APPENDIX I

DATA CLEANING AND CONSISTENCY CHECKS

CHECKS UNDERTAKEN AT THE DPC

This appendix contains a complete compilation of all checks undertaken at the DPC. Each check is presented in the following way:

Problem Number (Affected files): Description of the problem

> Applied cleaning rule: Corrections have been undertaken

> Applied cleaning rule: A warning has been given, but no corrections have been undertaken.

Please Note: The cleaning rule reports the action undertaken at the DPC if no solution could be found by inspecting the instruments and forms and if no instructions on how to handle these cases were given by the national center.

COLUMN SHIFT CHECK

This cleaning step identifies potential column shifts. It identifies all observations in which valid data have been found in the CHECKn variables, which should be blank.

101 (CG, TG, SA, SG, SR): The listed observations include one CHECKn variable that contains a value.
> If the reason for the column shift could not be identified, all data between the effected CHECKn variable and the preceding one have been coded to 'Not administered'.

102 (CG, TG, SA, SG, SR): The listed observations include more than one CHECKn variable that contains values.
> If the reason for the column shifts could not be identified, all data between the last correct CHECKn variable and this one have been coded to 'Not administered'.

IDENTIFICATION CHECK - STUDENT FILE

This set of cleaning steps identifies all problems with IDs within an observation. This includes deviations from the hierarchical ID system described in the Main Study Manual as well as missing or incomplete identification.

201 (SA, SG, SR): The listed observations include a deviation from the hierarchical ID system. The Class ID, Student ID, and, if applicable, Teacher IDs all match, but all are inconsistent with the School ID.
> The School ID has been made consistent with the other IDs.

202 (SA, SG, SR): The listed observations include a deviation from the hierarchical ID system. The School ID, Student ID, and, if applicable, all Teacher IDs match, but all are inconsistent with the Class ID.
> The Class ID has been made consistent with the other IDs.

203 (SA, SG, SR): The listed observations include deviations from the hierarchical ID system. The Class ID does not match the class-specific part of the Student ID.

If it was not possible to identify the class to which the student belongs, the student has been treated as nonparticipating and has been assigned a weight of zero.

204 (SA, SG, SR): The listed observations include a deviation from the hierarchical ID system. The School ID, Class ID, and, if applicable, all Teacher IDs match, but all are inconsistent with the Student ID.

The Student ID has been made consistent with the other IDs.

205 (SG): The listed observations include a deviation from the hierarchical ID system. The School ID, Class ID, Student ID, and all but one of the Teacher IDs match.

For the nonmatching ID, the school-specific part of the Teacher ID has been made consistent with the other IDs.

206 (SG): The listed observations include Teacher IDs (ILTEACH1-6) that have a different number of digits from the other IDs.

If the correct Teacher ID could not be identified, the Teacher ID has been dropped and the corresponding variables ILTEACHx and ILLINKx have been coded to 'Not administered' (99998) and (98) respectively.

207 (SG): The listed observations include Teacher IDs (ILTEACH1-6) whose corresponding Teacher Link Numbers (ILLINK1-6) have been coded to 'Not administered' (98).

If no Teacher Link Number could be identified from the *Teacher Tracking Forms*, a new Teacher Link Number has been assigned. In this unlikely case, all course-related data would be lost and only person-related data would be available.

208 (SA, SG, SR):): The listed observations include more than one deviation from the hierarchical ID system.

If it was not possible to recreate the correct identification, the student has been treated as nonparticipating and a weight of zero has been assigned.

IDENTIFICATION CHECK - SCHOOL FILE

This set of cleaning steps identifies all inconsistencies between identification variables within the School File. This includes all inconsistencies between the IDs of replaced schools and originally sampled schools, as well as their Participation Indicators. Strata with only one school have also been flagged.

301 (CG): The listed schools are indicated as replacement schools (ILREPLAC contains a valid ID), but their Participation Indicators (ITPARTx) show that they are participating as originally sampled schools.

ITPARTx has been recoded to 'First replacement school' (1) or 'Second replacement school' (2).

302 (CG): The listed schools are indicated as replacement schools (ILREPLAC contains a valid ID).

The replaced school indicated by ILREPLAC has been entered.

303 (CG): The listed schools are indicated as having been replaced by other schools (their IDs were found in ILREPLAC for other schools), but their Participation Indicators (ITPARTx) are different from 'Nonparticipating' (0).

If data were entered for a corresponding school, the corresponding entry for ILREPLAC has been recoded to 'Not administered' (998). If no data were entered for a corresponding school, the Participation Indicator has been recoded to 'Nonparticipating' (0).

304 (CG): The listed schools are indicated as being 'Nonparticipating' (0), but no replacement schools have been entered.
> A warning has been given, but no corrections have been undertaken.

305 (CG): The listed schools belong to strata which contain only one school.
> A warning has been given, but no corrections have been undertaken.

IDENTIFICATION CHECK - TEACHER FILE

This set of cleaning steps identifies all problems with IDs within an observation. This includes deviations from the hierarchical ID system described in the Main Study Manual as well as missing or incomplete identification. Please note that all recoding has been performed under consideration of its consistency with other files.

401 (TG): The listed observations include a deviation from the hierarchical ID system. The Teacher ID and Class IDs all match, but all are inconsistent with the School ID.
> The School ID has been made consistent with the other IDs.

402 (TG): The listed observations include a deviation from the hierarchical ID system. The School ID and Class IDs all match, but all are inconsistent with the Teacher ID.
> The Teacher ID has been made consistent with the other IDs.

403 (TG): The listed observations include a deviation from the hierarchical ID system. The School ID, Teacher ID, and Class IDs all match, but all are inconsistent with one of the Class IDs.
> This Class ID has been made consistent with the other IDs.

404 (TG): The listed observations include a Teacher ID whose length is inconsistent with that of the other IDs.
> A new consistent Teacher ID has been assigned.

405 (TG): The listed observations include a Class ID whose length is inconsistent with that of the other IDs.
> This Class ID has been made consistent with the other IDs.

406 (TG): The listed observations include more than one deviation from the hierarchical ID system.
> The effected cases have been dropped from the file.

407 (TG): The listed observations do not include Teacher Link Numbers.
> A dummy link number has been assigned. Teacher-student linkage is not possible.

408 (TG): The listed observations include a Class ID twice.
> The redundant entry has been set to 'Not applicable' (99998).

409 (TG): The listed observations do not include Class IDs.
> A warning has been given, but no corrections have been undertaken.

410 (TG): The listed observations have a different subject (IDSUBJCT) from the booklet administered. Only applicable to Population 2.
> The Subject ID has been made consistent with the administered questionnaire.

DUPLICATE ID CHECK

This set of cleaning steps identifies all problems with IDs between observations. This includes duplicate IDs or records and inconsistent IDs within homogenous groups; e.g., inconsistent grade identification within a class.

501 (CG, TG, SA, SR, SG): The listed observations contain exactly the same IDs and data.
   One of the observations has been dropped.

502 (CG, TG, SA, SR, SG): The listed observations contain exactly the same IDs, but different data.
   A dummy ID has been be created for at least one of the observations.

504 (CG, TG, SA, SR, SG): The listed observations include a Population ID (IDPOP) that does not match the file.
   The Population ID has been corrected to agree with the source file.

505 (CG, TG, SA, SR, SG): The listed observations belong to the same school, but their Stratum IDs are different.
   The Stratum ID has been recoded to agree with the majority of entries.

506 (SG): The listed observations belong to the same class, but their Grade IDs are different.
   The Grade ID has been recoded to agree with the majority of entries and the teacher data.

