability criterion was not achieved initially, discrepancies were discussed, and necessary modifications were made to the code definition. Reliability was then attempted on a different, randomly selected set of lessons. The reliability procedures were similar to those used in the TIMSS 1995 assessment to code students' responses to the open-ended tasks (Mullis, Jones, and Garden 1996; Mullis and Martin 1998). The analyses of teacher responses included in this report are based on codes that met or exceeded this criterion.

For the five extended-response items describing teachers' educational backgrounds, each lesson was reviewed by the questionnaire coding team. This procedure ensured that each lesson would be reviewed and judged by a team member familiar with that country's educational system. The teams were required to come to consensus on the codes for each lesson, referring to documents describing each country's educational system and consulting with the national research coordinators to resolve any disagreements.

### Video Coding Reliability

The members of the Science Content Coding Team each established reliability through consensus coding of all the team members.

Percentage agreement was used to estimate inter-rater reliability and the reliability of codes applied by the International Video Coding Team within and across countries for all variables presented in the report. The procedures were based on those previously used and documented for the TIMSS 1995 Video Study and as described in the literature (Bakeman and Gottman 1997). Percentage agreement allows for consideration of not only whether coders applied the same codes to a specific action or behavior, for example, but also allows for consideration of whether the coders applied the same codes within the same relative period of time during the lesson. That is, the reliability of coding in this study was judged based on two general factors: (1) that the same code was applied and (2) that it was applied during the same relative time segment in the lesson. Thus, it was not deemed appropriate to simply determine that the same codes were applied, but that they were applied to the same point in the lesson (here referred to as time segment) as well.

The calculation of percentage of agreement in this study is defined as the proportion of the number of agreements to the number of agreements and disagreements (Bakeman and Gottman 1997). Table A.3 reports the reliability of applying codes to the video data at two points: at or very near the beginning of applying codes (initial reliability) and at the midpoint of applying codes to the video data (midpoint reliability). Coders established initial reliability on all codes in a coding pass prior to their implementation. After the coders finished coding approximately half of their assigned set of lessons (in most cases about 40-50 lessons), coders established midpoint reliability. The minimum acceptable reliability score for each code (averaging across coders) was 85 percent. Individual coders or coder pairs had to reach at least 80 percent reliability on each code.[5]

Initial reliability was computed as agreement between coders and a master document. A master document refers to a lesson or part of a lesson coded by consensus by the Science Code Development Team. To create a master, the country associates independently coded the same lesson and then met to compare their coding and discuss disagreements until consensus was achieved. Masters were used to establish initial reliability. This method is considered a rigorous and cost-effective alternative to inter-coder reliability (Bakeman and Gottman 1997).

Midpoint reliability for each of the 11 coding dimensions was computed as agreement between pairs of coders. By halfway through the coding process, coders were considered to be more expert in the code definitions and applications than the Science Code Development Team. Therefore, in general, the most appropriate assessment of their reliability was deemed to be a comparison among coders rather than to a master document. Each midpoint reliability check involved pair coding of five randomly-selected lessons, one from each country. For each coding dimension, or pass, a different set of five lessons was randomly selected. When there were disagreements between pairs of coders, the Science Code Development Team met to resolve the disagreement. Pair-rater agreement was also used to establish initial reliability in some of the later coding passes, but only for those codes for which coders helped to develop coding definitions.

In each of 11 coding dimensions, a minimum of 15 lessons were coded independently by two or more international coding team members: ten lessons used in the initial training and reliability tests and five lessons at midpoint. Because consensus coding was used in the content coding dimensions, all of the lessons were examined by two or more coders for science content codes.

A percentage agreement reliability statistic was computed for each coder by dividing the number of agreements by the sum of agreements and disagreements (Bakeman and Gottman 1997). Average reliability was then calculated across coders and across countries for each code.

Codes were dropped from the study if 85 percent reliability could not be achieved. As indicated in table A.3, all codes presented in the report met or exceeded the minimum acceptable reliability standard established for this study.

[5] The minimum acceptable reliability score for all codes (across coders and countries) was 85 percent. For coders and countries, the minimum acceptable reliability score was 80 percent. That is, the reliability of an individual coder or the average of all coders within a particular country was occasionally between 80–85 percent. In these cases clarification was provided as necessary, but re-testing for reliability was not deemed appropriate.

In cases where coders did not reach the established reliability standard, they were re-trained and re-tested using a new set of lessons. Coders who still could not achieve the reliability standard did not code the given pass. Coders not achieving 85 per cent reliability at midpoint were also re-trained and re-tested. In addition, all previously coded lessons in that dimension were checked by a code development team member and changes made as appropriate. Coders who did not reach the defined standard even after re-training were not permitted to code for that dimension or any future dimensions that depended on knowledge of that dimension.

What counted as an agreement or disagreement depended on the specific nature of each code, and is explained in detail in Lemmens, Garnier, and Roth (forthcoming). Some codes required coders to indicate a time. In these cases, coders' time markings had to fall within a predetermined margin of error. This margin of error varied depending on the nature of the code, ranging from 10 seconds to 2 minutes. Rationales for each code's margin of error are provided in Lemmens, Garnier, and Roth (forthcoming).

Exact agreement was required for codes that had categorical coding options. In other words, if a code had four possible coding categories, coders had to select the same coding category as the master. In most cases, coders had to both mark a time (i.e., note the in- and/or out-point of a particular event) and designate a coding category. In these cases, it was first determined whether coders reliably marked the same or nearly the same in- and out-points, within the established margin of error. If reliability could not be established between coders based on marking the in- and out-time of codes, then reliability for the actual coding category was not calculated. In these cases, as explained above, coders were re-trained and re-tested using a different set of lessons.

Percentage agreement was used to estimate inter-rater reliability and the reliability of the codes within and across countries for all the variables presented in this report. Percentage agreement allowed us to take into account the markings of both in- and out-points of the codes applied to the videotaped lessons when computing the reliability for a code. All three marks (i.e., in-point, out-point, and label) were included in the calculation. Percentage agreement was selected to calculate reliability for all codes because most codes included marking times as well as labels.

While initial and midpoint reliability rates are reported, coders were monitored throughout the coding process to avoid reliability decay. If a coder did not meet the minimum reliability standard, additional training was provided until acceptable reliability was achieved. The data reported only include data from coders who were evaluated as reliable.

A variety of additional quality control measures were put in place to ensure accurate coding. These measures included: 1) discussing difficulties in coding reliability lessons with the science code development team and/or other coders, 2) checking the first two lessons coded by each coder, either by a code developer or by another coder, and 3) discussing hard-to-code lessons with code developers and/or other coders.

Table A.3 lists the initial and midpoint reliability scores for each code, averaged across coders.

**TABLE A.3. Initial and midpoint reliability statistics for each science code applied by the International Coding Team, by code: 1999**

| Code | Initial reliability[1] (percent) | Midpoint reliability[2] (percent) |
|---|---|---|
| Lesson structure | | |
| Lesson (LSSN) | 96 | 99 |
| Science instruction (SI) | 94 | 95 |
| Science organization (ORG) | 90 | 90 |
| Non-science (NS) | 94 | 92 |
| Technical difficulties (TD) | 96 | 96 |
| Classroom talk | | |
| Public talk (PUBL) | 92 | 94 |
| Teacher-student interaction (TSI) | 95 | 98 |
| Social structure | | |
| Individual work (AP1) | 92 | 94 |
| Pair work (AP2) | 97 | 97 |
| Group work (AP3) | 98 | 99 |
| Other work (AP4) | 95 | 98 |
| Activity structure | | |
| Copying notes (CN) | 93 | 98 |
| Divided class work (DC) | 100 | 100 |
| Silent reading (IR) | 96 | 96 |
| Whole-class work (PDF) | 99 | 100 |
| Independent seatwork work (WA) | 97 | 95 |
| Independent practical work (WP) | 92 | 91 |
| Whole-class seatwork activities (PD) | 100 | 98 |
| Whole-class practical activities (PPD) | 99 | 98 |
| Purpose | | |
| Administrative purpose (ADM) | 99 | 97 |
| Assessing student learning (AS1) | 96 | 94 |
| Going over assessment (AS2) | 94 | 96 |
| Developing new content (DEV) | 95 | 94 |
| Assigning homework (HW1) | 96 | 96 |
| Going over homework (HW2) | 96 | 95 |
| Review (REV) | 98 | 100 |
| Students coming to the front of class (SCF) | 97 | 94 |
| Homework start in class (HWS) | 95 | 97 |
| Type of homework (HWT) | 99 | 98 |
| Students pace their own work (PAC) | 98 | 99 |
| Independent practical activities | | |
| Writing (LW) | 86 | 86 |
| Diagrams (DD) | 87 | 94 |
| Graphs (GRP) | 97 | 97 |
| Mathematics calculations (MP) | 95 | 98 |

See notes at end of table.

**Table A.3. Initial and midpoint reliability statistics for each science code applied by the International Coding Team, by code: 1999—Continued**

| Code | Initial reliability[1] (percent) | Midpoint reliability[2] (percent) |
|---|---|---|
| Content development | | |
| Density of science ideas (2) | 87 | 94 |
| Making connections/ acquiring information (22) | 92 | 93 |
| Goal statements (2) | 100 | 100 |
| Summary statements (3) | 96 | 100 |
| Focus of lesson (content/activity) (18) | 98 | 100 |
| Textbook use (19) | 100 | 100 |
| First-hand data (8) | 98 | 98 |
| Phenomena (17) | 97 | 98 |
| Visual representations (11) | 99 | 98 |
| Rigor (20) | 99 | 98 |
| Learning environment | | |
| Rooms (RM) | 98 | 100 |
| Computers (C) | 97 | 98 |
| Chalkboards (CB) | 99 | 98 |
| Overhead projectors (OH) | 97 | 94 |
| Adult teaching assistants (TA) | 100 | 100 |
| Video recorders (TC) | 90 | 93 |

[1] Initial reliability refers to reliability established on a designated set of lessons before coders began work on their assigned lessons.
[2] Midpoint reliability refers to reliability established on a designated set of lessons after coders completed approximately half of their total assigned lessons.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

## Data Reliability

Estimates produced using data from the TIMSS 1999 Video Study are subject to two types of error, sampling and nonsampling errors. Nonsampling errors can be due to errors made in the collection and processing of data. Sampling errors can occur because the data were collected from a sample rather than a complete census of the population.

### *Nonsampling Errors*

Nonsampling error is a term used to describe variations in the estimates that may be caused by population coverage limitations, nonresponse bias, and measurement error, as well as data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, the differences in respondents' interpretations of the meaning of the questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

In general, it is difficult to identify and estimate either the amount of nonsampling error or the bias caused by this error. In the TIMSS 1999 Video Study, efforts were made to prevent such errors from occurring and to compensate for them when possible. For example, the design phase entailed a field test that evaluated items as well as the implementation procedures for the survey.

Another potential source of nonsampling error was respondent bias, which occurs when respondents systematically misreport (intentionally or unintentionally) information in a study. One potential source of respondent bias in this survey was social desirability bias. For example, teachers may report that they assign more homework than would be observed through classroom observation. If there were no systematic differences among specific groups under study in their tendency to give socially desirable responses, then comparisons of the different groups will accurately reflect *differences* among groups. In order to minimize bias, all items were subjected to field tests. Readers should be aware that respondent bias may be present in this survey as in any survey. It was not possible to state precisely how such bias may affect the results.

### *Sampling Errors*

Sampling errors occur when the discrepancy between a population characteristic and the sample estimate arises because not all members of the reference population are sampled for the survey. The size of the sample relative to the population and the variability of the population characteristics both influence the magnitude of sampling error. The sample of science classrooms from the 1998–1999 school year was just one of many possible samples that could have been selected. Therefore, estimates produced from the TIMSS 1999 Video Study sample may differ from estimates that would have been produced from other samples. This type of variability is called sampling error because it arises from using a sample of science classrooms in 1998–1999, rather than all science classrooms in that year.

