

Chapter 8

Quality Control During Data Collection: Refining for Rigor



Lauren Musu, Sandra Dohr, and Andrea Netten

Abstract Studies undertaken under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) use rigorous quality control processes to help ensure high quality data and facilitate cross-country comparisons of study results. This chapter provides an overview of the quality control processes implemented during data collection including the production of detailed manuals to standardize data collection procedures and the monitoring of how these procedures are implemented through national and international quality control programs. National and international quality control procedures help to confirm the validity of the data by monitoring data collection efforts. National quality control programs are carried out by participating countries with specific guidance provided by IEA and the international study centers (ISCs). International quality control monitoring ensures that sampling procedures are followed at participating school, classroom, and student levels, monitors on-site data collection to check whether the test administration procedures and security guidelines set by IEA and the ISCs are met, and provides information on circumstances that occurred during data collection that could influence the data reliability and quality. This chapter provides a history of the development of these different quality control procedures and information on their implementation within IEA studies. The chapter concludes with a series of recommendations for potential improvements to consider for the future.

Keywords Data collection · Large-scale assessment · Quality assurance · Quality control · Test administration

L. Musu (✉) · S. Dohr · A. Netten

International Association for the Evaluation of Educational Achievement (IEA), Amsterdam, The Netherlands

e-mail: l.musu@iea.nl

S. Dohr

e-mail: s.dohr@iea.nl

A. Netten

e-mail: a.netten@iea.nl

8.1 Introduction

Quality control is an essential component of international large-scale assessment. Rigorous quality control processes help ensure high quality data and facilitate cross-country comparisons of study results. Although quality control can and does exist at all stages, including before, during, and after data collection, the term is often used within IEA studies to refer to the monitoring of data collection procedures in schools and classrooms, especially at the international level. Quality control procedures during data collection serve several purposes:

- to standardize data collection by providing detailed procedural manuals for countries to follow;
- to check on-site whether the test administration procedures and security guidelines set by IEA and the international study centers (ISCs) and outlined in these manuals are followed;
- to ensure that sampling procedures are adhered to at participating school, classroom and student levels; and
- to provide information on any circumstances that occurred during data collection that could influence the data quality.

Deviations from the standardized procedures outlined in the manuals are a threat to reliability and validity. These procedures are put in place to ensure data collection occurs in a comparable way across countries with limited disruptions to the process. Therefore, national and international quality control procedures help to confirm the validity of the data by monitoring data collection efforts and ensuring appropriate participation from the sampled schools, classrooms, and students.

The terminology used to describe the process for monitoring data quality varies in the research literature, but quality control and quality assurance are the most commonly used labels. Some studies use these terms interchangeably, but they often refer to slightly different aspects of the processes for ensuring data are valid and reliable. Within IEA studies, as well as in other large-scale research, quality control procedures are often part of a larger quality assurance program. Quality assurance generally refers to the full spectrum of procedures that are implemented while quality control refers to the measurement of certain characteristics of data collection procedures to ensure that certain standards are met (Statistics Canada 2010). Various aspects of quality control and assurance are used throughout IEA studies. For example, response and range checks are used during data processing to look for implausible values or evidence of data tampering. These quality control measures are covered in other chapters, but this chapter focuses on quality control materials and procedures for test administration, that is, when the assessments and surveys are administered in the sampled schools and classrooms.

Quality control procedures for data collection and test administration consists of three major components:

- (1) Production and distribution of standardized manuals managed by IEA and the ISC;
- (2) National quality control procedures and monitoring managed by participating countries; and
- (3) International quality control procedures and monitoring managed by IEA.

There are three major aspects of quality control during test administration, with various tasks and responsibilities for the organizations and entities involved in the data collection (Table 8.1).

All three elements of quality control during data collection play an important role in ensuring that countries follow standardized procedures and monitoring whether those procedures are implemented in a uniform way across countries. In describing ideal quality control procedures for large-scale assessment, Cresswell (2017) noted that quality management during data collection should include “the development and agreement of standardized implementation procedures, the production of manuals which reflect agreed procedures, the recruitment and training of personnel in administration and organization—especially the test administrator and the school coordinator, and the quality monitoring processes—recruiting and training quality monitors to visit schools and national centers” (p. 195). IEA studies follow these best practices with the use of collaboratively developed manuals, and national and international quality monitoring programs.