507 (TG): The listed teachers teach the same classes, but the Grade IDs are different.
   The Grade ID has been recoded to agree with the majority of entries and the student data.

508 (SG): The listed observations belong to the same class, but their Language IDs are different.
   The Language ID has been recoded to agree with the majority of entries and the student data.

509 (SG): The listed observations belong to the same class, but their Stream IDs are different.
   The Stream ID has been recoded to agree with the majority of entries and the student data.

510 (SG): The listed observations belong to the same class, but their Test Administrator IDs (ITADMINI) are different.
   A warning has been given, but no corrections have been undertaken.

TEST INDICATOR CHECK

This set of cleaning steps identifies inconsistencies between the Test Indicator Variables (TOKENxx) and the data variables. The code '3' should be assigned if the corresponding instrument was administered. All other codes indicate that the instrument was not administered or that the instrument has been lost. Please note that the TOKENxx variables represent the availability of data. The information given with the cleaned TOKENxx variables has been transcribed into the Participation Indicators ITPARTx (which summarizes the availability of data, the participation status, and the exclusion status) after all cleaning steps were finalized. Therefore, the TOKENxx variables became redundant after cleaning and have been removed from the files.

601 (CG, TG, SG, SA, SR): The listed observations include at least one TOKENxx variable that is coded invalidly and less than five valid data have been found.
   The TOKENxx variable has been recoded to 'Nonparticipating' (2 for teachers, 1 for other respondents).

602 (CG, TG, SG, SA, SR): The listed observations include at least one TOKENxx variable that is coded invalidly, but more than five valid data have been found.
   The TOKENxx variable has been recoded to 'Participating' (3).

603 (CG, TG, SG, SA, SR): The listed observations include at least one TOKENxx variable that is coded to 'Participating' (3) and less than 3 corresponding data variables are different from 'Not administered'.
   The TOKENxx variable has been recoded to 'Nonparticipating' (2 for teacher, 1 for other respondents).

604 (CG, TG, SG, SA, SR): The listed observations include at least one TOKENxx variable that is coded to 'Not participating' (2) for teachers, (1) for all others and more than ten valid data have been found.
   The TOKENxx variable has been recoded to 'Participating' (3).

605 (SG, SA, SR): The listed observations include at least one TOKENxx variable that has been coded to 'Booklet lost' (2), but valid data have been found.
   The corresponding TOKENxx variable has been recoded to 'Participating' (3).

606 '(CG, TG, SG, SA, SR): The listed observations include at least one TOKENxx variable that is coded to 'Participating' (3) but only missing data (different from 'Not administered') have been found.
   These cases should be checked carefully. If NRCs did not indicate changes to these cases, it is assumed that the respondent returned the questionnaire but left everything blank, e.g., because he/she did not cooperate. A warning has been given, but no corrections have been undertaken.

610 (SG, SA, SR): The listed students participated in only one testing session.
   These cases have been flagged because this can be an indication of data loss and should be checked. No corrections are necessary.

611 (SG, SA, SR): The listed students did not participate in a testing session. These cases should not be entered into the Student Achievement File or the Reliability File.
   These cases have been flagged. Because dummy records have been created for all nonparticipating students, a warning has been given, but no corrections have been undertaken.

612 (CG): The listed schools are indicated as 'Nonparticipating' (0) by the Participation Indicator Variable ITPARTx, but valid data have been found.
   The Participation Indicator has been recoded to 'Participating' (3).

613 (CG): The listed schools are indicated as 'Participating' (1, 2, or 3) by the Participation Indicator Variable ITPARTx, but no valid data have been found and the corresponding TOKENxx variable indicates that the school did not participate.
   The TOKENxx or ITPARTx variable has been recoded to agree with to the student and teacher data for this school.

BOOKLET ID CHECK

This set of cleaning steps identifies inconsistencies between the Booklet ID (IDBOOK) and the corresponding Test Indicator Variables (TOKENxA, TOKENxB). The Booklet ID indicates which booklet has been administered to the students. The corresponding TOKENxx variables should be coded to '2' or '3' and all other TOKENxx variables should be coded to '1'. Please note that the TOKENxx were cleaned before, which means they represent the existence of data. All mispunches of TOKENxx have been corrected during the

Test Indicator Check; data are available. If a TOKENxx is different from '3', there are no data available.

701 (SA, SR): The listed observations include a Booklet ID that has an invalid code, and no TOKENxx variables are coded to '3'.
   IDBOOK has been coded to 'No booklet assigned' (0).

702 (SA, SR): The listed observations include a Booklet ID that has an invalid code, and one TOKENxx variable is coded to '3'.
   IDBOOK has been made consistent with the TOKENxx variable, i.e. it has been coded to agree with the data found in the record.

703 (SA, SR): The listed observations include a Booklet ID that has an invalid code, and two TOKENxx variables corresponding to one booklet are coded to '3'.
   IDBOOK has been made consistent with the TOKENxx variable, i.e. it has been coded to agree with the data found in the record.

704 (SA, SR): The listed observations include a Booklet ID that has an invalid code, and two or more TOKENxx variables corresponding to different booklets are coded to '3'.
   All data have been recoded to 'Not administered' and IDBOOK has been coded to 'No booklet assigned' (0).

705 (SA, SR): The listed observations include a Booklet ID that has a valid code, but one (and only one) TOKENxx variable (which does not correspond to the appropriate booklet) are coded to '3'.
   IDBOOK has been made consistent with the data.

706 (SA, SR): The listed observations include a Booklet ID that has a valid code, but two TOKENxx variables (which correspond to the same booklet) are coded to '3'. The booklet to which the two TOKENxx variables correspond is different from the one indicated by IDBOOK.
   IDBOOK has been made consistent with the data.

707 (SA, SR): The listed observations include a Booklet ID that has a valid code, but TOKENxx variables, which correspond to a different booklet from that indicated by IDBOOK, have been identified. The data conflict as to which booklet was administered.
   All data have been recoded to 'Not administered'; the data corresponding to IDBOOK have been marked as 'Lost'.

LINKAGE CHECK STUDENT GENERAL FILE -> STUDENT ACHIEVEMENT FILE

This set of cleaning steps identifies inconsistencies between the Exclusion Indicator IDEXCLUD, the Participation Indicators ITPART1 (first testing session), ITPART2 (second testing session), ITPART3 (Student General File), and the corresponding TOKENxx variables. Therefore, the linkage between the Student General File and the Student Achievement File will also be checked. Before correcting inconsistencies between TOKENxx variables in one file and Participation Indicator Variables ITPART1 and ITPART2 in the other, the ITPARTx linkage has to be checked carefully.

1001 (SA, SG): The listed observations include Participation Indicators ITPARTx that identify the particular student as 'Participating' (3), but no matching data have been found in the Student Achievement File. Either

the Participation Indicator is coded incorrectly or the linkage between the Student General File and the Student Achievement File is incorrect.

A dummy record has been created for the achievement data, which have been marked as 'Lost'; the corresponding TOKENxx variables have been coded to 'Booklet/data lost' (2).

1002 (SA, SG): The listed observations include the Participation Indicators ITPARTx that identify the particular student as 'Nonparticipating' (0, 1, or 2), but a matching record, which contains valid data, has been found in the Student Achievement File. Either the Participation Indicator ITPARTx is coded incorrectly or the linkage between the Student General File and the Student Achievement File is incorrect.

The Participation Indicator ITPARTx has been made consistent with the data, i.e., it has been coded to 'Participating' (3).