The standard error is a measure of the variability due to sampling when estimating a statistic. Standard errors for estimates presented in this report were computed for each country using the jackknife technique. Standard errors can be used as a measure for the precision expected from a particular sample.

Standard errors for all of the estimates are included in appendix C to this report. These standard errors can be used to produce confidence intervals. There is a 95 percent chance that the true average lies within the range of 1.96 times the standard errors above or below the estimated score. For example, it was estimated that 58.0 percent of U.S. science instruction time was devoted to public talk, and this statistic had a standard error of 4.0. Therefore, it can be stated with 95 percent confidence that the actual percentage of U.S. science instruction time devoted to public talk for the total population in 1998–1999 was between 50.16 and 65.84 percent (1.96 X 4.0= 7.84; confidence interval = 58.0 +/- 7.84).

## Data Entry and Cleaning Procedures

Most codes for the TIMSS 1999 Video Study were entered directly into the multimedia database, so that the videotapes and English transcripts could be linked directly with specific codes. The data then were exported either in spreadsheet format for statistical analyses, or in table format for further study by specialist coding groups. In some cases, where the vPrism software was not usable with particular types of coding, codes were entered into an Excel spreadsheet.

Codes from Dimensions 1–7, 9, 10, and 12 were entered directly into a vPrism database. Codes from Dimensions 8 and 11 were entered into an Excel database and the transcripts were analyzed in a custom-made text analysis software program.

A data cleaning process was put in place for both the vPrism and Excel databases. For the vPrism data, coders first recorded their coding decisions in writing onto printed lesson transcripts. Then they entered this information into vPrism. Lastly, coders exported the vPrism data for each lesson and compared it to their markings on the transcripts. In this way, data entry errors were immediately noted and corrected. In addition, errors detected through preliminary data analyses were examined and corrected. For example, coding that was outside of a possible range was detected and extreme outliers on particular codes were studied.

For the Excel data, coders first recorded their coding decisions in writing onto a printed spreadsheet for each lesson. Then they entered this information into Excel. Every tenth lesson was checked for accuracy, and errors were corrected.

Once they were cleaned, all of the data were aggregated to the lesson level, with each coding dimension in a separate datafile.  The full sample and replicate weights were then appended to each file. Finally, statistical analyses were run using the weighted data in Wesvar and/or SPSS.

## Transcription and Translation

The videotapes of science lessons were digitized and entered into a multimedia database. This made the videotapes available through a network server. All non-English videotapes were transcribed and translated into English. Translation of the videotapes was handled through a team of translators who were hired on the basis of their fluency in both English and in the language of instruction being studies.  A science background was also a strong determinant of hire.

Each translator and transcriber participated in a 2-week training period, during which they were instructed in the transcription convention requirements and operation of the specialized vPrism software. Details of the procedure can be found in the TIMSS 1999 Video Study Transcription and Translation Manual, included in appendix A of the *TIMSS 1999 Video Study Technical Report: Volume 1: Mathematics Study* (Jacobs et al. 2003).

Each videotaped lesson was processed and reviewed by two transcribers prior to its final processing and review by the transcription manager.  Every audible utterance by the teacher and students was translated into English from the original language.  The initial translation of each lesson was reviewed up to three times before its review by a second translator.  The second translator made any necessary adjustments to the translation by comparing it to the original videotaped lesson.  Each lesson was reviewed in its entirety up to six times (three times per translator).  As an additional quality control measure, completed translations were selected at random and checked line-by-line by the transcription manager, with the assistance of a translator.

## Weighting

Sampling weights were developed to allow for the computation of statistically sound, nationally representative estimates. Weighting adjusts for various situations such as school nonresponse because data cannot be assumed to be randomly missing.  The base weight for each lesson/class selected was the reciprocal of the product of the school selection and classroom selection probabilities.  The lesson/class base weights have the following property: had all schools participated, then the sum of these weights across the entire sample within the country would give an unbiased estimate of the total number of lessons in a country (or close to an unbiased estimate when replacement schools were used). In the absence of nonresponse, the lesson/class base weights are a mechanism to provide valid generalizations from the sample to the national population.

To account for nonresponse in cases where a sampled school had one or more eligible classes but none was videotaped, a nonresponse adjustment was created. The idea behind nonresponse adjustments was to compensate for missing data from nonresponding schools by increasing the weights of similar responding schools. To accomplish this, schools were grouped into cells. There were three principles for forming cells: (1) schools within the same cell should be somewhat similar with respect to characteristics that might relate to the phenomena being studied; (2) there were at least six responding schools in each cell; and (3) as many cells could be formed as were reasonable given restraints 1 and 2.

Nonresponse cells were generally based on sampling stratification variables. The final weight for the lesson/class selected from a school was given as the product of the lesson/class base weight, $BW_i$, and the nonresponse adjustment factor for the cell to which the school belongs, $NRF_i$:

$$FW_i = BW_i \; x \; NRF_i$$

## Variance Estimation Using the Jackknife Technique

Sampling variances were computed for each country using the jackknife technique. This technique takes into account the design used to select the lesson/class samples as well as the effect on sampling variance due to the nonresponse adjustments. Nonresponse adjustments were computed in order to mitigate against any nonresponse bias. However, since these adjustments involved calculating ratios of sample estimates within cells and then applying these ratios to the weights, they also have an impact on the sampling variances of estimates derived from the study. The variance estimates obtained via the jackknife approach reflect this appropriately.

The jackknife technique is described in detail in Wolter (1985) and summarized in Rust (1985) and Rust and Rao (1996). The jackknife technique used in the TIMSS 1999 Video Study is essentially the same as that used in the 1995 and 1999 TIMSS assessment studies, and the TIMSS 1995 Video Study.

## Statistical Analyses

Most of the analyses presented in this report are comparisons of means or distributions across five countries for video data and questionnaire data. The TIMSS 1999 Video Study was designed to provide information about and compare science instruction in eighth-grade classrooms. For this reason, the lesson rather than the school, teacher, or student was the unit of analysis in all cases.

Analyses were conducted in two stages. First, means or distributions were compared across all available countries using either one-way ANOVA or Pearson Chi-square procedures. For some continuous data, additional dichotomous variables were created that identified either no occurrence of an event (code = 0) or one or more occurrences of an event (code = 1). Variables coded dichotomously were usually analyzed using ANOVA, with asymptotic approximations.

Next, for each analysis that was significant overall, pairwise comparisons were computed and significance determined by the Bonferroni adjustment. The Bonferroni adjustment was made assuming all combinations of pairwise comparisons. For continuous variables, Student's *t* values were computed on each pairwise contrast. Student's *t* was computed as the difference between the two sample means divided by the standard error of the difference. Determination that a pairwise contrast was statistically significant with $p<.05$ was made by consulting the Bonferroni *t* tables published by Bailey (1977). For categorical variables, the Bonferroni Chi-square tables published in Bailey (1977) were used.

The degrees of freedom were based on the number of replicate weights, which was 50 for each country. Thus, in any comparison between two countries there were 100 replicate weights, which were used as the degrees of freedom.

A significance level criterion of .05 was used for all analyses. All differences discussed in this report met at least this level of significance, unless otherwise stated. Terms such as "less," "more," "greater," "higher," or "lower," for example, are applied only to statistically significant comparisons. The inability to find statistical significance is noted as "no measurable differences detected" or a similar phrase. In this latter case, failure to find a statistically significant difference should not be interpreted to mean that the estimates are the same or similar; rather, failure to find a difference may be due to measurement or sampling error.

All tests were two-tailed. Statistical tests were conducted using unrounded estimates and standard errors, which also were computed for each estimate. Standard errors for estimates shown in figures in the report are provided in appendix C.

The analyses reported here were conducted using data weighted with survey weights, which were calculated specifically for the classrooms in the TIMSS 1999 Video Study (see Rust forthcoming for a more detailed description of weighting procedures).
The coefficient of variation (CV) was calculated for all reported estimates. In cases where the CV was found to be .50 or greater, the estimate was marked as unstable (!) in all tables and figures. Comparisons among unstable estimates are not made in this report. The CV was calculated by dividing the standard error of the estimate by the estimate.

# APPENDIX B
## Participants in the TIMSS 1999 Video Study of Science Teaching

## Director of TIMSS 1999 Video Study of Science Teaching

Kathleen J. Roth

## Associate Director of TIMSS 1999 Video Study of Science Teaching

Stephen L. Druker

## Directors of TIMSS 1999 Video Study

Ronald Gallimore
James Stigler

## National Research Coordinators

*Australia*
Jan Lokan
Barry McCrae
John Cresswell
*Czech Republic*
Jana Strakova
*Japan*
Shizuo Matsubara
Yasushi Ogura
*Netherlands*
Klaas Bos
Hans Pelgrum
*United States*
Patrick Gonzales

## U.S. Steering Committee

Rodger Bybee
James J. Gallagher
Kathleen Hogan
Jim Minstrell
Senta Raizen

## Chief Analyst

Helen Garnier

## Associate Analyst

Meike Lemmens

## Director, Public Release Lessons

Catherine Chen

## Country Associates

*Australia*
David Rasmussen
*Czech Republic*
Svetlana Trubacova
*Japan*
Takako Kawanaka
Yukari Okamoto
*Netherlands*
Dagmar Warvi
*United States*
Catherine Chen

## International Video Coding Team

*Australia*
Azaro
Mark Durston
Paul Fischer
Akemi Phillips
David Rasmussen
*Czech Republic*
Renata Ferrari
Jana Hatch
Svetlana Trubacova
*Japan*
Yui Omine
Jun Yanagimachi
Kazumi Yoshihara
*Netherlands*
Meike Lemmens
Yasmin Penninger
Dagmar Warvi
Tom Young
*United States*
Akemi Phillips
Marcie Gilbert
Cynthia Simington

## Science Content Coding Team

Ivonne Budianto
Catherine Chen
Constance Christensen
Stephen Druker
Patrick Lam
Angelica Mejia
Alvaro Mercado
Vladislav Mikulich
Mark Valderrama

## Consultants

Charles W. Anderson
Judith Edgington
Karen Givvin
Hilary Hollingsworth
Jennifer Jacobs
Gregory Kelly
Yukari Okamoto
Jo Ellen Roseman
Justus J. Schlichting
Nanette Seago
Edward L. Smith

## Editors

Patrick Gonzales
Erin Pahlke
Lisette Partelow

## Analysis Team

Helen Garnier
Meike Lemmens
Kathleen Roth

## Text Analysis Group

Keith Cascio
Stephen Druker
Don Favareau
Ronald Galimore
Takako Kawanaka
Bruce Lambert
Meike Lemmens
David Lewis
Fang Liu
Samer Mansukhani
Genevieve Patthey-Chavez
Rodica Waivio
Dagmar Warvi
Clement Yu

## Questionnaire Development Team

Sister Angelo Collins
Helen Garnier
Kathleen Hogan
Jennifer Jacobs
Kathleen Roth

## Questionnaire Coding Team

Catherine Chen
Stephen Druker
Helen Garnier
Leanne Klein
David Rasmussen
Kathleen Roth
Svetlana Trubacova
Dagmar Warvi

## Proofreaders

Catherine Chen
Renata Ferrari
Kamaliah N. Lewis
Amee Mehta
Yukari Okamoto
Akemi Phillips
David Rasmussen
Svetlana Trubacova
Dagmar Warvi
Jun Yanagamachi