Table 8.1 Roles and responsibilities for quality control procedures during test administration

Procedure	IEA and the ISC	Participating countries
Manuals	IEA and the ISC develop standardized manuals to detail data collection procedures	Countries are responsible for translating and adapting the provided manuals to necessary languages and their national context
National quality control	IEA and the ISC develop and provide manuals for use in national quality control procedures	Countries have primary responsibility for implementing a national quality control program using guidelines and recommendations provided by IEA and the ISC
International quality control	Materials and training for IQCMs are developed collaboratively by IEA and the ISC. IQCMs solely report to IEA and the ISC	Countries coordinate with the IQCM to provide requested materials and select schools for monitoring visits

Notes ISC = international study center; IQCM = international quality control monitor

8.2 Manuals

To help ensure that national centers, schools, and test administrators are familiar with the required procedures for data collection, detailed manuals are developed and distributed to countries. The use of manuals is one of the earliest procedures implemented during IEA studies to ensure standardization across countries in the data collection. This section describes the development of these manuals and their implementation.

8.2.1 *Early Quality Control Procedures and the Development of Manuals*

The first IEA studies were conducted by a team of researchers in different countries who met at regular intervals to plan and implement the studies. The Pilot Twelve-Country Study, which took place in 1960 (IEA 2020a) is considered the first IEA study. The study was sponsored by the UNESCO Institute for Education and many of the founding members of IEA (e.g., Douglas Pidgeon, Benjamin Bloom, Robert Thorndike, and Torsten Husén) took part in the design, implementation, and analysis of the study. While the Pilot Twelve-Country Study was a milestone in the development of international large-scale assessment in general, there is little mention of quality control procedures in the reporting of the study results. In fact, the authors cautioned against overinterpretation of the results in the report and chose not to show total scores in a way that facilitated comparison between countries due to concerns about data comparability (Foshay et al. 1962).

Standardized procedures for data collection were developed in the form of detailed procedural manuals beginning with the next IEA study, the First International Mathematics Study in 1964 (IEA 2020a). Publications for this study also acknowledged the importance of standardization when describing the administration of the data collection procedures. “It was extremely important to ensure that as far as possible uniform methods of procedure were employed in the testing programme in all countries” (Postlethwaite 1967, p. 46). To accomplish this, a small committee of the researchers involved in organizing the study developed three manuals: one for the national centers, one for individuals coordinating data administration at each school, and one for the individuals administering the test. Additional information and instructions were also sent to participating countries in the form of circular letters and bulletins.

Similar to the First International Mathematics Study, individuals involved in planning the different assessments for the Six Subject Survey in 1970–1971 (IEA 2020a) produced detailed manuals that were distributed to the national centers, the school coordinators, and the test administrators. While closely scripted administrative procedures were described in the manuals and other documentation was provided to schools, oversight of the assessment administration was left to the discretion of

the participating schools under the assumption that the procedures outlined in these manuals were being followed.

Subsequent studies continued to produce detailed manuals and the content and structure of the manuals has expanded over time. These manuals form the backbone of IEA data collections in that they carefully explain the data collection and survey administration procedures that should occur within schools and classrooms.

8.2.2 Current Implementation of Manuals

Over time, IEA and the ISCs have developed increasingly detailed manuals for use by participating countries. These manuals include survey operations procedures (SOP) manuals, a school coordinator manual, a test administrator manual, and manuals for national and international quality control monitors. The manuals detail procedures for the test administration and data collection that have been agreed upon by IEA, the ISC, and the national research coordinators (NRCs). All manuals provided to national study centers are in the English language but can be translated to a national language as needed by the national study centers themselves.

SOP manuals outline the process of data collection from beginning to end. Each manual details a specific part of the study process such as sampling, preparing assessment instruments, and scoring items after the assessments have been completed. Manuals for school coordinators, test administrators, and national quality control monitors are included as supplementary materials with relevant sections of the SOP. To guide NRCs through the process, SOPs are released on a staggered basis to coincide with major data collection milestones.

School coordinators are responsible for ensuring that sampled classes, teachers, and students actually participate in the assessment. They also oversee the distribution, completion, and collection of testing materials and questionnaires. The manuals for school coordinators provide detailed instructions on the ways in which these tasks should be completed, allowing for some individual variation between countries due to contextual factors such as confidentiality laws. The manuals include extensive details on the role of the school coordinators, including completion of class and student listing forms and tasks for securing materials prior to testing, distributing them on the testing day, and returning them to the national center after testing is complete.

Test administrators are responsible for administering the assessments. Test administrators must ensure that each student receives their specific testing materials and that the assessments are given in a standardized way across countries. This includes following a specific script with instructions for students taking the assessment.

Manuals for national and international quality control monitors describe the roles and responsibilities of those positions. The manuals for national quality control monitors include a description of the roles and responsibilities and sample classroom observation forms that can be used during school visits. Manuals for

international quality control monitors also include a description of roles and responsibilities and observation forms. In addition, the manuals for the school coordinators and test administrators are provided so that international quality control monitors can ensure that these individuals are adhering to procedures when they do their classroom observations.