1003 (SA, SG): The listed observations include a record in the Student Achievement File for which all TOKENxx for the tested booklet are coded to 'Nonparticipating' (1) and for which the corresponding ITPARTx variables identify the student as 'Nonparticipating' (0, 1, or 2). Records without data should not be entered into the Student Achievement File.

No corrections are necessary.

1004 (SA, SG): The Participation Indicators ITPARTx for the listed observations identify the student as 'Nonparticipating' (0, 1, or 2) and no valid data have been found. At least two of the Participation Indicators ITPARTx mark the student as excluded, but the Exclusion Indicator IDEXCLUD is coded to 'Not excluded' (9).

The Participation Indicators ITPARTx have been recoded to 'Student absent' (2) instead of 'Excluded' (0).

1005 (SA, SG): The Participation Indicators ITPARTx for the listed observations identify the student as 'Participating' (3) and valid data have been found, but the Exclusion Indicator IDEXCLUD is coded differently from 'Not excluded' (9).

The Participation Indicators ITPARTx have been recoded to 'Participating, but excluded' (5).

1006 (SA, SG): The Participation Indicators ITPARTx for the listed observations identify the student as 'Nonparticipating' (0, 1, or 2) and no valid data have been found, but the Exclusion Indicator IDEXCLUD has an invalid code.

A warning has been given, but no corrections have been undertaken.

1007 (SA, SG): All but one of the Participation Indicators ITPARTx for the listed observations are consistent both with the data found and with the Exclusion Indicator IDEXCLUD. Valid data have been found for the corresponding session.

The Participation Indicator ITPARTx has been made consistent with the data.

1008 (SA, SG): All but one of the Participation Indicators ITPARTx for the listed observations are consistent both with the data found and with the Exclusion Indicator IDEXCLUD. One Participation Indicator ITPARTx is coded to 'Nonparticipating' and the corresponding booklet is indicated by TOKENxx as being 'Lost' (2).

The TOKENxx variable has been coded to 'Nonparticipating' (1).

1009 (SA, SG): All but one of the Participation Indicators ITPARTx for the listed observations are consistent both with the data found and with the Exclusion Indicator IDEXCLUD. No data have been found for the corresponding session, but the booklet is not indicated by TOKENxx to be 'Lost' (2).

The Participation Indicator ITPARTx has been coded to 'Booklet/data lost' (4).

1010 (SA, SG): The listed observations include inconsistencies between Participation Indicators. Different codes were given for nonparticipation (0 = 'Student excluded', 1 = 'Student left school', or 2 = 'Student absent').
These codes have been made consistent. Either 'Student left school' or 'Student absent' was used twice and the third Participation Indicator has been made consistent, or all Participation Indicator Variables which indicate nonparticipation have been recoded to 'Student absent' (2). If a student was excluded and no valid data have been found, all Participation Indicators have been recoded to 'Student excluded' (0).

1011 (SA, SG): The listed observations include several inconsistencies between Participation Indicators, Exclusion Indicators, and data.
Inconsistent indicators have been made consistent with the data. Participation indicators ITPARTx and TOKENxx have been recoded to 'Participating' (3) if data were found. ITPARTx has been coded to 'Student absent in session' if no data were found.

1012 (SA, SG): The listed observations include inconsistencies between the Class ID (School ID) in the Student Achievement File and the Class ID (School ID) in the Student General File.
As with many other problems, this problem creates at least two messages in the cleaning report. The rules for the Identification Check, where these cases have also been reported, have been applied.

LINKAGE CHECK RELIABILITY FILE -> STUDENT ACHIEVEMENT FILE

This set of cleaning steps identifies all inconsistencies between the Reliability Coding File and the corresponding Student Achievement File.

1111 (SA, SR): The listed observations in the Student Achievement File do not match those in the corresponding Reliability Coding File.
The affected cases have been dropped from the Reliability Coding File.

1112 (SA, SR): The Reliability Coding Files and the Student Achievement Files for the listed observations indicate that different booklets were administered.
The affected cases have been dropped from the Reliability Coding File.

1113 (SA, SR): The listed observations are missing some data for the Reliability Coding File, although the Student Achievement File contains corresponding data.
No corrections could be performed.

1114 (SA, SR): The booklet set (indicated by ITBSET in the Reliability Coding File) for the listed observations' is different from the booklet set indicated in the Student Achievement File.
No corrections could be performed. It should be checked whether the data were obtained from the same booklets.

1115 (SA, SR): The listed observations indicate the same coder for both the Reliability Coding File and the Student Achievement File.
No corrections could be performed. It should be checked if the reliability coding followed the correct procedures.

1116 (SA, SR): The listed observations include inconsistencies between the Class ID (School ID) within the Student Achievement File and the Class ID (School ID) in the Reliability Coding File.
This problem creates at least two messages in the cleaning report. The rules for the Identification Check, where these cases have also been reported, have been applied.

1121 (SA, SR): The listed observations include a record in the Student Achievement File, which corresponds to one in the Reliability Coding File, which contains invalid data (all booklets have been lost or the student did not participate). These students should not be selected for reliability coding.
   The affected cases have been dropped from the Reliability Coding File.

LINKAGE CHECK STUDENT ACHIEVEMENT FILE -> STUDENT GENERAL FILE

This cleaning step identifies records in the Student Achievement File which can not be merged to the Student General File.

1201 (SA, SG): The listed observations do not include any matching observations within the Student General File.
   A dummy record has been created for the background data and all Participation Indicators have been made consistent with the data, i.e., the indicators corresponding to the background session have been coded to 'Booklet lost'. Indicators that could be derived from the achievement data, e.g., the Grade ID from the Class ID, have been corrected.

STUDENT - TEACHER LINKAGE CHECK

This set of cleaning steps identifies linkage problems between the Teacher and Student Files.

1311 (SG, TG): The listed observations include a combination of Teacher ID and Teacher Link Number which can not be found in the Teacher Files.
   A dummy teacher has been created so that complete linkage is possible. All indicators in the dummy teacher record have been set to 'Teacher did not participate'.

1312 (SG, TG): The listed observations include a class linked to the teacher (in ILCLASS1-3) which can not be found in the Student Files.
   The affected variable ILCLASSx has been recoded to 'Not administered'.

1313 (SG, TG): The listed observations include a Grade ID that is inconsistent between the teacher and the students linked to him/her.
   The Grade IDs have been made consistent for students and teachers to agree with the majority of data.

1314 (SG, TG): The following teachers' students are from classes which were not given in ILCLASS1, ILCLASS2, or ILCLASS3.
   All class IDs corresponding to classes in which students are linked to the particular teacher will be added to ILCLASS1-3.

1315 (SG, TG): The following teachers are linked to classes (ILCLASS1-3), which are represented in the datafiles, but no student is linked to the corresponding teacher. (Variables ILTEACH1-6, ILLINK1-6)
   IDs have been made consistent.

1321 (SG): No math teacher is indicated for the listed students.
   No corrections could be performed.

1322 (SG): No science teacher is indicated for the listed students.
   No corrections could be performed.

1323 (SG): No teacher is linked to the listed students.
   No corrections could be performed.

SCHOOL - STUDENT LINKAGE CHECK

This set of cleaning steps identifies linkage problems between the Student and School Files.

1411 (CG, SG): The listed schools are marked as 'Not participating' (0), but student data have been found in the Student File.
> The participation status has been coded to 'Participating' (3) to agree with the data found.