## Field Test Team

Karen Givvin
Jennifer Jacobs
Takako Kawanaka
Christine Pauli
Jean-Paul Reeff
Nick Scott
Svetlana Trubacova

## Chief Videographers

Maria Alidio
Takako Kawanaka
Scott Rankin

## Videographers

Sue Bartholet
Talegon Bartholet
Michaela Bractalova
Gabriel Charmillot
Matthias Feller
Ruud Gort
Christopher Hawkins
Kurt Hess
Rowan Humphrey
Narian Jagasia
LeAnne Kline
Tadayuki Miyashiro
Silvio Moro
Selin Ondül
Mike Petterson
Stephen Skok
Sikay Tang
Giovanni Varini
Sofia Yam
Jiri Zeiner
Andreas Zollinger

## Field Test Videographer

Ron Kelly

## Computer Programming

Paul Grudnitski
Daniel Martinez
Ken Mendoza
Carl Manaster
Rod Kent

## Transcription and Translation Directors

Lindsey Engle
Don Favareau
Wendy Klein
Petra Kohler
David Olsher
Susan Reese

## Transcribers and Translators

*Australia*
Marco Duranti
Hugh Grinstead
Amy Harkin
Tammy Lam
Tream Anh Le Duc
James Monk
Aja Stanman
Elizabeth Tully
Daniella Wegman

*Czech Republic*
Barbara Brown
Silvie Fabikova
Jana Hatch
Peter Kasl
Vladimir Kasl
Jirina Kyas
Alena Mojhova
Vaclav Plisek

*Japan*
Kaoru Koda
Yuri Kusuyama
Ken Kuwabara
Emi Morita
Angela Nonaka
Naoko Otani
Jun Yanagimachi

*Netherlands*
Hans Angessens
Annemiek deHaan
Tony DeLeeuw
Neil Galanter
Maaike Jacobson
Maarten Lobker
Yasmin Penninger
Linda Pollack
Silvia Van Dam

*United States*
Ginger (Yen) Dang
Jake Elsas
Jordan Engle
Steven Gomberg
Barry Griner
Jaime Gutierrez
Sydjea Johnson

Keith Murphy
Kimberly Nelson
Raoul Rolfes
Tosha Schore
Budie Suriawidjaja

### Administrative Support

Maria Alidio
Cori Busch
Ellen Chow
Olivier de Marcellus
Melanie Fan
Tammy Haber
Christina Hartmann
Gail Hood
Rachael Hu
Brenda Krauss
Samuel Lau
Phil Makris
Yukiyo Miyajima
Francesca Pedrazzini-Pesce
Liz Rosales
Rossella Santagata
Eva Schaffner
Yen-lin Schweitzer
Cynthia Simington
Kathya Tamagni-Bernasconi
Vik Thadani
Laura Wagner
Sophia Yam
Andreas Zollinger

### Video Processing

Don Favareau
Tammy Haber
Petra Kohler
Brenda Krauss
Miriam Leuchter
David Martin
Alpesh Patel
David Rasmussen
Susan Reese
Liz Rosales
Steven Schweitzer
Mark Valderrama

### Video Clip Production Team

Catherine Chen
David Rasmussen
Kathleen Roth

# APPENDIX C
## Standard Errors for Estimates in Figures and Tables

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Figure 2.1 | Graduate degree | 3.6 | 0.0 | 3.7 | 5.8 | 6.7 |
| | Undergraduate degree | 4.1 | ‡ | 3.7 | 5.8 | 6.7 |
| | Below undergraduate degree | 2.3! | ‡ | ‡ | ‡ | ‡ |
| Text | Certified to teach grade 8 or higher | 5.7 | 1.8 | 0.0 | 0.2 | 5.9 |
| Text | Certified to teach lower than grade 8 | 1.6! | 1.8! | 0.0 | 0.2! | 5.8! |
| Table 2.1 | Science – Total | 3.3 | 2.3 | 0.0 | 1.1 | 6.5 |
| | Life sciences | 6.9 | 5.4 | 4.1 | 5.7 | 6.9 |
| | Physics | 4.7 | 4.8 | 4.4 | 5.8 | ‡ |
| | Chemistry | 4.9 | 5.2 | 5.0 | 6.0 | 1.8 |
| | Earth sciences | 4.6 | 5.4 | 3.2 | 3.3! | 2.7 |
| | General science | 2.2! | ‡ | 0.0 | ‡ | 3.2 |
| | Other than science | 3.3 | 2.3 | ‡ | ‡ | 6.5 |
| Table 2.2 | Years teaching | | | | | |
| | Mean | 1.2 | 1.2 | 0.8 | 1.3 | 1.3 |
| | Median | — | — | — | — | — |
| | Range | — | — | — | — | — |
| | Years teaching science | | | | | |
| | Mean | 1.1 | 1.1 | 0.8 | 1.1 | 1.3 |
| | Median | — | — | — | — | — |
| | Range | — | — | — | — | — |
| Table 2.3 | Lessons taught by teachers who took at least one science or science education course | 3.3 | 6.0 | 5.2 | 7.4 | 5.5 |
| | Average number of professional development activities | 0.3 | 0.1 | 0.1 | 0.2 | 0.3 |
| Table 2.4 | Classroom management and organization | 5.6 | 2.6 | 4.3 | 4.9 | 4.7 |
| | Cooperative group instruction | 6.0 | 3.2 | 3.4 | 5.5 | 6.2 |
| | Interdisciplinary instruction | 4.0 | 1.7 | ‡ | 1.9 | 6.8 |
| | Science instructional techniques | 5.5 | 4.9 | 4.9 | 6.3 | 7.2 |
| | Standards-based teaching | 5.6 | — | 5.3 | 3.9 | 7.2 |
| | Teaching higher-order thinking skills | 4.6 | ‡ | ‡ | 4.0 | 7.0 |
| | Teaching students from different cultural backgrounds | 4.0 | ‡ | ‡ | 3.4 | 5.9 |
| | Teaching students with limited proficiency in their national language | 2.4 | ‡ | ‡ | 2.7! | 4.8 |
| | Teaching students with special needs | 5.5 | 2.8 | 2.5 | 4.0 | 6.1 |
| | Use of technology | 5.9 | 4.9 | 5.5 | 5.4 | 5.2 |
| | Other professional development activities | 6.3 | 4.5 | 3.6 | 5.7 | 7.2 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Table 2.5 | All teaching and other school-related activities - Total | 1.4 | 1.1 | 1.5 | 1.2 | 3.0 |
| | Teaching science classes | 0.6 | 0.6 | 0.3 | 0.8 | 1.8 |
| | Teaching other classes | 0.4 | 0.5 | 0.2 | 0.8 | 1.0 |
| | Meeting with other teachers to work on curriculum and planning issues | 0.1 | 0.2 | 0.2 | 0.1 | 0.3 |
| | Work at school related to teaching science | 0.6 | 0.4 | 0.5 | 0.5 | 0.6 |
| | Work at home related to teaching science | 0.5 | 0.4 | 0.4 | 0.7 | 0.8 |
| | Other school-related activities | 0.9 | 1.0 | 1.2 | 0.5 | 1.1 |
| Table 2.6 | Performance expectations for science | | | | | |
| | Knowing and understanding science | | | | | |
| | Knowing science information | 5.6 | 5.3 | 4.5 | 5.0 | 5.5 |
| | Understanding scientific ideas | 6.3 | 2.6 | 5.5 | 5.9 | 5.3 |
| | Understanding the nature of science | 2.3! | ‡ | ‡ | ‡ | 2.1! |
| | Doing science | | | | | |
| | Carrying out a scientific experiment, project, or activity | 2.1! | 2.0 | 3.3 | 4.1 | 5.5 |
| | Developing generic thinking skills | ‡ | ‡ | 1.8! | 3.2 | 2.6! |
| | Learning laboratory skills | 3.7 | 3.0 | 3.1 | 4.2 | 2.7 |
| | Using scientific inquiry skills | 4.1 | 2.7 | 2.5 | 3.9 | 6.6 |
| | Context of science | | | | | |
| | Awareness of the usefulness of science in life | 5.0 | 3.7 | 3.2 | 5.3 | 5.1 |
| | Collaborative work in groups | ‡ | ‡ | ‡ | 3.9 | 3.7 |
| | Independent work | 2.0 | ‡ | 1.7! | 3.9 | 3.1 |
| Table 2.7 | Cooperative work with other teachers | 5.8 | 2.6 | 2.7! | 7.2 | 5.4 |
| | Curriculum guidelines | 6.2 | 2.5 | 4.4 | 6.7 | 5.2 |
| | External examinations or standardized tests | — | 1.9! | 2.3! | 2.6 | 4.4 |
| | Mandated textbook | 5.6 | 5.6 | 6.3 | 6.7 | 6.2 |
| | Teacher's comfort with or interest in the topic | 5.0 | 5.1 | 3.1 | 6.7 | 7.2 |
| | Teacher's assessment of students' interests or needs | 6.7 | 5.8 | 5.8 | 6.7 | 5.9 |
| Figure 2.2 | Agree | 4.8 | 4.8 | 3.5 | 4.2 | 5.9 |
| | No opinion | 4.1 | 4.7 | 5.2 | 3.3 | 2.8 |
| | Disagree | 2.3 | 2.6 | 5.2 | 2.8 | 5.3! |
| Text | Teacher satisfied videotaped lesson achieved goals | 3.3 | 2.9 | 6.0 | 4.7 | 2.4 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Text | Teacher not satisfied videotaped lesson achieved goals | 3.3 | 2.9 | 6.0 | 4.7 | 2.4 |
| Figure 2.3 | A fair amount or a lot | 4.4 | 4.7 | 3.5 | 5.5 | 4.9 |
| | A little | 4.2 | 5.1 | 5.5 | 6.7 | 4.9 |
| | Not at all | ‡ | 3.2 | 5.3 | 5.3 | ‡ |
| Text | All students required to take science course | 3.6 | 1.9 | 2.9 | 0.0 | 5.1 |
| Text | Students behaved better than usual | 5.8 | 4.0 | 5.0 | 4.0 | 4.4 |
| Text | Students behaved as usual | 5.7 | 5.4 | 5.5 | 4.8 | 5.6 |
| Text | Students behaved worse than usual | 2.7! | 5.4 | 2.5 | 3.9 | 4.6! |
| Text | More difficult | 1.8! | 3.1 | 3.1 | 3.6 | 3.1 |
| Text | About the same | 2.5 | 3.8 | 3.7 | 4.3 | 5.8 |
| Text | Less difficult | 1.7 | 2.2! | 2.6 | 2.4! | 5.5 |
| Figure 2.4 | Almost always | 4.9 | 5.8 | 5.0 | 4.5 | 5.8 |
| | Often | 6.9 | 5.7 | 5.0 | 5.4 | 6.4 |
| | Sometimes or seldom | 5.0 | 1.9! | 3.7 | 4.7 | 4.7 |
| Text | Lesson was better than usual | 3.7 | 4.4 | 4.4 | 3.0! | 2.5! |
| Text | Lesson was not influenced by camera | 6.3 | 4.5 | 5.2 | 4.5 | 3.1 |
| Text | Lesson was worse than usual | 5.3 | 5.0 | 3.7 | 3.6 | 2.3 |
| Figure 2.5 | Videotaped lesson | 4.0 | 6.3 | 13.5 | 3.2 | 5.8 |
| | Similar lessons | 1.9 | 1.9 | 11.5 | 1.8 | 6.1 |
| Text | Videotaped lesson was part of a sequence of lessons | 2.6 | 1.4 | 0.8 | 1.3 | 1.5 |
| Table 2.8 | Average number of lessons in unit | 1.0 | 0.7 | 0.8 | 2.5 | 0.8 |
| | Average placement of the videotaped lesson in unit | 0.8 | 0.4 | 0.7 | 1.6 | 0.5 |
| Table 3.1 | Mean | 1.8 | 0.2 | 0.4 | 0.9 | 1.6 |
| | Median | — | — | — | — | — |
| | Range | — | — | — | — | — |
| | Standard deviation | 1.1 | 0.2 | 0.3 | 1.8 | 2.2 |
| Table 3.2 | Mean | 1.6 | 0.2 | 0.4 | 0.9 | 1.5 |
| | Median | — | — | — | — | — |
| | Range | — | — | — | — | — |
| | Standard deviation | 1.1 | 0.2 | 0.2 | 1.6 | 2.2 |
| Figure 3.2 | Non-science | 0.4 | 0.1 | 0.2 | 0.4 | 0.4 |
| | Science organization | 0.8 | 0.3 | 0.4 | 0.7 | 0.7 |
| | Science organization and non-science | 0.8 | 0.3 | 0.4 | 0.7 | 0.7 |
| | Science instruction | 0.8 | 0.3 | 0.4 | 0.7 | 0.7 |
| Text | Lessons with 3 or more interruptions | 6.2 | 4.7 | 4.4 | 6.6 | 4.1 |
| Figure 3.3 | Outside interruptions | 6.2 | 2.8 | ‡ | 4.7 | 6.1 |
| | Non-science segments | 6.2 | 5.4 | 5.2 | 5.2 | 6.3 |
| | Science organization segments | 4.3 | 5.9 | 4.5 | 4.2 | 3.0 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Table 3.3 | Developing new content | 2.8 | 0.9 | 0.0 | 1.1 | 2.2 |
| | Reviewing previous content | 5.2 | 3.8 | 5.5 | 3.5 | 5.3 |
| | Going over homework | 0.5 | 1.9! | 1.5! | 6.1 | 5.5 |
| | Assessing student learning | 1.4! | 5.5 | 2.3 | 3.7 | 5.6 |
| | Other purposes | 0.9 | 1.4 | 1.2 | 0.0 | 2.8 |
| Text | Time devoted to going over homework | ‡ | ‡ | 1.9 | ‡ | ‡ |
| Table 3.4 | Developing new content | 3.0 | 2.3 | 1.0 | 2.9 | 2.4 |
| | Reviewing previous content | 2.9 | 2.0 | 0.9 | 0.5! | 1.8 |
| | Going over homework | 0.0 | 0.5! | ‡ | 2.5 | 1.2 |
| | Assessing student learning | ‡ | 1.4 | 0.2 | 0.6 | 0.9 |
| | Other purposes | 0.6 | 0.2 | 0.2 | 0.7 | 1.5 |
| Figure 3.4 | Developed new content only | 5.2 | 3.8 | 5.5 | 3.5 | 5.2 |
| | Developed new content and reviewed previous content | 4.5 | 3.9 | 5.5 | 3.5 | 5.4 |
| | Reviewed previous content only | ‡ | ‡ | ‡ | ‡ | ‡ |
| Text | Lessons with practical activities | 3.2 | 4.5 | 4.1 | 5.4 | 4.3 |
| Text | Lessons with seatwork activities | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| Figure 3.5 | Practical activities | 3.2 | 4.5 | 4.1 | 5.4 | 4.3 |
| | Seatwork activities | 2.7 | 1.7 | 3.5 | 3.9 | 3.6 |
| Text | Lessons with independent work | 0.1 | 3.0 | 1.1 | 4.1 | 2.4 |
| Text | Lessons with whole-class work | 0.5 | 0.0 | 0.0 | 1.6 | 1.5 |
| Text | Lessons with divided class work | ‡ | 4.2 | ‡ | 3.3 | 1.9 |
| Figure 3.6 | Whole-class work | 2.7 | 1.4 | 2.8 | 3.9 | 4.2 |
| | Independent work | 2.3 | 1.5 | 2.8 | 4.0 | 4.1 |
| | Divided class work | ‡ | 0.6 | ‡ | 1.5 | 0.5! |
| Table 3.5 | Whole-class practical activities | 4.6 | 4.4 | 4.9 | 6.6 | 5.3 |
| | Whole-class seatwork activities | 0.5 | 0.0 | 0.0 | 1.6 | 1.5 |
| | Independent practical activities | 6.1 | 4.6 | 5.5 | 5.8 | 6.4 |
| | Independent seatwork activities | 3.3 | 3.2 | 4.2 | 5.0 | 4.3 |
| Figure 3.7 | Whole-class practical activities | 1.3 | 1.1 | 1.0 | 1.4 | 0.7 |
| | Whole-class seatwork activities | 1.9 | 1.8 | 2.7 | 3.5 | 3.8 |
| | Independent practical activities | 3.0 | 1.2 | 3.4 | 3.9 | 3.8 |
| | Independent seatwork activities | 2.2 | 1.3 | 1.7 | 3.2 | 2.9 |
| Figure 4.1 | Earth science | 2.1 | ‡ | 2.8 | ‡ | 5.3 |
| | Life science | 4.5 | 4.3 | 3.4 | 5.4 | 4.5 |
| | Physics | 5.8 | 4.4 | 4.6 | 5.4 | 5.5 |
| | Chemistry | 4.3 | 4.4 | 4.6 | 3.4 | 4.1 |
| | Other areas | 2.9 | 3.1 | ‡ | 4.0 | 5.4 |
| Figure 4.2 | Lessons that addressed canonical knowledge during public talk | 2.2 | 0.0 | 1.0 | 4.6 | 5.5 |
| Figure 4.3 | Public talk time devoted to canonical knowledge | 2.3 | 1.9 | 2.7 | 3.2 | 2.8 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Figure 4.4 | Lessons that incorporated real-life issues during public talk | 4.3 | 3.0 | 5.8 | 5.7 | 5.2 |
| Figure 4.5 | Public talk time devoted to real-life issues | 4.3 | 3.0 | 5.8 | 5.7 | 5.2 |
| Figure 4.6 | Lessons that addressed procedural and experimental knowledge during public talk | 3.9 | 5.1 | 2.6 | 5.9 | 5.8 |
| Figure 4.7 | Public talk time devoted to procedural and experimental knowledge | 1.5 | 1.5 | 1.9 | 2.1 | 2.0 |
| Figure 4.8 | Lessons that included classroom safety knowledge during public talk | 5.4 | 4.3 | 5.6 | 3.9 | 5.9 |
| Text | Public talk time devoted to classroom safety knowledge | 0.2 | 0.2 | 0.5 | 0.2! | 0.5! |
| Text | Lessons that addressed nature of science knowledge during public talk | 1.9 | ‡ | 2.7 | 1.6! | 2.9 |
| Text | Public talk time devoted to nature of science knowledge | 0.0 | ‡ | 0.1! | 0.0 | 0.2! |
| Text | Lessons that addressed meta-cognitive knowledge during public talk | 5.1 | 4.5 | 3.8 | 4.3 | 5.6 |
| Text | Public talk time devoted to meta-cognitive knowledge | 0.3 | 0.1 | 0.1! | 0.2 | 0.2 |
| Figure 5.1 | Teacher | 5.9 | 5.4 | 4.3 | 3.7 | 3.8 |
| | Textbook/workbook | 4.3 | 5.0 | 4.6 | 5.8 | 7.1 |
| | Worksheet | 5.9 | ‡ | 5.2 | 4.8 | 5.0 |
| | Other source | 4.3! | ‡ | 1.8! | ‡ | 3.0 |
| Figure 5.2 | Doing activities without the opportunity to learn science content | 3.2 | ‡ | 2.3 | 3.1 | 5.8 |
| | Learning science content | 3.2 | 0.0 | 2.3 | 3.1 | 5.8 |
| Figure 5.3 | High number of public canonical ideas | 4.7 | 5.2 | 2.8 | 4.8 | 6.0 |
| Figure 5.4 | Science terms | 1.3 | 3.2 | 1.2 | 1.6 | 2.4 |
| | Highly technical science terms | 0.9 | 2.3 | 0.8 | 0.7 | 1.6 |
| Figure 5.5 | Making connections | 6.2 | 4.6 | 4.7 | 5.3 | 6.5 |
| | Acquiring facts, definitions, and algorithms | 6.2 | 4.6 | 4.7 | 5.3 | 6.5 |
| Figure 5.6 | Making connections through inquiries | 6.1 | 3.6 | 4.9 | 4.2 | 5.7 |
| | Making connections through applications | 3.0 | 3.6 | 4.1 | 3.2 | 3.8 |
| | Making connections through unidentified approaches | ‡ | ‡ | ‡ | 3.0! | ‡ |