Since countries have varying degrees of familiarity with administering large-scale assessments, the different manuals are designed to provide all the details necessary to carry out the data collection procedures. All countries are asked to follow the procedures detailed in the manuals without significant deviation to ensure consistency, and training is provided for NRCs so that they understand the structure and content of the manuals and the procedures contained therein. National and international quality control monitoring also help to ensure that the procedures described in the manuals are carried out as specified.

8.3 National Quality Control Procedures

IEA and the ISC recommend that countries implement a national quality control program in order to monitor data collection efforts. Countries also want to monitor the quality of their data collection efforts so that they can intervene when problems are discovered and so that they have confidence in the data collected within their country. National quality control programs were developed individually in some countries before the international quality control program existed, but recent studies base guidelines for national quality control on the international program. Although similar in purpose and scope, the national quality control monitoring program and international quality control monitoring programs are designed to be separate but complementary to one another. For example, for IEA's Trends in International Mathematics and Science Study (TIMSS) 2015, NRCs were required to send national quality control monitors to a 10% sample of the schools to observe the test administration and document compliance with prescribed procedures. These site visits were in addition to the visits to 15 schools conducted by the international quality control monitors (Johansone and Wry 2016).

8.3.1 Development of National Quality Control Procedures

As studies became increasingly large and more complex, oversight of studies at the international level was helped by the establishment of ISCs and management at the national level was helped by the appointment of NRCs. The Second International Mathematics Study of 1980–1982 (IEA 2020a) was the first IEA study to explicitly mention the appointment of NRCs, then referred to as national technical officers, in each country (Garden 1990). While earlier studies used national study centers to

help coordinate the data collection, there was not always one individual person coordinating the work at the national level. NRCs were usually trained in how to perform their duties. In recalling the evolution of IEA, Alan Purves, Chair of IEA from 1986 to 1990, explained “there was on-the-job training for the national technical officers, as they were called. Usually [experts from IEA] visited each of the centers for several days” (Purves 2011, p. 546). Such training focused on the implementation of data collection procedures at the national level. However, training was not a requirement and was generally not given in a standardized way to all NRCs.

The use of international study centers (ISCs) and NRCs continued with the Second International Science Study in 1983–1984 (IEA 2020a) and the Reading Literacy Study in 1990–1991 (IEA 2020b). While international quality control was still lacking, some countries were implementing stricter independent quality control procedures for data collection at the national level. For example, for the Reading Literacy Study, the United States (US) chose to hire field staff with no associations with the schools themselves to serve as test administrators. This allowed the coordinating center to train the field staff and thus try to ensure more standardized procedures. As stated in the US technical report, “It was felt that data collected in this way would be far more comparable than that collected under an infinite number of differing conditions” (Binkley and Rust 1994, p. 41). While these procedures led to increased confidence that data were collected in a standardized way, they were also admittedly costly and were only implemented in this exact way in the US. Other countries implemented their own quality control procedures, but there were no checks implemented across countries to ensure that the standardized procedures were being followed.

TIMSS 1995 was the first study that explicitly laid out recommended procedures for quality control at the national level (Martin et al. 1996b). The recommendations for national-level procedures closely mirrored those that were being implemented at the international level during this same study. It was recommended that NRCs arrange for quality control observers to visit a sample of schools on the day of testing. To help facilitate this, IEA and the ISC developed a manual and accompanying forms based on the international materials that could be adapted for use at the national level. While NRCs could implement their own procedures for national quality control, they were encouraged to use the materials provided.

8.3.2 Implementation of National Quality Control Procedures

As part of the materials provided for participation in an IEA study, NRCs are given detailed information on how they can implement a national quality control program. Similar to the procedures in TIMSS 1995, IEA and the ISC produce detailed manuals on how to implement a national quality control program that will complement the international program. For example, TIMSS 2015 instructed NRCs to send national

quality control monitors to observe the test administration and document whether required procedures were followed in 10% of participating schools (Johansone and Wry 2016).

These national quality control monitor manuals are the primary resource provided by IEA and the ISC for national quality control. They are designed to train quality control monitors to observe test administration procedures in their country. For the most part, countries use the national quality control monitor manuals, but they are given flexibility in the best way to implement the program to meet the needs of their country. Some countries choose to implement altered national quality control procedures and sometimes countries are unable to implement the program as prescribed due to lack of funding. For example, in TIMSS 2019, one country with centrally trained test administrators who were totally independent of the sampled schools felt it was sufficient to observe a smaller percentage of these administrators in the field. Despite some difficulties or changes to procedures, the majority of countries implement national quality control procedures as specified in documentation provided by IEA and the ISC.