1412 (CG, SG): The listed schools are indicated to be 'Participating' (1, 2, or 3), but no student data have been found in the Student File.
> No corrections could be performed.

1413 (CG, SG): The listed schools have a different Stratum ID than the students linked to them.
> The Stratum IDs have been made consistent for the school, the teacher, and all students, to agree with the majority of data.

1414 (CG, SG): The data for the listed schools includes more students than are enrolled in the schools (ITMOS).
> No corrections could be performed.

1421 (SG): The listed observations are linked to schools which have not been found in the School File.
> A dummy school has been created.

SCHOOL - TEACHER LINKAGE CHECK

This set of cleaning steps identifies linkage problems between the Teacher and School Files.

1511 (CG, TG): The listed schools have different Stratum IDs from the teachers linked to them.
> The Stratum IDs have been made consistent for the school, the teacher, and all students to agree with the majority of data.

1512 (TG): The listed schools are marked as 'Not participating' (0), but teacher data have been found in the Teacher File.
> Corrections have been undertaken, depending on whether students were found for the corresponding schools or not.

1513 (TG): The listed schools are indicated to be 'Participating' (1, 2, or 3), but no teacher data have been found in the Teacher File.
> Corrections have been undertaken, depending on whether students were found for the corresponding schools or not.

1521 (TG): The listed observations are linked to schools which have not been found in the School File.
> Corrections have been undertaken, depending on whether students were found for the corresponding schools or not.

RANGE VALIDATION CHECK

This cleaning step identifies variables with values which do not match the range validation criteria as specified in the electronic codebook (structure database).

2001 The listed observations include noncategorical variables with values which do not match the range validation criteria.

> These variables have been set to 'Invalid'.

2002 The listed observations include categorical variables with values which do not match the range validation criteria.

> These variables have been set to 'Invalid'.

STUDENT SPLIT VARIABLE CHECK

In some question blocks, students have been allowed to answer using one of two response options per question (e.g., 'Yes' or 'No' to questions on home possessions). These question blocks have been coded using dichotomous variables. Distinguishing between 'Missing/nonresponse' and 'No' is a problem. Respondents often do not use the 'No' option and mark only options with 'Yes'. For questions with a large number of options, it is assumed that in case, in which no 'No' option was used but 'missing' was coded, the respondent meant 'No'. This algorithm can not work for questions with only a few items.

2101 (SG): The listed cleaning step identifies all problems in the question block 'Do each of these people live at home with you most or all of the time?

> If at least one of the dichotomous variables has been coded to 'Yes' (1) and all other variables have been coded to 'Missing', then all variables coded to 'Missing' have been recoded to 'No'.

2102 (SG): The listed cleaning step identifies all problems in the question block 'Do you have any of these items at home?

> If at least one of the dichotomous variables has been coded to 'Yes' (1) and all other variables have been coded to 'Missing', then all variables coded to 'Missing' have been recoded to 'No' (2).

TEACHER SPLIT VARIABLE CHECK

The following problems have been combined in this category:

In some question blocks, teachers were allowed to answer using one of two response options per question (e.g., 'Yes' or 'No' to questions on grades taught). These question blocks were coded using dichotomous variables. Distinguishing between 'Missing/nonresponse' and 'No' is a problem. Respondents often do not use the 'No' option and mark only options with 'Yes'. For questions with a large number of options, it is assumed that in case, in which no 'No' option was used but several times 'missing' was coded, the respondent meant 'No'. This algorithm can not work for questions with only a few items.

In some question blocks, lists were given (e.g., in the Section 'Opportunity to Learn'). The elements of the lists could be' Checked' or 'Not checked'. Distinguishing between 'Not checked' and 'Missing/nonresponse' is not possible. It is assumed, that in cases where 'Missing' was coded and at least one option was checked, 'Not checked' should have been coded.

In other question blocks, teachers were asked to assign numbers (e.g., order or time). Distinguishing between '0' and 'Missing/nonresponse' is a problem. Respondents often do not use the '0' option and enter only numbers for options which are applicable. For questions with a large number of options, it is assumed that if no '0' option was used but several times 'missing' was coded, the respondent meant '0'. This algorithm can not work for questions with only a few items.

2201 (TG): In the question blocks concerning the grades taught, teachers were allowed to answer using one of two response options per question ('Yes' or 'No'). These question blocks were coded using dichotomous variables. Distinguishing between 'Missing/nonresponse' and 'No' is a problem.
If at least one of the dichotomous variables are coded to 'Yes' (1) and all other variables are coded to 'Missing' (9), then all variables coded to 'Missing' have been recoded to 'No'(2).

2202 (TG): In the question block concerning the textbooks used, teachers were allowed to answer using one of two response options per question ('Yes' or 'No'). These question blocks were coded using dichotomous variables. Distinguishing between 'Missing/nonresponse' and 'No' is a problem.
If at least one of the dichotomous variables are coded to 'Yes' (1) and all other variables are coded to 'Missing' (9), then all variables coded to 'Missing' have been recoded to 'No'(2).

2203 (TG): In the question block concerning the topics in the last lesson (A/BTBMTOxx and BTBSTOxx), teachers were allowed to answer using one of two response options per question ('Yes' or 'No'). These question blocks were coded using dichotomous variables. Distinguishing between 'Missing/nonresponse' and 'No' is a problem.
If at least one of the dichotomous variables is coded to 'Yes' (1) and all other variables are coded to 'Missing' (9), then all variables coded to 'Missing' have been recoded to 'No'(2).

2204 (TG): Teachers were asked to list the activities of the last class hour in order.
If at least one number is assigned, but no 0 has been detected, all 'Missing' have been recoded to 0.

2206 (TG): Teachers were asked to assign times to the activities in the last lesson. They should enter 0 if they did not do a certain activity
If at least one number is assigned, but no 0 has been detected, all 'Missing' have been recoded to 0.

2208 (TG): This cleaning step identifies inconsistencies in the lists concerning the opportunity to learn a topic. (Population 2: Section C) For each topic, two lists were given, one which should have been answered if the topic was taught and the second which should have been answered if the topic was not taught. Both lists have been checked separately for internal consistency
If at least one variable in a list is coded to 'Checked' (2), all variables coded to 'Missing' (9) have been recoded to 'Not checked' (1).
If no variables in a list are coded to 'Checked' (2), all variables coded to 'Not checked' (1) have been recoded to 'Missing' (9).

SCHOOL SPLIT VARIABLE CHECK

The listed problems have been combined in this category: In some question blocks, principals were allowed to answer using one of two response options per question (e.g., 'Yes' or 'No' to questions on grades found in the school). These question blocks were coded using dichotomous variables. Distinguishing between 'Missing/nonresponse' and 'No' is a

problem. Respondents often do not use the 'No' option and mark only options with 'Yes'. For questions with a large number of options, it is assumed that if no 'No' option was used but 'missing' was coded, the respondent meant 'No'. This algorithm can not work for questions with only a few items.

In other question blocks, principals were asked to assign numbers (e.g., time). Distinguishing between '0' and 'Missing/nonresponse' is a problem. Respondents often do not use the '0' option and enter only numbers for options which are applicable. For questions with a large number of options, it is assumed that if no '0' option was used but several times 'missing' was coded, the respondent meant '0'. This algorithm can not work for questions with only a few items.

2301 (CG): This cleaning step identifies all problems in the question block 'Which of following grade levels are found in your school?'
If at least one of the dichotomous variables is coded to 'Yes' (1) and all other variables are coded to 'Missing' (9), then all other variables coded as 'Missing' (9) have been recoded to 'No' (2).