See notes at end of table.

**TABLE C.1.** **Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Figure 5.7 | Doing activities with no conceptual links | 3.2 | ‡ | 2.3 | 3.1 | 5.8 |
| | Learning content with weak or no conceptual links | 5.7 | 5.8 | 5.0 | 5.6 | 7.0 |
| | Learning content with strong conceptual links | 6.1 | 5.8 | 5.3 | 5.6 | 5.4 |
| Figure 5.8 | Goal statements | 2.2 | 2.7 | 4.5 | 5.0 | 6.1 |
| | Summary statements | 5.6 | 4.4 | 4.9 | 3.0! | 3.8 |
| Figure 5.9 | Goal statement includes main idea presented as a research question | 6.2 | 4.9 | 4.9 | 6.0 | 4.8 |
| | Goal statement includes main idea presented as a known outcome | 6.4! | ‡ | ‡ | 5.9! | ‡ |
| | Goal statement includes topic only | 5.0 | 5.0 | 3.1 | 5.5 | 5.4 |
| | Goal statement includes only activity or page number | 5.1 | ‡ | 2.4 | 5.7 | 3.2 |
| Figure 5.10 | Both goal and summary statements of any type | 5.5 | 4.5 | 5.3 | 3.0! | 3.2 |
| | Both goal and summary statements include more than naming a topic | 4.9 | 3.2 | 5.0 | ‡ | ‡ |
| Figure 5.11 | Challenging content | 4.6! | 4.3 | 2.6 | 5.0 | 5.8 |
| | Basic and challenging content | 5.1 | 5.5 | 4.3 | 6.6 | 5.6 |
| | Basic content | 5.7 | 4.2 | 4.8 | 6.8 | 6.4 |
| Figure 5.12 | Lessons that publicly presented scientific laws and theories | 5.8 | 4.6 | 3.4 | 4.7 | 5.9 |
| Figure 6.1 | First-hand data | 3.9 | 5.6 | 3.5 | 6.1 | 6.6 |
| | Phenomena | 1.9 | 5.7 | 5.2 | 5.4 | 6.2 |
| | Visual representations | 4.8 | 2.5 | 2.1 | 5.0 | 6.6 |
| Text | Lessons that used at least 2 types of visual representations | 6.6 | 4.5 | 5.3 | 6.1 | 6.3 |
| Text | Lessons that used at least 3 types of visual representations | 3.1 | 4.9 | 4.1 | 2.7 | 3.2 |
| Figure 6.2 | More than one set of first-hand data | 5.9 | 4.7 | 4.7 | 5.1 | 5.5 |
| | More than one phenomenon | 5.5 | 4.5 | 5.0 | 4.3 | 4.7 |
| | More than one visual representation | 6.6 | 5.6 | 5.3 | 5.6 | 6.7 |
| Figure 6.3 | Lessons that supported all main ideas with first-hand data, phenomena, and visual representations | 5.1 | 4.9 | 5.4 | 4.0 | 4.7 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Table 7.2 | Created models | 1.8! | ‡ | ‡ | ‡ | 3.2 |
| | Displayed or classified objects | 2.7! | ‡ | ‡ | ‡ | 4.7! |
| | Used tools, procedures, and science processes | 3.0 | 1.9! | ‡ | 2.4 | 1.6 |
| | Conducted an experiment | 3.7 | ‡ | ‡ | 2.7! | ‡ |
| | Produced or observed phenomena | 5.4 | 4.0 | 5.4 | 4.8 | 5.3 |
| Figure 7.1 | No set-up talk | ‡ | ‡ | ‡ | 3.1! | 2.1! |
| | Set-up talk about procedures | 6.1 | 2.8 | 5.2 | 4.4 | 4.2 |
| | Set-up talk about procedures and ideas | 5.6 | 4.0 | 5.5 | 2.6 | 5.7 |
| Figure 7.2 | Verified knowledge | 4.1 | 3.6 | ‡ | 2.1! | 5.6 |
| | Followed procedures | 4.9 | 1.8! | 3.8 | 3.1 | 4.3 |
| | Explored a question | 5.0 | 2.8 | 5.7 | 4.4 | 2.7 |
| Figure 7.3 | Main conclusion was discussed | 5.0 | 3.5 | 5.0 | ‡ | ‡ |
| | Several conclusions were discussed | 5.2 | 2.8 | 3.1 | 2.1! | 5.8 |
| | Observations and data were discussed | 4.3 | ‡ | 3.5 | ‡ | 2.4 |
| | Outcomes were not discussed | 5.2 | 2.0! | 3.5 | 5.4 | 5.2 |
| Figure 7.4 | Methods critiqued or evaluated | 4.6 | 2.2! | 3.9 | 2.0 | 2.1 |
| | New questions to be investigated discussed | 3.3 | ‡ | 4.8 | ‡ | ‡ |
| Table 7.3 | Generated the research question | 2.0! | ‡ | ‡ | ‡ | ‡ |
| | Designed procedures for investigation | 3.7 | ‡ | 2.4! | ‡ | 1.9 |
| | Made predictions | 3.5 | ‡ | 5.0 | 2.5! | 3.1 |
| | Interpreted the data or phenomena | 6.1 | 4.5 | 5.2 | 5.1 | 6.3 |
| | Collected and recorded data | 5.5 | 3.2 | 5.7 | 5.6 | 5.1 |
| | Organized or manipulated data collected independently | 3.2 | ‡ | ‡ | 3.4 | 3.4 |
| | Organized or manipulated collected data guided by teacher or textbook | 5.0 | 1.9! | 5.1 | 3.3 | 4.5 |
| Figure 7.5 | Students made predictions | ‡ | 2.9 | 3.4 | ‡ | ‡ |
| | Students interpreted data or phenomena | 4.8 | 5.6 | 3.5 | 4.1 | 2.6 |
| Text | Students give reasons for predictions | 2.4 | ‡ | 2.6 | ‡ | ‡ |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Figure 8.1 | Individual work | 4.8 | 5.0 | 4.5 | 5.7 | 4.8 |
| | Pair/group work | 6.1 | 5.2 | 5.0 | 6.1 | 6.0 |
| Text | Changing social participation structure | 1.6! | 1.1! | 1.9! | 2.1! | 1.9! |
| Figure 8.2 | Individual work | 2.5 | 1.0 | 1.2 | 3.0 | 2.5 |
| | Pair/group work | 3.0 | 1.5 | 3.3 | 4.0 | 3.1 |
| Figure 8.3 | Individual work during independent practical activities | 1.5! | ‡ | ‡ | 0.4! | ‡ |
| | Pair/group work during independent practical activities | 3.0 | 1.2 | 3.4 | 3.7 | 3.6 |
| | Individual work during independent seatwork activities | 2.1 | 1.0 | 1.1 | 3.1 | 2.2 |
| | Pair/group work during independent seatwork activities | 1.2 | 0.7 | 1.2 | 1.8! | 2.6 |
| Table 8.1 | Total | 3.0 | 1.5 | 3.3 | 4.0 | 3.1 |
| | Sitting together | 2.9 | 1.5 | 3.4 | 3.7 | 3.4 |
| | Sharing materials | 2.9 | 1.4 | 3.4 | 3.8 | 3.1 |
| | Talking among students | 3.0 | 1.5 | 3.3 | 4.0 | 3.1 |
| | Working on tasks requiring collaboration | 1.3 | ‡ | 1.0 | ‡ | 2.0 |
| | Assigning roles to group members | 1.4 | ‡ | 0.7 | ‡ | 2.5 |
| | Creating science group products | 2.0 | 0.8 | 2.4 | 2.5! | 3.0 |
| | Working in all mixed gender groups | 0.9! | ‡ | 2.1 | ‡ | 1.5 |
| Figure 9.1 | Public discussions | 1.2 | 1.5 | 1.0 | 2.0 | 2.3 |
| | Public presentations | 2.5 | 1.4 | 2.1 | 2.3 | 3.3 |
| Figure 9.2 | Private teacher-student talk | 2.0 | 0.6 | 1.7 | 2.9 | 2.6 |
| | Private student-peer talk | 2.4 | 0.7 | 1.1 | 3.0 | 2.1 |
| Figure 9.3 | Other words | 1.2! | 0.4 | 0.1 | 1.3! | 0.7 |
| | Student words | 0.6 | 0.5 | 1.0 | 1.4 | 1.1 |
| | Teacher words | 1.0 | 0.6 | 0.9 | 1.7 | 1.2 |
| Figure 9.4 | 5+ word student utterances during public talk | 1.3 | 1.1 | 1.6 | 1.6 | 1.3 |
| | 5+ word student utterances during private teacher-student talk | 1.7 | 3.4 | 1.5 | 2.0 | 1.9 |
| Figure 9.5 | Took notes during whole-class work | 0.7 | 0.6 | 0.6 | 0.4 | 0.4 |
| | Selected answers during independent work | 2.3 | 1.2 | 2.5 | 2.2 | 2.2 |
| | Generated written responses during independent work | 3.1 | 1.2 | 3.2 | 4.3 | 3.3 |
| Text | Time for students to write about science | 5.9 | 11.0 | 72 | 9.3 | 11.0 |
| Text | Lessons in which students were expected to write at least a paragraph | 3.5 | ‡ | ‡ | 2.4 | 4.9 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Text | Lessons in which students generated written responses | 6.0 | 5.0 | 4.9 | 5.7 | 5.3 |
| Text | Lessons in which students independently selected answers | 5.7 | 5.1 | 5.3 | 4.8 | 6.4 |
| Text | Lessons in which students took notes | 6.2 | 5.8 | 5.8 | 4.1 | 3.7 |
| Figure 9.6 | Graphs | 1.5! | ‡ | 3.3 | 3.7 | 3.6 |
| | Diagrams | 4.6 | 2.7 | 3.2 | 4.9 | 4.5 |
| | Mathematical calculations | 3.5 | 4.5 | 3.7 | 5.7 | 5.8 |
| Figure 9.7 | Reading aloud together | 0.1! | 0.2! | 0.1 | 0.4! | 0.2! |
| | Reading silently | 1.7 | 0.1! | 0.5! | 3.7 | 2.6 |
| Text | Lessons with silent reading tasks | 5.1 | 1.8! | 2.8 | 6.1 | 4.9 |