In addition to supporting quality control monitoring procedures at the national level, IEA and the ISC support standardized procedures at the national level by providing both online training and direct presentations to NRCs on appropriate procedures. As part of this training, the detailed manuals for test administrators and school coordinators described earlier in the chapter are provided and discussed with NRCs.

At the end of the data collection and submission process, NRCs are required to provide a summary report to IEA and the ISC describing their national quality control activities. In addition, NRCs provide feedback to IEA and the ISC through the survey activities questionnaire (SAQ). This questionnaire is meant to document study procedures at the national level, from sampling all the way through submitting the final data. NRCs were initially asked questions from the SAQ during a structured interview with the international quality control monitor (IQCM). In recent years, the SAQ has always been distributed electronically to NRCs by the ISC once all the data from a country has been received.

The purpose of the SAQ is to gather information from the NRC and other national center staff on how well procedures and materials worked and what can be improved in the future. The SAQ asks about sampling procedures and manuals and includes questions on contacting and recruiting schools, focusing on how schools were contacted and how school coordinators were trained. Subsequent sections of the SAQ include questions about how assessment materials were adapted and translated, how materials were distributed to schools, and whether there were difficulties in the actual administration of the assessments. In addition, there are sections asking about scoring the assessments and preparing and submitting the final data. This detailed set of questions allows the ISC and IEA to get a sense of national-level quality control procedures and identify areas where there may be potential issues or aspects that can be improved in the future. Information from the SAQ is often reported alongside information from the international quality control monitoring program in technical reports to provide a more in-depth picture of the data collection process as a whole.

8.4 International Quality Control

Quality control during test administration includes an international quality control monitoring component in which independent observers visit a sample of classrooms to ensure that standardized procedures and test security guidelines are being followed. Individuals who are independent of the national study centers are hired and trained by IEA to conduct these monitoring procedures. In recent years, international quality control has been implemented in a standardized way across studies, although differences exist because all quality control programs are tailor-made to accommodate the specific needs for each study. IQCMs (although the term international quality observers [IQOs] is used in some studies) are individuals hired in each country to observe (independently from the national center) the actual data collection in a sub-sample of all selected schools in their country and record whether the standardized procedures are followed.

8.4.1 *Development of International Quality Control Procedures*

As described in Sect. 8.2, the development and sharing of standardized manuals was the primary mode of quality control for data collection during the earliest studies. However, even with these manuals, challenges to ensuring uniform data collection and high data quality across countries were common in earlier IEA studies. When referring to the Six Subject Survey, Benjamin Bloom, one of the founding members of IEA, commented, “Inevitably there are difficulties in ensuring that the right tests, etc. get to the right students and that all understand exactly what it is they have to do. In surveys that cross country boundaries, especially where many different languages are involved, administrative problems are magnified and great care in planning is necessary if errors are to be avoided” (Bloom 1969, pp. 10–11).

Although the training and materials described earlier in this chapter helped to ensure additional standardization for data collection, there was still no explicit oversight at the international level to provide information about what happened within schools. This lack of oversight presented challenges. In one of the publications on the results of the Second International Mathematics Study, R. A. Garden, former NRC for New Zealand, commented, “During the study it was the negative aspects which dominated our lives - the National Research Coordinators (NRCs) who did not follow instructions, the postal delays, the misunderstandings, the unreadable data tapes, the miscoded data, and so on” (Garden 1990, p. 1).

In response to some of these issues, the Third International Mathematics and Science Study of 1995 was the first IEA study to implement a coordinated quality control program at the international level. Boston College was the ISC for TIMSS 1995 and oversaw the development of international quality control procedures within the study. Funding, always an issue in previous IEA studies, was resolved in TIMSS

1995 when Boston College received a grant from the US Department of Education to complete data collection for the study (Mullis and Martin 2018). Albert Beaton headed the study at Boston and brought with him a wealth of experience in psychometrics, data collection, and analysis from his many years at Educational Testing Service (ETS) and his work on the US National Assessment of Education Progress (NAEP). Two other experienced researchers also joined the Boston College team, namely Michael Martin and Ina Mullis. Michael Martin was a former NRC for Ireland's international studies, and Ina Mullis, like Albert Beaton, had worked with ETS and NAEP (Schwille 2011). This shared experience with national and international assessment and psychometrics helped shape the work on TIMSS 1995.