2302 (CG): This cleaning step identifies all cases in which the respondent was asked to write a 0 for 'None', but he/she ignored this and left all fields blank. The listed variables have been checked:(1) number of full-time equivalents A/BCGFTE1-6;(2) principal's activities A/BCGAC01-14;(3) percentage of students coming from different background (A/BCBGSTD1-8.)
In all cases in which no '0', one or more 'Missing', and valid answers were given, all 'Missing' have been recoded to '0'.

2303 (CG): The listed cleaning step identifies all problems in the question block 'On what basis are pupils admitted to your school ?' (Population 1 SCQ1-29x, Population 2 SCQ2-30x)
If at least one of the dichotomous variables is coded to 'Yes' (1) and all other variables are coded to 'Missing' (9), then all other variables coded as 'Missing' (9) have been recoded to 'No' (2).

STUDENT FILTER VARIABLE CHECK

These cleaning steps identify all problems between filter variables and dependent variables in the Student File(s). In a first cleaning step, the filter variable has been made consistent with the dependent variable(s). In a second cleaning step all dependent variables coded to 'Missing' have been recoded to 'Logically not applicable' if the filter variable was coded to 'No'.

2401 (SG): This cleaning step identifies all problems between the statement concerning 'Born in country' and 'Age when student came to country. ('Note: this check requires the following checks (I) Check of consistency in 'Age of Student' obtained from tracking information and questionnaire information. (II) Check of consistency in 'Date of Testing'. (III) Check of consistency between 'Age of student' and 'Age when student came to country'.)
If the student's age is valid but the student indicated in the filter question that he/she was born in country (1), then the filter variable has been recoded to 'No' (2).

2402 (SG): The listed cleaning step identifies all problems between the statement concerning 'Born in country' and 'Age when student came to country. (Note: this check requires the following checks: (I) Check of consistency in 'Age of Student' obtained from tracking information and questionnaire information. (II) Check of

consistency in 'Date of Testing'.(III) Check of consistency between 'Age of student' and 'Age when student came to country'.

If the student's age when he/she came to the country was invalid and the student indicated in the filter question that he/she was 'Born in country' (1), then the student's age when he/she came to the country has been recoded to 'Logically not applicable' (96).

2403 (SG): This cleaning step reports all inconsistencies between the filter question concerning <SCIENCE SUBJECT> and completion of questions in the corresponding part of the *Student Questionnaire*. (Note: <BIO> refers to the biology part, <CHE> refers to the chemistry part, <EAR> refers to the earth science part, <PHY> refers to the physics part.

If the student states in the filter question that he/she does not study <SUBJECT> this year (1), but answers more than 3 questions concerning the <LESSON>, the filter has been recoded to 'Yes' (2).

If the student states in the filter question that he/she does not study <SUBJECT> this year (1), but answers more than 2 questions concerning the <SUBJECT>, a warning has been given, but no corrections have been undertaken.

2404 (SG): The listed cleaning step identifies all inconsistencies between the filter question concerning <SUBJECT> and completion of less than 3 questions in the corresponding part of the *Student Questionnaire*. Note: <BIO> refers to the biology part, <CHE> refers to the chemistry part, <EAR> refers to the earth science part, <PHY> refers to the physics part.

If variables coded differently from 'Not administered' (8) have been found in the part under consideration, the variables have been recoded to 'Logically not applicable' (B).

TEACHER FILTER VARIABLE CHECK

This cleaning steps identifies all problems between filter variables and dependent variables in the Teacher Files. In a first cleaning step the filter variable has been made consistent with the dependent variable(s). In a second cleaning step all dependent variables coded to 'Missing' have been recoded to 'Logically not applicable' if the filter variable was coded to 'No'.

2501 (TG): This cleaning step identifies inconsistencies between a filter which indicates whether <SUBJECT> is taught this year and dependent variables which indicate at which grade levels <SUBJECT> is taught. (Note: <SUBJECT> can be mathematics or science.)

If the filter indicates that NO <SUBJECT> is taught (2), but at least one grade level variable indicates that <SUBJECT> is taught, then the filter has been recoded to 'Taught' (1).

If the filter indicates that <SUBJECT> is taught (1), but NO grade level variable indicates that <SUBJECT> is taught, then the filter has been recoded to 'Not taught' (2).

2502 (TG): This cleaning step identifies inconsistencies between a filter which indicates whether <SUBJECT> was taught this year and the number of hours formally scheduled for teaching it. (Note: <SUBJECT> can be mathematics or science.)

If the filter indicates that NO <SUBJECT> is taught (2), but time is officially scheduled to teaching it, a warning has been given, but no corrections have been undertaken.

2521 (TG): This cleaning step identifies inconsistencies between a filter which indicates whether <SUBJECT> was taught this year and the type of questionnaire which was administered for <SUBJECT>. (Note: <SUBJECT> can be mathematics or science.)

If the filter indicates that NO <SUBJECT> is taught (2), but a corresponding questionnaire is completed, a warning has been given, but no corrections have been undertaken.

2522 (TG): This cleaning step identifies inconsistencies between two filters which indicate that neither mathematics nor science is taught this year.

If both filters indicate that NO MAT and NO SCI is taught, a warning has been given, but no corrections have been undertaken.

2503 (TG): This cleaning step identifies inconsistencies between a filter which indicates that no <SUBJECT> is taught this year and dependent variables which indicate at which grade levels it is taught. Also, the identification variable IDSUBJCT has been included.(Note: <SUBJECT> can be mathematics or science.)

If the filter indicates that NO <SUBJECT> is taught (2), and NO grade level variables indicate that it is taught, but IDSUBJCT indicates that it is taught in the selected class, a warning has been given, but no corrections have been undertaken.

2531 (TG): This cleaning step identifies inconsistencies between a filter which indicates whether <SUBJECT> is taught this year and dependent variables which indicate at which grade levels <SUBJECT> is taught. (Note: <SUBJECT> can be mathematics or science.)

If the filter indicates that NO <SUBJECT> is taught (2), and the dependent variables were coded to 'Missing' (9) or 'Not taught' (2), the dependent variable has been recoded to 'Logically not applicable' (B).

2504 (TG): This cleaning step identifies inconsistencies between a variable which indicates whether a textbook is used in <SUBJECT> and dependent variables which indicate which textbooks are used. (Note: <SUBJECT> can be mathematics or science.)

If it is indicated that NO textbook is used (2), but textbooks are checked in the list or a name is entered, then the filter has been recoded to 'Yes' (1).

2505 (TG): This cleaning step identifies inconsistencies between a filter variable which indicates whether a textbook is used in<SUBJECT> and dependent variables which indicate which textbooks are used.(Note: <SUBJECT> can be mathematics or science.)

If the filter indicates that NO textbook is used (2) and dependent variables were coded to 'Missing', then dependent variables have been recoded to 'Logically not applicable'.

2506 (TG): This cleaning step identifies inconsistencies between a filter which indicates whether a textbook is used in <SUBJECT>and a dependent variable which indicates whether teaching is based on a textbook. (Note: <SUBJECT> can be mathematics or science.)

If it is indicated that NO textbook is used, but teaching is based on textbooks, then a warning has been given, but no corrections have been undertaken.

2507 (TG): This cleaning step identifies inconsistencies between a filter variable which indicates whether a textbook is used in<SUBJECT> and a dependent variable which indicates whether teaching is based on a textbook. (Note: <SUBJECT> can be mathematics or science.)