| Figure 9.8 | Talk about science | 2.2 | 1.6 | 2.6 | 2.7 | 3.2 |
| | Write about science | 2.6 | 1.6 | 2.9 | 4.0 | 3.8 |
| | Read about science | 1.7 | 0.2! | 0.5! | 3.6 | 2.6 |
| Text | Time during seatwork activities for students to talk about science | 2.6 | 1.6 | 1.8 | 3.2 | 3.3 |
| Text | Time during seatwork activities for students to write about science | 2.2 | 1.4 | 1.5 | 3.1 | 2.8 |
| Text | Time during seatwork activities for students to read about science | 1.5 | 0.2! | 0.2 | 3.2 | 2.2 |
| Figure 10.1 | Lessons in which at least one real-life issue was raised | 4.1 | 3.0 | 5.7 | 5.4 | 4.3 |
| Figure 10.2 | Time during which real-life issues were raised | 2.2 | 1.3 | 2.1 | 3.4 | 4.2 |
| Figure 10.3 | At least one real-life issue used to develop science ideas | 5.0 | 3.7 | 5.9 | 6.3 | 5.8 |
| | At least one real-life issue mentioned as topic-related sidebar | 6.3 | 4.7 | 6.0 | 6.3 | 6.0 |
| Figure 10.4 | Real-life issues mentioned as topic-related sidebars | 1.1 | 1.1 | 0.8 | 1.3 | 0.9 |
| | Real-life issues used to develop science ideas | 1.7 | 0.8 | 1.8 | 3.1 | 4.2 |
| Figure 10.5 | Lessons that had at least one motivating activity | 6.0 | 4.1 | 4.6 | 5.7 | 5.5 |
| Figure 10.6 | Time allocated to motivating activities | 2.7 | 0.8 | 1.6 | 1.8 | 4.7 |
| Figure 10.7 | Three types of activities | 4.1 | 2.6 | 2.8 | 3.2! | 5.1 |
| | Two types of activities | 5.6 | 4.6 | 5.2 | 6.3 | 5.8 |
| | One type of activity | 4.9 | 4.6 | 5.7 | 6.7 | 5.3 |
| Text | Lessons with routine lesson openers | 1.5! | 0.0 | 2.2 | 0.0 | 5.8 |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Figure 11.1 | Lessons in which students created organized science notebooks | 4.0 | 2.0 | 5.9 | 5.7 | 5.3 |
| Figure 11.2 | Lessons in which students used textbooks and/or workbooks | 5.3 | 5.5 | 4.8 | 3.8 | 5.8 |
| Figure 11.3 | Computers available in the classroom | 3.9 | 2.3! | 1.9! | 5.3 | 6.6 |
| | Computers used by students | ‡ | ‡ | 1.6! | ‡ | 4.7! |
| Figure 11.4 | Public grading | 2.3 | 5.1 | ‡ | 3.7 | 2.3! |
| | Public assessment | ‡ | 4.2 | ‡ | ‡ | ‡ |
| | Public work | 2.9 | 4.4 | 4.4 | 2.4 | 3.9 |
| Text | Lessons in which students made presentations | 2.3 | 3.6 | 2.2! | 3.3! | 3.0 |
| Figure 11.5 | Lessons that included at least one student-initiated science question | 5.9 | 3.6 | 4.3 | 5.8 | 7.2 |
| Figure 11.6 | Student-initiated science questions per eighth-grade science lesson | 33.9 | 10.8 | 26.9 | 109.2 | 55.5 |
| Text | Lessons in which students generated own research questions | 2.0! | 0.0 | 0.0 | 1.6! | 0.0 |
| Text | Lessons in which students designed procedures for investigation | 3.7 | 1.1! | 2.4! | 1.6! | 1.9 |
| Text | Lessons in which students collected data | 4.5 | 5.4 | 4.6 | 5.8 | 6.6 |
| Figure 11.7 | Lessons in which the teacher assigned homework for future lessons | 5.9 | 5.4 | 2.9 | 6.3 | 6.6 |
| Figure 11.8 | Work on new content only | 5.7 | 5.2 | 3.5 | 6.0 | 6.1 |
| | Mixed | ‡ | 2.4! | ‡ | 5.7 | ‡ |
| | Review previously covered content only | ‡ | 3.9 | ‡ | ‡ | 1.2 |
| Figure 11.9 | Reviewing homework | 0.5 | 1.9! | ‡ | 6.1 | 5.5 |
| | Working on homework assignments in class | 6.3 | 2.7 | 3.1 | 5.6 | 5.7 |
| | Reviewing homework and working on homework assignments in class | ‡ | ‡ | ‡ | 5.2 | 2.8 |
| Figure 11.10 | Students worked at their own pace on long-term assignments | 5.0 | ‡ | ‡ | 5.5 | 4.9 |
| Text | Students expected to check their own work | 0.0 | 0.0 | 1.1! | 6.2 | 1.1! |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Tables E.1-E.5 | Earth science | | | | | |
| | Building and breaking of earth's surface | ‡ | ‡ | ‡ | ‡ | 1.3! |
| | Planets in the solar system | ‡ | ‡ | ‡ | ‡ | 2.4! |
| | Rocks and soil | ‡ | ‡ | ‡ | ‡ | 2.9 |
| | Weather and climate | ‡ | ‡ | 2.8 | ‡ | 2.3 |
| | Life science | | | | | |
| | Animals | ‡ | ‡ | 1.8! | ‡ | ‡ |
| | Evolution, speciation, and diversity | ‡ | ‡ | ‡ | ‡ | 2.0! |
| | Disease | ‡ | ‡ | ‡ | ‡ | 3.1! |
| | Organs and tissues | 2.6! | 4.7 | 3.0 | 4.5 | ‡ |
| | Plants and fungi | 2.4 | ‡ | ‡ | ‡ | ‡ |
| | Reproduction | 1.8! | ‡ | ‡ | 1.9! | ‡ |
| | Sensing and responding | ‡ | ‡ | 1.7! | 2.0! | ‡ |
| | Variation and inheritance | ‡ | ‡ | ‡ | ‡ | 1.7! |
| | Physics | | | | | |
| | Electricity | 4.2 | 3.0 | 4.5 | 1.8! | 1.8! |
| | Energy types, sources, and conversions | 4.0 | ‡ | ‡ | ‡ | ‡ |
| | Fluid behavior | 1.7! | 1.8! | ‡ | ‡ | ‡ |
| | Heat and temperature | ‡ | 2.0! | ‡ | 3.1 | ‡ |
| | Light | 2.9 | ‡ | ‡ | 4.0 | ‡ |
| | Magnetism | ‡ | ‡ | 1.9! | ‡ | ‡ |
| | Physical changes | ‡ | 2.6 | ‡ | ‡ | ‡ |
| | Physical properties | ‡ | ‡ | ‡ | ‡ | 1.6! |
| | Sound and vibration | ‡ | ‡ | ‡ | 4.7 | ‡ |
| | Types of forces | 3.3 | ‡ | ‡ | 2.3! | ‡ |
| | Chemistry | | | | | |
| | Atoms, ions, and molecules | ‡ | 1.7! | ‡ | ‡ | 3.2! |
| | Chemical changes | 1.6! | 2.2! | 4.5 | 2.7! | ‡ |
| | Chemical properties | 3.3 | 2.3! | ‡ | 2.2! | 1.7! |
| | Classification of matter | ‡ | 3.5 | ‡ | ‡ | 2.3! |
| | Other areas | | | | | |
| | Interactions of science, technology, and society | 1.9! | 2.3 | ‡ | ‡ | 2.1! |
| | Nature of scientific knowledge | ‡ | ‡ | ‡ | ‡ | 3.4! |
| | Science and mathematics | ‡ | 2.1! | ‡ | 1.5! | ‡ |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Table E.6 | Earth science | | | | | |
| | Making connections | 1.5! | ‡ | 2.3 | ‡ | 1.9 |
| | Acquiring facts, definitions, and algorithms | ‡ | ‡ | ‡ | ‡ | 5.1 |
| | Life science | | | | | |
| | Making connections | 3.9 | ‡ | 2.3 | ‡ | ‡ |
| | Acquiring facts, definitions, and algorithms | 3.8 | 4.2 | 2.8 | 5.4 | 4.4 |
| | Physics | | | | | |
| | Making connections | 5.7 | 3.9 | 4.8 | 3.8 | 5.1! |
| | Acquiring facts, definitions, and algorithms | 5.2 | 3.9 | 2.2 | 5.4 | 2.7 |
| | Chemistry | | | | | |
| | Making connections | 4.2 | 3.3 | 4.9 | 2.4! | 3.1 |
| | Acquiring facts, definitions, and algorithms | 1.8 | 3.4 | 3.3 | 2.4! | 3.2 |
| | Other areas | | | | | |
| | Making connections | 1.7! | ‡ | ‡ | 3.0! | 4.2 |
| | Acquiring facts, definitions, and algorithms | 2.4 | 3.1 | ‡ | 2.7 | 4.8 |
| Figure E.1 | Focus on algorithms and techniques | 3.6 | 4.4 | 3.8 | 4.5 | 5.9 |
| | Focus on sequences of events | ‡ | 3.1 | ‡ | ‡ | ‡ |
| | Focus on discrete bits of information | 5.7 | 4.8 | 3.3 | 6.1 | 6.2 |
| | Focus on unidentified approaches | ‡ | ‡ | ‡ | ‡ | ‡ |
| Table E.7 | Earth science | | | | | |
| | Challenging or a mix of basic and challenging | ‡ | ‡ | ‡ | ‡ | 3.8 |
| | Basic | 1.9! | ‡ | 2.6 | ‡ | 4.1 |
| | Life Science | | | | | |
| | Challenging or a mix of basic and challenging | 3.2 | 4.3 | 3.3 | 5.5 | 3.7 |
| | Basic | 3.9 | 2.8 | 3.0 | 2.0 | 3.5! |
| | Physics | | | | | |
| | Challenging or a mix of basic and challenging | 6.1 | 4.2 | 3.5 | 4.7 | 5.5 |
| | Basic | 4.4 | 2.1 | 4.5 | 5.5 | 2.5 |
| | Chemistry | | | | | |
| | Challenging or a mix of basic and challenging | 2.3! | 3.9 | 2.9 | ‡ | 3.6 |
| | Basic | 3.6 | 2.4 | 4.3 | 2.8 | ‡ |