Although NAEP was administered within a national context, the study presented some challenges that were similar to those of large-scale international assessments. It thus provided a relevant example of ways to examine and implement quality control across countries. As the number of countries involved in IEA studies continued to increase, their level of comfort and familiarity with implementing large-scale assessment varied widely. The team at Boston College were able to provide leadership on this drawing on the experience of NAEP. In addition to the on-the-ground expertise from Boston College, the US funders for the study wanted to have confidence that the results were comparable across countries. They therefore requested that more rigorous oversight of quality control procedures be included at the international level as a condition of the funding.

While the data collection procedures themselves did not change significantly for TIMSS 1995, the level of oversight for these procedures and the amount of training provided to coordinators at the national and international levels did increase. Similar to previous studies, detailed manuals outlining standardized procedures for data collection were developed collaboratively and used to guide the data collection, although additional manuals were developed and, in many cases, they included greater levels of detail than that provided for prior studies. What was also unique for TIMSS 1995 was that IQCMs were employed and centrally trained to perform classroom-level observations of the data collection as it was taking place (Martin et al. 1996a). Boston College helped organize five different training sessions in various locations around the world so that all IQCMs had the opportunity to attend a session.

The duties of the IQCMs for TIMSS 1995 were standardized across all countries and communicated during the training sessions. NRCs and IQCMs prepared classroom observation tracking forms for each school and classroom under the guidance of Boston College (Martin et al. 1996a). In addition, IQCMs for TIMSS 1995 were asked to interview the NRC about all aspects of the data collection using a structured interview. The interview covered the topics of sampling, experiences working with school coordinators, translation of instruments, preparing test booklets (including sending them to schools and arranging their return), procedures for national-level quality control monitoring, coding of open-ended assessment items, and recording and submitting the final data (Martin et al. 1996a). The questions from this structured interview were later used to develop the survey activities questionnaire (SAQ) that is still in used across IEA studies.

Another new development in the monitoring and standardization of quality control came soon after TIMSS 1995 with the release of the technical standards for IEA studies, which were published in 1999 (Martin et al. 1999). The IEA technical standards focus on the international design and management of the studies, but also address aspects of national implementation that are important for collecting high quality, internationally comparable data.

Quality assurance and control are the primary focus of two of the technical standards. The first relevant standard is the “Standard for developing a quality assurance program.” This standard specifies that operational documentation prepared by the ISC should emphasize quality control as integral to all aspects of a study, particularly data collection activities. The purpose of this technical standard notes that “[quality control] is particularly important for activities such as test administration, which may be conducted by school personnel and therefore outside the control of study staff” (Martin et al. 1999, p. 27). The guidelines recommend making visits to a sample of data collection sites. The data collection monitoring is described as essential for national centers to implement, but also highly recommended at the international level to ensure that unbiased and trained observers can report on the extent to which the sampled schools and classrooms follow the specified procedures.

The second technical standard to specifically address quality control is the “Standard for implementing data collection quality control procedures” and it states that, “Quality control should be an integral part of the study at both the national and international levels. Quality control encompasses both the internal mechanisms that are built into each stage of data collection to ensure that procedures are implemented correctly, and external reviews administered by staff members who are separate from the staff being evaluated” (Martin et al. 1999, p. 59). The implementation of this standard is important in ensuring that the data collection procedures meet the study requirements set by the ISC. The guideline for implementation of this standard emphasizes the ways in which quality control should be built into many steps of the data collection process, for example, hiring and training qualified quality control monitors to assist in observing the administration of the data collection. The guidelines further state that both the ISC and the national center should conduct separate quality control monitoring checks.

International quality control monitoring occurred in all of the IEA studies that followed TIMSS 1995 with a few exceptions. For the Civics Education Study of 1999, there was not enough funding to implement an international quality control monitoring program. However, the ISC (Humboldt University of Berlin) advised NRCs to implement broader national level quality control procedures. NRCs were asked to phone 25% of the tested schools to interview the school coordinator about how testing was done, whether there were any problems encountered, and whether there were any deviations from the testing procedure outlined in the manual (Humboldt University of Berlin, unpublished internal report 1999). The ISC provided formal guidelines for the telephone interviews, along with instructions for how NRCs could select a simple random-sample of the participating schools.

Understanding the development of international quality control procedures also requires an understanding of the development of IEA as an organization. Both structural and financial changes to IEA, as well as the countries involved in IEA studies, have led to changes and developments over time. One country representative provided the following anecdote in regard to early data collection. “This [visits to schools] was a very hard job, some researchers had to reach the schools on horseback since no other means of transportation existed” (Purves 2011, p. 552). The difficulties encountered in reaching and communicating with schools were a potential barrier to implementing a coordinated international quality control monitoring program in the earlier studies. Advances in technology, such as the increased use of email, video chats, and webinars, also helped to facilitate coordination of the studies and the international quality control program.