If it is indicated that NO textbook is used and teaching is not based on textbooks (Missing), then 'Missing' has been recoded to 'Logically not applicable' (B).

2508 (TG): This cleaning step identifies inconsistencies between a filter which indicates whether homework was given in the recent <CLASS> and a dependent variable which indicates the time necessary to do the homework. (Note: <CLASS> can be the mathematics class or science class.)

If the filter indicates that NO homework was assigned (2), but a time was entered, the filter has been recoded to 'Yes' (1).

2509 (TG): This cleaning step identifies inconsistencies between a filter variable which indicates whether homework was given in the recent <CLASS> and a dependent variable which indicates the time necessary to do the homework. (Note: <CLASS> can be the mathematics class or the science class.)

If the filter indicates that NO homework was assigned (2) and the dependent variable is coded to 'Missing' (999), the dependent variable has been recoded to 'Logically not applicable' (996).

2510 (TG): This cleaning step identifies inconsistencies in a filter indicating whether or not a topic was taught and two dependent sets of variables ('Yes list' and 'No list').

If both lists are checked and the filter is set to 'No' (2), a warning has been given, but no correction has been undertaken.

If the 'Yes list' is checked, but not the 'No list', and the filter is set to 'No' (2), the filter has been recoded to 'Yes' (1).

If the 'No list' is checked, but not the 'Yes list', and the filter is set to 'Yes' (1), the filter has been recoded to 'No' (2).

2511 (TG): This cleaning step identifies inconsistencies in a filter indicating whether or not a topic was taught and two dependent sets of variables ('Yes list' and 'No list').

If both lists are checked and the filter is set to 'Yes' (1), a warning has been given, but no correction has been undertaken.

2512 (TG): This cleaning step identifies inconsistencies in a filter indicating whether or not a topic was taught and two dependent sets of variables ('Yes list' and 'No list').

If only one list is checked and the filter is consistent with the list checked, but variables in the other list are coded to 'Missing' (9), 'Missing' has been recoded to 'Logically not applicable' (B).

2513 (TG): This cleaning step identifies inconsistencies in a filter indicating whether or not science is taught mainly as a separate subject and two dependent variables indicating the time science is taught (one variable if the filter states that science is taught as a separate subject and one variable if the filter states that science is not taught as a separate subject).

If both times have been assigned, a warning is given, but no correction has been undertaken.

If the 'Yes'-time is assigned, and the filter is set to 'No' (2), the filter has been recoded to 'Yes' (1)

If the 'No'-time is assigned, and the filter is set to 'Yes' (1), the filter has been recoded to 'No' (2).

2514 (TG): This cleaning step identifies inconsistencies in a filter indicating whether or not science is taught mainly as a separate subject and two dependent variables indicating the time science is taught (one variable if the filter states that science is taught as a separate subject and one variable if the filter states that science is not taught as a separate subject)

If only one time is assigned and the filter is consistent with the assigned time, but the second time is coded to 'Missing' (999) or 0, the 'Missing' or 0 has been recoded to 'Logically not applicable' (996).

SCHOOL FILTER VARIABLE CHECK

This cleaning steps identifies all problems between filter variables and dependent variables in the School File. In a first cleaning step, the filter variable has been made consistent with the dependent variable(s). In a second cleaning step, all dependent variables coded to 'Missing' have been recoded to 'Logically not applicable' if the filter variable was coded negatively.

2601 (CG) This cleaning step identifies all cases in which the instructional time is the same for both the upper grade and the lower grade.

If the instructional time is the same for both grades, but different times were entered for the lower and the upper grade, the filter (A/BCBGINST) has been recoded to 'Not checked' (1).

If the instructional time is not the same for both grades according to the filter variable, but all times were entered identically for both grades, the filter (A/BCBGINST) has been recoded to 'Checked' (2).

2602 (CG): This cleaning step identifies all cases in which the school week is not divided into instructional periods according to the filter variable, but the dependent questions have been answered.

The filter variable A/BCBGDIVI has been recoded to 'Yes' (1).

2603 (CG): This cleaning step identifies all cases in which the school week is not divided into instructional periods according to the filter variable, but the dependent variables have been coded to 'Missing' instead of 'Not applicable'.

The dependent questions have been recoded to 'Logically not applicable'.

2604 (CG): This cleaning step identifies problems with filter and dependent questions for the remedial teaching of math or science and the special enriched teaching of math or science. The following filter questions and dependent variables have been checked: (1) variables on teaching remedial math: A/BCBMRMDL <-> A/BCBMRMD1-4; (2) variables on teaching remedial science: A/BCBSRMDL <-> A/BCBSRMD1-4; (3) variables on special enrichment activities in math: A/BCBMENRH <-> A/BCBMENR1-4;(4) variables on special enrichment activities in science: A/BCBSENRH <-> A/BCBSENR1-4.

If at least one dependent question was answered with 'Yes' (1), but the filter question was answered with 'No' (2), the filter question has been recoded to 'Yes' (1.)

If at least one dependent question was answered with 'No' (2) and none was answered with 'Yes' (1) and the filter question has been answered with 'No' (2), the dependent questions have been recoded to 'Logically not applicable' (B).

2605 (CG): This cleaning step identifies problems with filter and dependent questions for the upper grade courses in math or science. The listed filter and dependent variables have been checked: (1) variables on upper grade courses in mathematics: A/BCBMUSCO <-> A/BCBMUC1-62,A/BCBMUFC1-8; (2) variables on upper grade courses in science: A/BCBSUSCO <-> A/BCBSUC1-62,A/BCBSUFC1-8.

If the filter variable (A/BCMUSCO A/BCSUSCO) was coded to 'All students take the same courses' (1), but questions answered indicate that students take different courses, the filter variable has been recoded to 'Different courses' (2).

If the filter variable was coded to 'Different courses' (2), but the dependent variables indicate that students take the same courses, the filter variable has been recoded to 'All students take the same courses' (1).

Answers given for both blocks of dependent questions have been flagged.

2606 (CG): This cleaning step identifies all cases in which dependent questions were incorrectly coded as 'Missing' (9), but the filter question indicates that the questions should not be answered. The listed filter

and dependent variables have been checked: (1) variables on teaching remedial math: A/BCBMRMDL <-> A/BCBMRMD1-4; (2) variables on teaching remedial science: A/BCBSRMDL <-> A/BCBSRMD1-4; (3) variables on special enrichment activities in math: A/BCBMENRH <-> A/BCBMENR1-4; (4) variables on special enrichment activities in science: A/BCBSENRH <-> A/BCBSENR1-4; (5) variables on upper grade courses in mathematics: A/BCBMUSCO <-> A/BCBMUC1-62,A/BCBMUFC1-8; (6) variables on upper grade courses in science: A/BCBSUSCO <-> A/BCBSUC1-62,A/BCBSUFC1-8.

Incorrectly coded dependent variables have been recoded to 'Logically not applicable' (B).

INCONSISTENCY CHECK-STUDENT FILE

This set of cleaning steps identifies all problems between data variables.

3101 (SG): This cleaning step identifies all problems with date of testing. Different or missing testing dates within a class have been detected.

If the year of testing is missing for the whole class, the year has been recoded to 1994 for countries in the southern hemisphere and to 1995 for countries in the northern hemisphere.

If the month of testing is missing for the whole class, a warning has been given, but no correction has been undertaken.

If the testing dates are missing for single students, they have been replaced by the values found for the other students in the class (if they are otherwise consistent).

If the date of testing differs for students within a class, a warning has been given, but no corrections have been undertaken.