See notes at end of table.

**TABLE C.1. Standard errors for estimates shown in figures and tables, by country[1]—Continued**

| Table/Figure | Category | AUS | CZE | JPN | NLD | USA |
|---|---|---|---|---|---|---|
| Figure E.2 | 3-dimensional models | 4.8 | 4.3 | 2.2 | 4.6 | 2.6 |
| | Graphic organizers | 5.9 | 5.9 | 5.2 | 6.5 | 6.6 |
| | Diagrams | 6.9 | 4.3 | 3.4 | 6.4 | 6.4 |
| | Formulas | 2.2 | 4.9 | 4.5 | 3.7 | 4.0 |
| | Other visual representations | 3.9 | 4.4 | 2.8 | 3.1 | 3.7 |

!Interpret data with caution. Estimate is unstable.
‡Reporting standards not met. Too few cases to be reported.
—Not available.
[1]AUS=Australia; CZE=Czech Republic; JPN=Japan; NLD=Netherlands; and USA=United States.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

# APPENDIX D
## Definitions of Constructs and Variables Used in Analyses

The definitions of constructs and variables, and, in some cases, the methods by which they are reported, are organized by the chapter in which they first appear. Only those constructs and variables which are not already defined in the chapters are included here.

## Chapter 2

### Educational Preparation

Teachers were asked about their training in and preparation for teaching science. Since comparisons of degree type are difficult to compare cross-nationally due to differences in the education systems of the participating countries, definitions developed by the Organization for Economic Cooperation and Development (OECD 1997) specifically for international comparative purposes were used to help categorize teachers' educational backgrounds in the questionnaires into three educational levels. This categorization schema is known as the International Standard Classification of Education (ISCED). Teachers' reports of their educational attainment were assigned to three categories. The first level includes upper secondary and post-secondary educational programs that require a minimum of high school completion, matriculation, or a vocational certificate (ISCED levels 3 and 4). Teachers' indications that they hold a degree from a secondary vocational school or a high school diploma fall into this category. This level also includes some college attendance, such as attaining an associate degree, without attaining a bachelor's degree. The second level incorporates postsecondary programs that last at least 3 years and prepare students for entry to graduate programs (ISCED level 5A). The third and highest level covers programs that result in an advanced research qualification or degree and include submission of original research such as a thesis or dissertation (ISCED level 6). This category includes such programs as master's degree and doctoral degree. Any of these three education levels could incorporate teacher training.

### Certification to Teach Science

Preparation for teaching eighth-grade science includes certification to teach science as well as having a science content background. To describe teachers' certification backgrounds, teachers were asked to list the subject areas and corresponding grade levels in which they were certified to teach. Each response was divided into two mutually exclusive groups: (1) teacher's certification in science included eighth-grade or (2) teacher's certification in science was not identified for grade level, did not include eighth-grade, or certification was identified in another subject area.

### Years Teaching in General and Teaching Science

Teachers were asked to identify how many years they had been teaching in general, and also how many years they had been teaching science (not limited to grade 8).

### Professional Development Opportunities

Teachers were asked to describe the science courses and professional development activities they participated in during the two years up to and including the day of videotaping.

### Teachers' Learning Goals for Science Lessons

Teachers were asked to identify, in their own words, their main goals for students in the eighth-grade science lessons that were subsequently videotaped. Teachers' responses subsequently were coded to a set of goals based on common themes across the teachers' responses. Teachers could identify more than one goal for a lesson; all goals identified by a teacher were coded. Goals were grouped according to three major expectations: knowing and understanding science (four goals), doing science (six goals), and developing students' attitudes toward science and participating in science (five goals).

### Typicality of Planning for the Lesson

Once they agreed to be videotaped, the eighth-grade science teachers could have spent more effort in planning for the videotaped lesson than they normally would have spent for a typical science lesson, although they were asked specifically to do nothing special. To get an indication of the degree to which the videotaped teachers may have put more effort into their lesson planning specifically for the purpose of the study, the teachers were asked to report how many minutes they spent planning for the videotaped lesson and how many minutes they typically spent planning for similar science lessons.

## Chapter 3

### Lesson Interruptions

Comparing countries on the occasions when science instruction was interrupted is a way of assessing whether lessons maintained a continuous focus on science instruction or provided breaks from the science instruction focus. Interruptions by an outside source, by non-science segments, or by science organization segments were examined. Examples of outside interruptions are announcements over the intercom, telephone calls that require the teacher's attention, fire drills, and visitors from outside the classroom who require the teacher's attention. Non-science segments could occur at the start or end of the lesson without interrupting or providing a break from science instruction, but occurrences of three or more of these events would most likely involve a mid-lesson interruption to science instruction. Time spent on science organization is sometimes needed to move the flow of the science lesson from one activity to another. These organizational periods do not necessarily disrupt the lesson as would outside interruptions or non-science interruptions, but they are more likely to interrupt the lesson flow if they occur multiple times. These three indicators of lesson interruptions are analyzed together in order to describe and compare how countries organize their lessons with a minimum of interruptions.

### Whole-Class and Independent Work

Science activities, both practical and non-practical, were observed to take place as a whole class or as an independent student activity. Whole-class work occurs when science instruction and related information are provided to or worked on together by the entire class. In whole-class work, all students are expected to pay attention to the same activity that is led by the teacher, a student, a small group of students, or another source (e.g., videotape, assistant teacher). Independent activities involve students working on their own, either individually or in small groups. At times, science lessons are conducted with part of the class working together under the direction of the teacher and part of the class working independently. For example, the teacher may assign half the class to work on answering questions individually, while she showed the rest of the class a demonstration. In this case, some students worked independently while the other students worked together under the direct supervision of the teacher.

## Chapter 5

### Science Terms

Science terms were identified by a team of six scientists who reviewed and categorized the words generated by a computer-assisted analysis. Science terms can range from terms commonly used outside the classroom (e.g., energy, force, kidney) to highly technical terms. The use of highly technical science terms is another indicator of the density of content in the eighth-grade science lessons. The list of science words was reviewed by a team of six scientists to identify highly technical science words.