Changes within the organizational structure of IEA itself also contributed to the increased possibilities for implementing an overarching quality control program. In the early years of the organization, membership was voluntary and there was no formal structure for funding the various studies. Administrative costs for individual studies during the earlier years were often provided by a single organization and individual countries were responsible for funding the collection within their own country. While countries still fund their own individual data collections, a formal fee structure was implemented in the 1990s to help IEA cover the administrative costs of the various studies. Kimmo Leimu, a former NRC for Finland, summarized the impact of this development on project management and oversight: “With the number of actively participating systems increasing up to some 70, recent studies and years have witnessed IEA’s development into a more comprehensive organization both nationally and internationally, despite the fact that national fees have become an indispensable condition for participation. At the same time the projects have become more carefully controlled from beginning to end, with ever-increasing formal procedures detailed at each stage through planning, development, fieldwork implementation, and reporting. In recent years, an international quality monitoring element of test management has been added. An efficient data processing unit enables smooth state-of-the-art analyses of the massive data sets” (Leimu 2011, p. 599). Although the international quality control for TIMSS 1995 was funded primarily by the US Department of Education, subsequent studies have been able to draw on the financial resources provided by the formal IEA funding structure that was implemented during the 1990s. The increasing numbers of participating countries also helps contribute to funding study oversight at the international level.

This expansion of involvement of different countries has added some additional challenges in more recent studies. As Hans Wagemaker, Executive Director of IEA from 1997 to 2014, commented, “[f]or IEA, the inclusion of the broader range of countries with distinctive local circumstances has meant the development of new ways of working to ensure that all countries can participate and that studies continue to achieve the highest technical standards” (Wagemaker 2011, pp. 268–269). Experience in administering large-scale assessment varied widely across countries, especially as countries joined IEA studies for the first time. It was therefore of increasing importance that standardized procedures were clearly documented in the

procedural manuals developed for the studies. Further, the use of independent quality control monitors at the international level helped to ensure that these procedures were being implemented in participating schools and classrooms. While the challenges of a larger and more diverse group of countries involved in the studies require close monitoring of quality, they also enable IEA studies to collect and report data on a broader spectrum of education systems around the world.¹

8.4.2 Implementation of International Quality Control Procedures

The objective of the international quality control monitoring program is to document data collection procedures and to verify that NRCs, school coordinators, and test administrators are following the standardized procedures for data collection. In order to select IQCMs, NRCs are asked to nominate or recommend an individual to serve in this role for their country. All nominations are screened by IEA to ensure that each individual meets the criteria for being an independent observer. For instance, nominees should not be a member of the national study center or a family member or friend of the NRC. The IQCM is often a school inspector, ministry official, or retired school teacher. In many instances, IQCMs are retained across study cycles and continue to serve in this role for subsequent studies.

IQCMs are required to be fluent in both English and the main language of administration, and should have easy access to and experience working on a computer. Additionally, IQCMs need the flexibility to perform their tasks within the required timeline. This often results in a lot of work needing to be completed within a short time frame. To help accomplish this, IQCMs sometimes work together with assistants to help with the classroom observations. Assistants are most common in large countries or countries where the assessments occur on only one or a few days. This helps ensure that IQCMs or their assistants are able to visit the specific schools that are selected for quality control monitoring. Ideally, assistants come from different areas of the country so that a broader geographic spread of schools can be included in the observations.

IQCMs are trained in face-to-face sessions on the standardized procedures for conducting the observations. On average, the IQCM training sessions last between one and two days. Trainers, usually from IEA or the ISC, provide a detailed manual outlining the roles and responsibilities of the IQCM, providing information on the survey operation procedures and assessment design, and including copies of the international questionnaires in English. These materials help ensure that observations and interviews are conducted according to a defined protocol and that responses are

¹The authors would like to acknowledge several individuals who provided their time and expertise on the history of quality control procedures during data collection. Specifically, special thanks are extended to Dirk Hastedt, Paula Koršňáková, Barbara Malak-Minkiewicz, Michael Martin, Ina Mullis, Tjeerd Plomp, Heiko Sibbens, Jack Schwille, and Judith Torney-Purta.

documented on standardized forms. It also helps to familiarize IQCMs with the procedures as they are supposed to be implemented.

Once IQCMs are selected and trained, they conduct school visits and classroom observations. There are three primary purposes of the tasks performed by IQCMs. The first purpose is to validate the sampling within the country. Specifically, it is important to know that the sampled schools, classrooms, and students are the ones actually participating in the assessment. The second purpose is to ensure standardized test administration and data security procedures set by the ISCs are being followed. The third is to provide information on occurrences during data collection that could have an influence on the data quality.