3102 (SG): This cleaning step identifies problems with 'Date of birth' obtained from tracking information and questionnaire information.

If both tracking and questionnaire information are missing, a warning has been given, but no corrections have been undertaken.

3103 (SG): This cleaning step identifies problems with 'Date of birth' obtained from tracking information and questionnaire information. 'Dates of birth' which are 'Incomplete' but not 'Missing' in the questionnaire information and are 'Incomplete' or 'Missing' in the tracking information have also been reported.

If tracking information is available and questionnaire information is missing, then the questionnaire variables have been recoded to the tracking information.

If tracking information is available, but is different from questionnaire information (not missing), a warning has been given, but no corrections have been undertaken.

3104 (SG): This cleaning step identifies all problems with student's sex obtained from tracking information and from the *Student Questionnaires.*

If tracking information is available and questionnaire information is missing, the questionnaire information has been replaced by the tracking information and vice versa.

If the information is different in both sources (and not missing), a warning has been given, but no corrections have been undertaken.

3105 (SG) This cleaning step identifies all problems between 'Age of student' and 'Age when student came to country'. (Note: this check requires the following checks: (I) Check of consistency in 'Age of student' obtained from tracking information and questionnaire information;(II) Check of consistency in 'Date of testing'. If 'Age of student' can not be computed from the data because 'Date of test' is not available, '94 is used in southern hemisphere countries and '95 in northern hemisphere countries.

If the student is younger than indicated in 'Age when student came to country', then 'Age when student came to country' has been coded to 'Invalid' (97).

3106 (SG): This cleaning step identifies inconsistencies between 'Number of people living at home' and the question block 'Do each of these people live at home with you most or all of the time?' Students where the number of questions answered with 'Yes' exceeds the 'Number of people living at home' by one have been identified and counted. It is assumed, that they forgot to include themselves in the number of people living at home.

A warning has been given, but no corrections have been undertaken.

3107 (SG): This cleaning step identifies inconsistencies between 'Number of people living at home' and the question block 'Do each of these people live at home with you most or all of the time?' Students where the number of questions answered with 'Yes' exceeds the 'Number of people living at home' by more than one have been identified and counted.

A warning has been given, but no corrections have been undertaken.

3108 (SG): This cleaning step identifies all problems with <SUBJECT>: 'The teacher gives us homework' and the dependent questions concerning homework in <SUBJECT> lessons.(Note:<MAT> refers to mathematics lessons <SCI> refers to science lessons)

If at least one variable (but not all) concerning 'Homework in <SUBJECT> lessons' is coded differently from 'Never' (1 or 2), but the variable concerning 'Homework given in <SUBJECT> lessons' is coded to 'Never' (3), then a warning has been given, but no corrections have been undertaken.

3109 (SG) This cleaning step identifies all problems with <SUBJECT> 'The teacher gives us homework' and the dependent questions concerning homework in <SUBJECT> lessons. (Note:<MAT> refers to mathematics lessons <SCI> refers to science lessons)

If all variables concerning 'Homework in <SUBJECT> lessons' are coded differently from 'Never' (1 or 2), but the variable concerning 'Homework given in <SUBJECT> lessons' is coded to 'Never' (3), then the variable concerning 'Homework given in <SUBJECT> lessons' has been coded to 'Invalid' (I).

3110 (SG) This cleaning step identifies all problems with <SUBJECT>: 'The teacher gives us homework' and the dependent questions concerning homework in <SUBJECT> lessons. (Note:<MAT> refers to mathematics lessons <SCI> refers to science lessons)

If at least one (but not all) variable concerning 'Homework in <SUBJECT> lessons' is coded differently from 'Never' (1, 2, or 3), but the variable concerning 'Homework given in <SUBJECT> lessons' is coded to 'Never' (4), a warning has been given, but no corrections have been undertaken.

3111 (SG) This cleaning step identifies all inconsistency problems for <SUBJECT> 'The teacher gives us homework' and the dependent questions concerning homework in <SUBJECT> lessons. (Note:<MAT> refers to mathematics lessons <SCI> refers to science lessons).

If all variables concerning 'Homework in <SUBJECT> lessons' are coded differently from 'Never' (1, 2, or 3), but the variable concerning 'Homework given in <SUBJECT> lessons' is coded to 'Never' (4), then the variables concerning 'Homework given in <SUBJECT> lessons' has been coded to 'Invalid' (I).

3112 (SG) This cleaning step reports all inconsistencies between students' responses to 'Work in <SUBJECT> lessons in small groups' and 'Begin a new topic in <SUBJECT> by working in small groups'. Note: <MAT>

refers to mathematics lessons <SCI> refers to science lessons <BIO> refers to biology lessons <CHE> refers to chemistry lessons <EAR> refers to earth science lessons <PHY> refers to physics lessons

If group work never happens in <SUBJECT> (4), but group work is used when students begin a new topic in <SUBJECT>, a warning has been given, but no corrections have been undertaken.

3141 (SG) This cleaning step identifies problems with numbers of girls and boys within classes.

If the number of boys or girls is less than 10 percent of the number of all students in class under consideration, a warning has been given, but no changes have been undertaken.

INCONSISTENCY CHECK-TEACHER FILE

This set of cleaning steps identifies problems between data variables.

3201 (TG): This cleaning step identifies all inconsistencies in 'Number of boys in <CLASS>' and 'Number of girls in <CLASS>' (Note: <CLASS> can be the mathematics class or the science class. A/BTBMBOY and A/BTBMGIRL refer to the mathematics class. BTBSBOY and BTBSGIRL refer to the science class.)

If both variables were coded to 0, then both have been recoded to 'Invalid'.

If one variable was coded to 0 and the other was coded to 'Missing', a warning has been given, but no corrections have been undertaken.

3202 (TG): This cleaning step identifies inconsistencies in the variable block concerning different achievement levels in the<CLASS>. If the sum of all percentages exceeds 110, a warning is given. (Note: <CLASS> can be the mathematics class or the science class.)

If the sum of all percentages exceeds 110, a warning has been given, but no corrections have been undertaken.

3203 (TG): This cleaning step identifies inconsistencies in the lists concerning <SUBJECT> topics. Teachers were asked to indicate the number of periods a topic has been taught (<VAR>). They also should have indicated whether they will begin or continue teaching the topic (<VAR_A>), whether or not the topic is taught this year (<VAR_B>), and whether the topic was taught in a previous year (<VAR_C>). (Note: <SUBJECT> can be mathematics or science.)

If no periods were assigned to the topic (<VAR> 'Missing' (9)) and all other variables concerning the topic were coded to 'Not checked' (1), all 'Not checked' (1) have been recoded to 'Missing' (9).

If at least one variable was coded to 'Checked' (2), all variables coded to 'Missing' (9) have been recoded to 'Not checked' (1).

If the topic was taught a number of periods (<VAR> < 5), but it is indicated that the topic is not taught this year, (<VAR_B> 'Checked' (2)), <VAR_B> has been recoded to 'Not checked' (1).

If it is indicated that the topic is not taught this year (<VAR_B> 'Checked' (2)) and also that teaching the topic will be continued or begun (<VAR_A> 'Checked' (2)), then a warning has been given, but no corrections have been undertaken.

3231 (TG): This cleaning step identifies inconsistencies in the lists concerning <SUBJECT> topics. Teachers were asked to indicate the number of periods a topic has been taught (<VAR>). They also should have indicated whether they will begin or continue teaching the topic (<VAR_A>), whether or not the topic is taught this year (<VAR_B>), and whether the topic was taught in a previous year (<VAR_C>)

If the value for time assigned to the topic (<VAR>) was coded to 'Not administered' (8) and all other variables concerning the topic were coded to 'Not checked' (1), 'Checked' (2), or 'Missing' (9), the value for time assigned to the topic (<VAR>) has been recoded to 'Missing' (9).