### Acquiring Facts, Definitions, and Algorithms

The primary way in which facts, definitions, and algorithms were used to develop science content was identified for each lesson using the following definitions:

- ***Algorithms and techniques:*** Science information is presented primarily through problem solving or procedural tasks that rely on linear reasoning. Problems are straightforward (for example, calculate the volume of a cube) rather than open-ended (for example, determine which kind of water filtration plant would be best for the local community). Teachers first show the students the problem, the steps needed to solve the problem with linear reasoning, and the answer. Students then practice applying the procedures to a similar set of problems.
- ***Sequences of events:*** Science information is presented primarily as facts describing processes or stages. For example, a teacher describes how blood travels through the body beginning with the process of blood traveling through the heart and continuing with the process of oxygen being received as blood travels through the lungs. Students also may participate by drawing or labeling diagrams that represent the process of blood flow.
- ***Discrete bits of information:*** Science information is presented as isolated and unrelated definitions, facts, processes, and/or procedures. The teacher presents the information as separate and unconnected in an "all about the topic" mode. For example, a teacher may talk about different elements on the Periodic Table, describing each element and its everyday uses without any conceptual or theoretical organization.
- ***Unidentified approaches:*** The teacher helps students acquire facts, definitions, and algorithms in a way that is not primarily defined as solving problems, describing sequences of events, or presenting discrete bits of information.

## Chapter 7

### Discussion of Results

Based on observations, four types of public discussion of the outcomes of practical activities were identified:

- ***Discussing observations and data:*** The class publicly shares, compares, and checks observations and data resulting from the practical activities, but does not discuss a possible conclusion or conceptual idea based on those outcomes (Video clip example 7.7).

- ***Discussing several conclusions:*** The class discusses multiple conclusions or ideas related to different parts of the practical activity. However, there is no attempt to identify and connect them to a single conclusion or idea that is supported by the evidence available. For example, one conclusion is drawn about what happens to food when it is mashed and mixed with saliva, and another conclusion is made about what happens to food when it is mixed with gastric juices. But these two conclusions are not linked together to create a big conclusion about the process of digestion involving chemical as well as physical breakdown of food.
- ***Discussing main conclusion:*** The class discusses how the outcomes of the practical activity are connected to and support a single or main conclusion or idea (📹 Video clip example 7.8).
- ***Not discussing outcomes:*** Nothing about the outcomes or results of the independent practical activity is discussed publicly.

## Chapter 9

### Teacher and Student Words

Computer-assisted analyses were applied to English-language transcripts (Australia and the United States) and translations (Czech Republic, Japan, and the Netherlands) of the eighth-grade science lessons. Analyses based on same-language transcripts allow for comparisons of speech across countries, thought not without potential bias. Transcribers and translators were fluent in English and the language of the countries they translated. A glossary was developed to standardize translation of special terms within each country. All translations were checked for accuracy by a second translator as well as a content expert.

It is important to note that the analyses are based on only those segments of public talk in which the teacher and student(s) could be heard. In cases where many students spoke at once or made remarks out of the range of the microphones, the transcripts noted that something was uttered, but did not include guesses about what was said. Because of instances such as this, estimates of the amount of student talk are likely to be lower than actually occurred.

## Chapter 11

### Organized Science Notebooks

In some science lessons, students were observed organizing their notes and other science work in a special science notebook. These notebooks became a record of the class activities, including notes as well as work on practical and seatwork activities. In many classes, the record was organized chronologically in a sewn notebook format, and any additional worksheets were pasted into the chronological notebook in the appropriate place. Thus, students created a chronological record, or text, of their experiences in science class. In other cases, loose-leaf, ringed binders were used, with special sections for different types of science class records. In all cases, however, students were responsible for keeping a special, organized science notebook. Loose-leaf, ringed binders used by students to keep their papers from all their different classes or binders that were not used in a systematic way were excluded from the analysis that follows.

## Research Questions, Procedures for Investigation, and Data Collection

Activities that could encourage students to take responsibility for their learning were defined as:

- ***Generating research questions:*** Students, either individually or in small groups, play a role in developing a research question that they will investigate in an independent practical activity. The students may have complete freedom, such as in the definition of a question for a science fair project, or be constrained to a particular topic area or a set of options provided by the teacher. For example, the teacher may allow students to generate a research question about what promotes mold growth, or the teacher may provide a list of five variables related to mold growth and ask students to pick one variable and generate a research question about it.
- ***Designing procedures for investigation:*** Students, either individually or in small groups, play a role in planning the procedures that will be used in an independent practical activity. The students may have complete freedom or be constrained by a set of options or materials provided by the teacher. For example, students are provided with bean seeds and related materials and told to design an investigation to explore the effect of light, different colors of light, temperature, gravity, or soil composition on plant growth.
- ***Collecting data:*** Students, either individually or in small groups, collect qualitative or quantitative data during independent or whole-class work through observation of phenomena and/or manipulation of physical objects. For example, students collect quantitative data about their pulse rates before and after exercise, or they generate qualitative descriptions resulting from the heating of different metals.

# APPENDIX E
## Additional Figures and Tables

## Content Subcategories

Tables E.1 to E.5 present estimates of the percentage of science lessons that addressed various science content subcategories. These tables provide more detailed information that that contained in figure 4.1 (see chapter 4). The science topics in the lessons were identified using the TIMSS Guidebook to Examine School Curricula (McNeely 1997), which provided a common, international frame of reference for talking about science content.

The content subcategories shown in tables E.1 to E.5 specify topics at the level typically used by the classroom teachers in describing the content of the lesson on the questionnaires (e.g., rocks and soil, organs and tissues, electricity, and chemical changes). Although multiple science topics may be included in any science lesson, only the primary science topic for each lesson was identified. The primary topic was defined as the topic that was addressed for the longest amount of science instruction time.

The following topic subcategories included too few cases to calculate reliable estimates in all five countries:

- earth science: atmosphere; beyond the solar system; bodies of water; composition; earth in the solar system; earth's history; evolution of the universe; ice forms; land forms; physical cycles;
- life science: animal behavior; biochemical processes in cells; biochemistry of genetics; biomes and ecosystems; cells; energy handling; habitats and niches; interdependence of life; life cycles; nutrition; other organisms;
- physics: dynamics of motion; explanation of physical changes; kinetic theory; quantum theory and fundamental particles; relativity theory; time, space, and motion; wave phenomena; and
- chemistry: crystals; electrochemistry; energy and chemical change; explanations of chemical changes; macromolecules, nuclear chemistry; organic and biochemical changes; rate of change and equilibriums; subatomic particles.

Tables E.6 and E.7, and figures E.1 and E.2 present estimates referred to in other parts of the report but not shown in a figure or table.

**TABLE E.1. Percentage distribution of Australian eighth-grade science lessons devoted to subordinated categories of earth science, life science, physics, chemistry, and other areas: 1999**

| Discipline and area | Percent of lessons |
|---|---|
| Life science | |
| Organs and tissues | 5! |
| Plants and fungi | 5 |
| Reproduction | 3! |
| Physics | |
| Electricity | 10 |
| Energy types, sources, and conversions | 10 |
| Fluid behavior | 3! |
| Light | 6 |
| Types of forces | 8 |
| Chemistry | |
| Chemical changes | 3! |
| Chemical properties | 8 |
| Other areas | |
| Interactions of science, technology, and society | 3! |

!Interpret data with caution. Estimate is unstable.
NOTE: Total does not sum to 100 because of data not presented for subcategories without reliable estimates. The following subcategories included too few cases to calculate reliable estimates: earth science: building and breaking of earth's surface; planets in the solar system; rocks and soil; weather and climate; life science: animals; disease; evolution, speciation, and diversity; sensing and responding; variation and inheritance; physics: heat and temperature; magnetism; physical changes; physical properties; sound and vibration; chemistry: atoms, ions, and molecules; classification of matter; other areas: nature of scientific knowledge; science and mathematics.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

**TABLE E.2. Percentage distribution of Czech eighth-grade science lessons devoted to subordinated categories of earth science, life science, physics, chemistry, and other areas: 1999**

| Discipline and area | Percent of lessons |
|---|---|
| Life science | |
| Organs and tissues | 19 |
| Physics | |
| Electricity | 8 |
| Fluid behavior | 3! |
| Heat and temperature | 4! |
| Physical changes | 6 |
| Chemistry | |
| Atoms, ions, and molecules | 3! |
| Chemical changes | 4! |
| Chemical properties | 5! |
| Classification of matter | 11 |
| Other areas | |
| Interactions of science, technology, and society | 5 |
| Science and mathematics | 4! |

!Interpret data with caution. Estimate is unstable.
NOTE: Total does not sum to 100 because of data not presented for subcategories without reliable estimates. The following subcategories included too few cases to calculate reliable estimates: earth science: building and breaking of earth's surface; planets in the solar system; rocks and soil; weather and climate; life science: animals; disease; evolution, speciation, and diversity; plants and fungi; reproduction; sensing and responding; variation and inheritance; physics: energy types, sources, and conversions; light; magnetism; physical properties; sound and vibration; types of forces; other areas: nature of scientific knowledge.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

**TABLE E.3. Percentage distribution of Japanese eighth-grade science lessons devoted to subordinated categories of earth science, life science, physics, chemistry, and other areas: 1999**

| Discipline and area | Percent of lessons |
|---|---|
| Earth science | |
| Weather and climate | 7 |
| Life science | |
| Animals | 3! |
| Organs and tissues | 13 |
| Sensing and responding | 3 |
| Physics | |
| Electricity | 28 |
| Magnetism | 4! |
| Chemistry | |
| Chemical changes | 33 |

!Interpret data with caution. Estimate is unstable.
NOTE: Total does not sum to 100 because of data not presented for subcategories without reliable estimates. The following subcategories included too few cases to calculate reliable estimates: earth science: building and breaking of earth's surface; planets in the solar system; rocks and soil; life science: disease; evolution, speciation, and diversity; plants and fungi; reproduction; variation and inheritance; physics: energy types, sources, and conversions; fluid behavior; heat and temperature; light; physical changes; physical properties; sound and vibration; types of forces; chemistry: atoms, ions, and molecules; chemical properties; classification of matter; other areas: interactions of science, technology, and society; nature of scientific knowledge; science and mathematics.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

**TABLE E.4. Percentage distribution of Dutch eighth-grade science lessons devoted to subordinated categories of earth science, life science, physics, chemistry, and other areas: 1999**

| Discipline and area | Percent of lessons |
|---|---|
| Life science | |
| Organs and tissues | 16 |
| Reproduction | 3! |
| Sensing and responding | 3! |
| Physics | |
| Electricity | 3! |
| Heat and temperature | 9 |
| Light | 10 |
| Sound and vibration | 14 |
| Types of forces | 4! |
| Chemistry | |
| Chemical changes | 5! |
| Chemical properties | 4! |
| Other areas | |
| Science and mathematics | 2! |

!Interpret data with caution. Estimate is unstable.
NOTE: Total does not sum to 100 because of data not presented for subcategories without reliable estimates. The following subcategories included too few cases to calculate reliable estimates: earth science: building and breaking of earth's surface; planets in the solar system; rocks and soil; weather and climate; life science: animals; disease; evolution, speciation, and diversity; plants and fungi; variation and inheritance; physics: energy types, sources, and conversions; fluid behavior; magnetism; physical changes; physical properties; chemistry: atoms, ions, and molecules; classification of matter; other areas: interactions of science, technology, and society; nature of scientific knowledge.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