IEA uses rigorous school and classroom sampling techniques to include a representative group of students within each country (see LaRoche and Foy 2016, and Weber 2018 for recent examples). For the majority of studies, sampling is conducted in three stages. First, the countries are asked to provide an exhaustive list of all eligible schools, from which the number of schools to be sampled are selected (usually 150 schools). Selected schools must then provide a list of all the classes that contain students from the target population. From this list, a classroom is selected and all students in the classroom should be included in the study, with a few exceptions for students with disabilities or those that do not speak the language of the assessment. The reliability and validity of the data collected depend on countries closely adhering to the sampling frame that they complete in conjunction with IEA and the ISC. It is therefore essential for IEA and the ISCs to ensure that the agreed upon sampling plans are followed within each country and that any deviations are noted and accounted for. To this end, the international quality control monitoring program provides an opportunity for IQCMs to visit a sample of schools and check that the school name and location match the sampling plan.

As part of their duties, IQCMs also ask school coordinators for information to help validate the within-school sampling. For example, they ask for a list of classes in the target grade(s) at that school and ask whether there are any students at the school that would not be included in these classes. These questions help to validate that the sampled classrooms and students provide the actual data.

Another of the other main aims of the international quality control program is to ensure data comparability by monitoring whether test administrators and school coordinators are following standardized procedures for data collection that are detailed in the manuals. To ensure the manuals are being followed overall within a country, IQCMs are asked to visit samples of individual classrooms on the day of data collection to observe the procedures and note any deviations from the standardized protocols.

The classroom observations during the data collection process are the central and most time-intensive aspect of the IQCMs' duties. IQCMs are generally instructed to visit a sample of either 10% of schools or 15 schools per country. This differs slightly depending on the particular study. For example, the most recent administrations of TIMSS and PIRLS both specified that 15 schools should be selected (Johansone and Wry 2016, 2017). In studies where multiple grade levels are included (e.g., TIMSS), 15 schools per grade should be selected in each country. Further, when one

or more benchmarking participants from the same country participate in a study, five additional school visits are required for each benchmarking entity. The most recent administrations of ICCS and ICILS specified that 10% of schools should be sampled (Koršňáková and Ebbs 2015; Noveanu et al. 2018).

Much of the content of the classroom observation records has remained consistent over time. However, in recent years electronic assessments have become more common across the studies and additional questions have been added to account for this new administration method. The international quality control monitoring process for computer-based assessments is closely aligned with the process for observing the more traditional pen-and-paper assessments. However, some parts of the observation protocol are altered in order to account for the electronic medium of the assessments. For example, the PIRLS 2018 observation record for the computer-based PIRLS modules included questions on whether any technical issues occurred during the testing session (e.g., whether any of the USB sticks were defective, whether the class needed to be split into multiple sections due to the computer availability in the school, and whether any technological problems occurred during the testing session; Johansone and Wry 2016).

The information from the classroom observation records can help inform what may have happened during the data collection to impact the results if issues of comparability do arise at any point during the data management and reporting process. IEA and the ISC receive, compile, and analyze the information from IQCMs to establish whether procedures were followed both within and across countries. In this way, the program can both illuminate specific instances within a country that may need to be examined more closely and identify systematic issues that may be occurring across countries. Although cheating is rare, the program can also help to prevent cheating and incentivize close adherence to study procedures by ensuring that countries know that the data collection will be monitored. Another reason to ensure that procedures are being followed is to check that the assessments and the questionnaires remain under strict control. This helps ensure that items remain confidential so that they can be used for trend comparisons in future studies. It is also important that the individual responses to survey items remain confidential.

In addition to observing the test administration procedures, IQCMs conduct interviews with the school coordinators in the selected schools. IQCMs ask how and when test items were delivered and how they were kept secure prior to the scheduled test administration. Finally, IQCMs collect the final version of all data collection materials from the NRC. These materials include the final manuals for the school coordinators and test administrators, student and teacher listing and tracking forms, and final copies of all questionnaires and assessments. The student and teacher listing and tracking forms are used for sampling validation, while the other materials are used to check the translation of the materials that were actually used during data collection procedures.

While seemingly straightforward, the international quality control monitoring program is essential in ensuring the quality of the data collected. It helps to provide a complete picture of what is actually happening within the schools and classrooms

themselves. This would be difficult or perhaps impossible to capture in any way other than having independent on-the-ground observers record this information.