3232 (TG): This cleaning step identifies inconsistencies in the lists concerning <SUBJECT> topics. Teachers were asked to indicate the number of periods a topic has been taught (<VAR>). They also should have indicated whether they will begin or continue teaching the topic (<VAR_A>), whether or not the topic is taught this year (<VAR_B>), and whether the topic was taught in a previous year (<VAR_C>). (Note: <SUBJECT> can be mathematics or science.)

If one (but not all) of the variables concerning the topic has been coded 'Not administered' (8), a warning has been given, but no corrections have been undertaken.

3204 (TG): This cleaning step identifies inconsistencies in the lists concerning <SUBJECT> topics. Teachers were asked to indicate the number of periods a topic has been taught. They also should have indicated whether they will begin or continue teaching the topic, if the topic is not taught, and if the topic was taught in a previous year. Note: <SUBJECT> can be mathematics or science.

If the respondent indicates that the topic was taught a number of periods, but all other variables concerning the topic were coded to 'Missing' (9), all 'Missings' have been recoded to 'Not checked' (1).

3205 (TG): This cleaning step identifies all inconsistencies between the sum of minutes assigned to teaching activities in <CLASS> and the duration of the <CLASS>.(Note: <CLASS> can be either the mathematics class or the science class).

If the total time assigned to activities exceeds the time available, a warning has been given, but no corrections have been undertaken.

3206 (TG): This cleaning step identifies all inconsistencies between order of teaching activities in <CLASS> and times assigned to teaching activities in <CLASS>. (Note: <CLASS> can be either the mathematics class or the science class).

If an activity was put in order, but 0 minutes were assigned to carrying out the activity, the time has been recoded to 'Invalid'.

If an activity was not put in order (0), but time was assigned to carrying out the activity, the order has been recoded to 'Invalid'.

3207 (TG): This cleaning step identifies inconsistencies in the variable blocks concerning the frequency with which different tasks are given as homework in <CLASS>.(Note: <CLASS> can be the mathematics class or the science class).

If the respondent indicates that more than zero tasks are assigned 'Rarely' to 'Always' and homework is never assigned for more than two tasks, a warning has been given, but no corrections have been undertaken.

If the respondent indicates for fewer than three tasks that they are assigned 'Never' (1) and for all other tasks that he/she does not assign homework and the variable indicating how often homework is assigned is coded to 'Never' (1), then 'Never' has been recoded to 'I do not assign homework' (5).

INCONSISTENCY CHECK-SCHOOL FILE

This set of cleaning steps identifies inconsistency problems between data variables.

3301 (CG): This cleaning step identifies all inconsistency problems with the grade levels found in a school.

If no grades can be found at a school and all grades are coded to 'No' (2), all variables have been recoded to 'Invalid' (I).

If 'Yes' (1) is not coded for any grade, but 'Missing' is found for at least one grade, a warning has been given, but no corrections have been undertaken.

3302 (CG): This cleaning step identifies all schools in which the grades in the school are not sequential.
A warning has been given, but no corrections have been undertaken.

3303 (CG): This cleaning step identifies all problems with the number of the classroom teachers indicated in Q. 3 & 4.
If the number of individual full-time classroom teachers is larger than the total number of full-time equivalent classroom teachers the observation is flagged, but no corrections have been undertaken
If the number of full-time classroom teachers and the number of part-time classroom teachers equals zero, a warning has been given, but no corrections have been undertaken.

3304 (CG): This cleaning step identifies all problems with the 'Percentage of the classroom teachers teach '. (Q.6).
If the sum of the percentage of teachers who teach more than three-quarters math (A/BCBMTEAC) and the percentage of teachers who teach no math (A/BCBMNONE) exceeds 100, a warning has been given, but no corrections have been undertaken.
If the sum of the percentage of teachers who teach more than three-quarters science (A/BCBSTEAC) and the percentage of teachers who teach no science (A/BCBSMNONE) exceeds 100, a warning has been given, but no corrections have been undertaken.
If the sum of the percentage of teachers who teach more than three-quarters math and science (A/BCBGTEAC) and the percentage of teachers who teach neither math nor science (A/BCBGNONE) exceeds 100, a warning has been given, but no corrections have been undertaken.

3305 (CG): This cleaning step identifies all problems with the total time the principal spends on activities (A/BCBGAC01-14).
If the sum of A/BCBGAC01-14 exceeds 280 hours per month, a warning has been given, but no corrections have been undertaken.
If the sum of A/BCBGAC01-14 equals zero, a warning has been given, but no corrections have been undertaken.

3306 (CG): This cleaning step identifies all problems with the question 'In your school, how many computers are ...'
If the total number of computers available for teachers and students (A/BCBGCOM1) is smaller than at least one of the numbers of computers used by teacher and students indicated in A/BCBGCOM2-4, a warning has been given, but no corrections have been undertaken.

3307 (CG): The listed observations have the sum of girls and boys for the whole school (A/BCBGGENR A/BCBGBENR), for the lower grade(A/BCBGLGER A/BCBGLBER) or for the upper grade (A/BCBGUGER A/BCBGUBER) equal to 0. The cases in which no children are enrolled in one target grade could be correct if the school is a lower grade or upper grade school only.
If the sum of boys and girls enrolled in a school equals zero, the number of boys and the number of girls enrolled have been recoded to 'Invalid' (9997).
If the sum of boys and girls enrolled in both target grades equals zero, the number of boys and the number of girls enrolled have been recoded to 'Invalid' (997) for both target grades.
If the sum of boys and girls enrolled in lower grades or in upper grade equals zero, but enrolled students are indicated for at least one target grade, the school has been flagged.

3308 (CG): This cleaning step identifies all problems with the number of girls/boys studying math or science compared to the number of girls/boys enrolled in the grade. It also identifies all problems with the number of girls/boys repeating a grade compared to the number of girls/boys enrolled in the grade.

   If the number of students repeating a grade, studying math, or studying science is larger than the number of students enrolled in the grade, a warning has been given, but no corrections have been undertaken.

3309 (CG): This cleaning step identifies all cases in which the instructional time is the same for both the upper grade and the lower grades, and the times were incorrectly entered into the column for the lower grade instead of the column for the upper grade.

   All values have been transcribed to the upper grade variables.

3310 (CG): This cleaning step identifies problems with the instructional times.

   If the number of instructional days per school year is 0 for both target grades the corresponding variables A/BCBGLDYY and A/BCBGUDYY have been recoded to 'Invalid'.

   If the number of total hours per week is 0 for both target grades the corresponding variables A/BCBGLTHW and A/BCBGUTHW have been recoded to 'Invalid'.

   If the number of instructional hours per week is 0 for both target grades the corresponding variables A/BCBGLIHW and A/BCBGUIHW have been recoded to 'Invalid'. If A/BCBGINST indicates that the instructional time is the same for both grades, only the upper grade variables have been recoded.

3311 (CG) This cleaning step identifies all cases in which the sum of the percentage of students in the most advanced course and the least advanced course exceeds 110. The following variables have been checked:(1) most and least advanced course in math (A/BCBMUC41 A/BCBMUC42); (2) most and least advanced course in science (A/BCBMUC41 A/BCBMUC42).

A warning has been given, but no corrections have been undertaken.