**TABLE E.5. Percentage distribution of U.S. eighth-grade science lessons devoted to subordinated categories of earth science, life science, physics, chemistry, and other areas: 1999**

| Discipline and area | Percent of lessons |
|---|---|
| Earth science | |
| Building and breaking of earth's surface | 2! |
| Planets in the solar system | 4! |
| Rocks and soil | 7 |
| Weather and climate | 5 |
| Life science | |
| Disease | 6! |
| Evolution, speciation, and diversity | 3! |
| Variation and inheritance | 3! |
| Physics | |
| Electricity | 3! |
| Physical properties | 3! |
| Chemistry | |
| Atoms, ions, and molecules | 5! |
| Chemical properties | 2! |
| Classification of matter | 4! |
| Other areas | |
| Interactions of science, technology, and society | 4! |
| Nature of scientific knowledge | 6! |

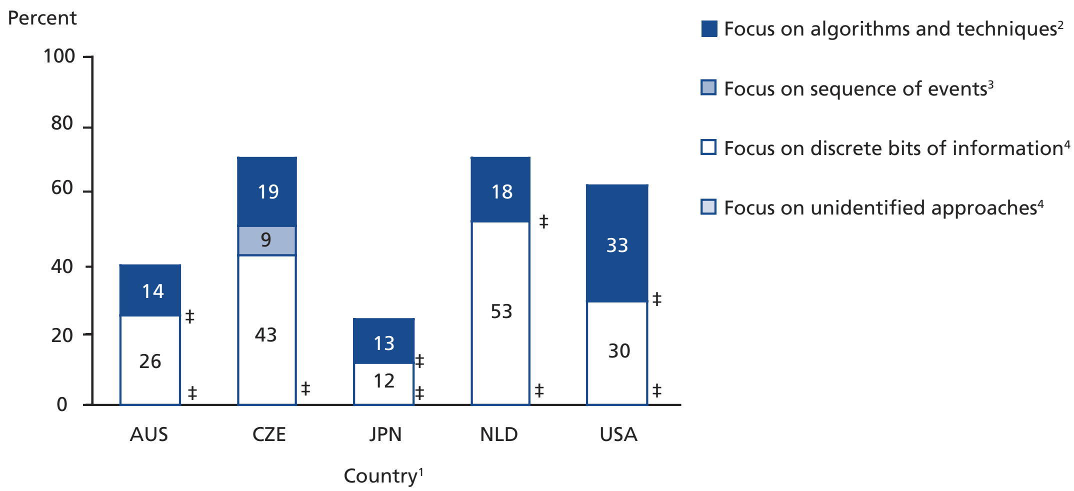
!Interpret data with caution. Estimate is unstable.
NOTE: Total does not sum to 100 because of data not presented for subcategories without reliable estimates. The following subcategories included too few cases to calculate reliable estimates: life science: animals; organs and tissues; plants and fungi; reproduction; sensing and responding; physics: energy types, sources, and conversions; fluid behavior; heat and temperature; light; magnetism; physical changes; sound and vibration; types of forces; chemistry: chemical changes; other areas: science and mathematics.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

**TABLE E.6. Percentage distribution of eighth-grade science lessons that developed content primarily by making connections and by acquiring facts, definitions, and algorithms, by content disciplines and country: 1999**

| | Country[1] | | | | |
|---|---|---|---|---|---|
| Discipline and teaching method | AUS | CZE | JPN | NLD | USA |
| Earth science | | | | | |
| Making connections[2] | 3! | ‡ | 5 | ‡ | 5 |
| Acquiring facts, definitions, and algorithms[3] | ‡ | ‡ | ‡ | ‡ | 23 |
| Life science | | | | | |
| Making connections[4] | 9 | ‡ | 10 | ‡ | ‡ |
| Acquiring facts, definitions, and algorithms[5] | 15 | 35 | 9 | 30 | 14 |
| Physics | | | | | |
| Making connections[6] | 32 | 13 | 31 | 15 | 9! |
| Acquiring facts, definitions, and algorithms[7] | 16 | 16 | 5 | 32 | 8 |
| Chemistry | | | | | |
| Making connections[8] | 11 | 13 | 27 | 5! | 7 |
| Acquiring facts, definitions, and algorithms[9] | 4 | 11 | 10 | 4! | 11 |
| Other areas | | | | | |
| Making connections[10] | 3! | ‡ | ‡ | 5! | 9 |
| Acquiring facts, definitions, and algorithms[11] | 5 | 9 | ‡ | 6 | 11 |

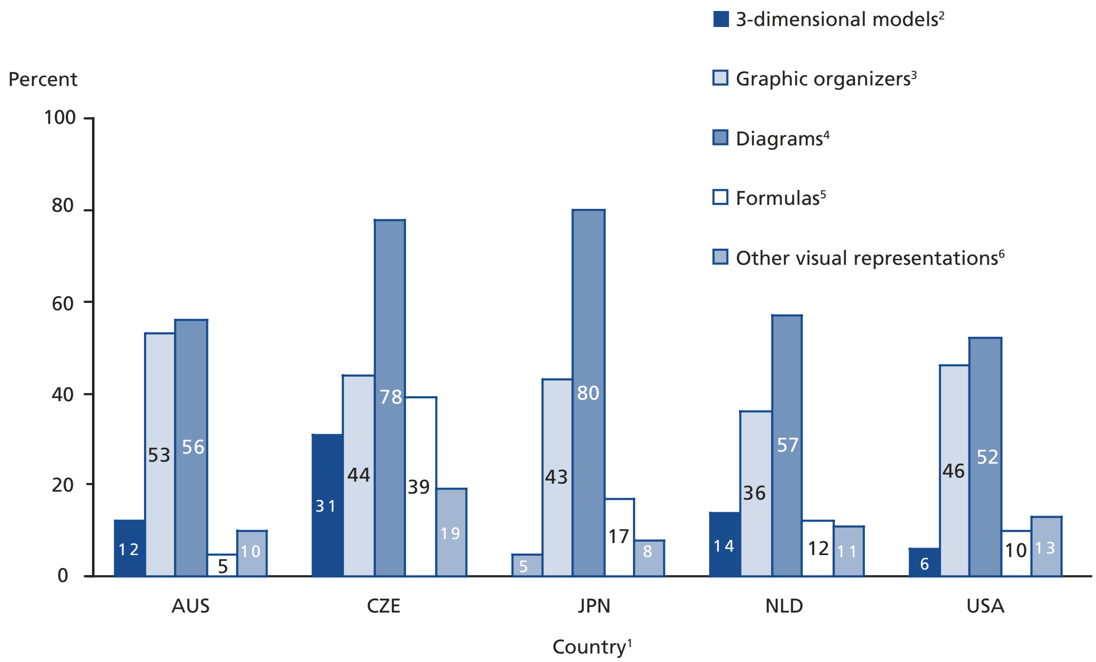
!Interpret data with caution. Estimate is unstable.
‡Reporting standards not met. Too few cases to be reported.
[1] AUS=Australia; CZE=Czech Republic; JPN=Japan; NLD=Netherlands; and USA=United States.
[2] Earth science: Making connections: No measurable differences detected.
[3] Earth science: Acquiring facts, definitions, and algorithms: No differences detected.
[4] Life science: Making connections: No measurable differences detected.
[5] Life science: Acquiring facts, definitions, and algorithms: CZE>AUS, JPN, USA; NLD>JPN.
[6] Physics: Making connections: AUS>CZE, USA; JPN>USA.
[7] Physics: Acquiring facts, definitions, and algorithms: NLD>JPN, USA.
[8] Chemistry: Making connections: JPN>NLD, USA.
[9] Chemistry: Acquiring facts, definitions, and algorithms: No measurable differences detected.
[10] Other areas: Making connections: No measurable differences detected.
[11] Other areas: Acquiring facts, definitions, and algorithms: No measurable differences detected.
NOTE: Totals may not sum to 100 because of rounding and data not reported.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.

**FIGURE E.1. Percentage distribution of eighth-grade science lessons that primarily developed science content by focusing on different approaches to acquiring facts, definitions, and algorithms, by country: 1999**




‡Reporting standards not met. Too few cases to be reported.
[1] AUS=Australia; CZE=Czech Republic; JPN=Japan; NLD=Netherlands; and USA=United States
[2] Focus on algorithms and techniques: No measurable differences detected.
[3] Focus on sequences of events: No measurable differences detected.
[4] Focus on discrete bits of information: NLD>AUS, JPN; CZE>JPN.
[5] Focus on unidentified approaches: No measurable differences detected.
NOTE: Only those lessons identified as developing science content primarily by acquiring facts, definitions, and algorithms are included in the analysis. See figure 5.5 for the total percentage of lessons that developed content by acquiring facts, definitions, and algorithms.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study (TIMSS), Video Study, 1999.

**FIGURE E.2. Percentage of eighth-grade science lessons that incorporated various types of visual representations to support science knowledge, by country: 1999**




[1] AUS=Australia; CZE=Czech Republic; JPN=Japan; NLD=Netherlands; and USA=United States.
[2] 3-dimensional models: CZE>JPN, USA.
[3] Graphic organizers: No measurable differences detected.
[4] Diagrams: JPN>AUS, NLD, USA; CZE>USA.
[5] Formulas: CZE>AUS, JPN, NLD, USA.
[6] Other visual representations: No measurable differences detected.
NOTE: A lesson may include more than one type of visual representation.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study (TIMSS), Video Study, 1999.

**Table E.7. Percentage distribution of eighth-grade science lessons that were judged to include challenging or a mix of basic and challenging content, and basic content, by content topics and country: 1999**

| Discipline and teaching method | Country[1] | | | | |
|---|---|---|---|---|---|
| | AUS | CZE | JPN | NLD | USA |
| Earth science | | | | | |
| Challenging or a mix of basic and challenging[2] | ‡ | ‡ | ‡ | ‡ | 13 |
| Basic[3] | 4! | ‡ | 6 | ‡ | 14 |
| Life science | | | | | |
| Challenging or a mix of basic and challenging[4] | 9 | 30 | 10 | 27 | 10 |
| Basic[5] | 14 | 6 | 9 | 4 | 7! |
| Physics | | | | | |
| Challenging or a mix of basic and challenging[6] | 26 | 24 | 15 | 20 | 11 |
| Basic[7] | 22 | 4 | 20 | 26 | 6 |
| Chemistry | | | | | |
| Challenging or a mix of basic and challenging[8] | 4! | 18 | 9 | ‡ | 16 |
| Basic[9] | 11 | 6 | 28 | 6 | ‡ |
| Other areas | | | | | |
| Challenging or a mix of basic and challenging[10] | ‡ | 7 | ‡ | ‡ | ‡ |
| Basic[11] | 6 | ‡ | ‡ | 10 | 19 |

‡Reporting standards not met. Too few cases to be reported.
!Interpret data with caution. Estimate is unstable.
[1] AUS=Australia; CZE=Czech Republic; JPN=Japan; NLD=Netherlands; and USA=United States.
[2] Earth science: Challenging or a mix of challenging and basic: No measurable differences detected.
[3] Earth science: Basic: No measurable differences detected.
[4] Life science: Challenging or a mix of challenging and basic: CZE>AUS, JPN, USA; NLD>JPN.
[5] Life science: Basic: No measurable differences detected.
[6] Physics: Challenging or a mix of challenging and basic: No differences detected.
[7] Physics: Basic: AUS, NLD>CZE, USA; JPN>USA.
[8] Chemistry: Challenging or a mix of challenging and basic: CZE, USA>AUS.
[9] Chemistry: Basic: JPN>AUS, CZE, NLD.
[10] Other areas: Challenging or a mix of challenging and basic: No difference detected.
[11] Other areas: Basic: No measurable differences detected.
NOTE: Totals may not sum to 100 because of rounding and data not reported. The tests for significance take into account the standard error for the reported differences. Thus, a difference between averages of two countries may be significant while the same difference between two other countries may not be significant.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Third International Mathematics and Science Study, Video Study, 1999.