8.5 Future Directions

The establishment and development quality control procedures during data collection has been an ongoing process since the early IEA years and it is still evolving. This is necessary because of the changing nature of large-scale assessments, especially as technology evolves and more studies and more countries make use of computer-based assessments. While these changes add new layers of complexity, they also offer the opportunity to reflect on what is working well and where policies and procedures may need to be adapted. Currently, information on quality control procedures is disseminated through technical reports or used by IEA and the ISC as a check to ensure that data are valid and reliable. In some ways, these rigorous procedures contribute to the fact that data comparability and high quality data are taken for granted. The program generally uncovers very few issues, but there would be no way to know whether issues exist without the program. Although few issues are usually noted, it is essential to continue documenting adherence to standardized procedures to ensure that studies maintain the consistently high quality for which they have come to be known.

In addition, the context surrounding the assessments themselves has changed over time. The early IEA studies were conducted by researchers for research purposes. Over the years, policymakers and country leadership have taken an increased interest in the results in many of the participating countries leading to assessments becoming more high stakes in those countries. This is important because political pressure to perform could influence the behavior of NRCs, school officials, test administrators, and even the students themselves. In a high stakes environment it is even more important to ensure that there are fully independent observers monitoring the data collection process.

While the quality control procedures described in this chapter are important in ensuring data quality, individual components are regularly evaluated to ensure the quality control monitoring during data collection continues to accomplish the intended purposes. It is also vital to consider new ways in which the information from quality control procedures can inform researchers and study participants. Advances in technology offer opportunities to consider new ways to streamline and improve the process of quality control monitoring.

One issue that is not currently addressed in the international quality control monitoring is what may be happening in schools prior to the day of testing. Organizing the assessment administration within schools is a complex and time-consuming process. Therefore, schools know several months ahead of time whether they have been selected for inclusion in the study. While this information should not influence the educational activities within the school in the time leading up to the actual data collection, there is currently no way to monitor whether this is actually the case;

namely whether some countries are “teaching to the test” or coaching their students prior to the data collection. Some aspects of the assessment design mitigate attempts to provide students with direct answers to questions. This helps prevent efforts to give students the correct responses to individual questions but does not preclude coaching prior to testing.

One possibility to screen for this would be to have IQCMs monitor activities leading up to the data collection more closely. The logistics of organizing any type of pre-assessment monitoring could be complicated and costly, so careful planning would be needed before implementing this type of expansion to the quality control program. In addition, quality monitoring processes that occur once data submission is complete help to check for anomalies in country-level data. For example, sudden and dramatic changes in the mean level of performance within a country would be cause for concern. However, it could still be helpful to know about what is happening in schools at earlier points in time in order to best determine how to address potential situations where this may occur. Ultimately, these efforts are self-defeating for countries because they prevent valid measurement of student performance, which in turn precludes countries from accurately evaluating their education policies and practices for potential changes or improvements.

Another potential issue is that no international quality control monitoring occurs during field testing. The data for field testing is not disseminated externally, but field testing can be seen as a trial run of sorts for the main data collection. Thus, observing the field test could inform IQCMs of issues that need to be resolved prior to the main assessment. This would give the ISC and IEA time to consult with NRCs to ensure that corrections to the procedures can be made in time for the main data collection. National quality control is recommended during field testing, so increased communication in regard to national quality control procedures so that IEA and the ISC are aware of issues that arise during the field test can help countries problem solve before the main study.

In addition to potential pre-assessment monitoring, the procedures implemented during the actual data collection could be enhanced. As mentioned earlier, computer-based assessments are becoming more common. While some of the quality control monitoring procedures have been adapted to account for this medium, there has been little use of the computer-based assessments themselves as a way to collect data on quality control. For example, questions about assessment start and end times could be answered using data stored when students begin and end the assessment. The use of log file data is currently under investigation as a source of information that would further enhance the data already being collected as part of the international quality control program, so this is an area of active development. In addition, some of this data is already used to monitor response patterns for anomalies during the data cleaning phase.

Electronic platforms could also be used to streamline and improve the process of receiving information from IQCMs. Currently, IQCMs fill out paper forms as they observe the data collection within the classrooms and interview the test administrators and school coordinators. IQCMs are then asked to enter that data electronically at a later time and mail the hard copies of the paper forms to IEA or the ISC once

all of their duties have been completed. The current system adds additional steps and time to the process. A possible option for future studies would be to move the observation records to an electronic system that could be completed in real time while the IQCMs are in the schools. This would allow for better monitoring by IEA and the ISC and would cut down on the different steps IQCMs need to complete. It could also allow for more multimedia type information to be uploaded with the observation records, such as photographs of the testing facilities, being careful not to show actual test administrators, teachers, or students. Such technological advances could also be shared with countries for use during national quality control.

Quality control monitoring during data collection plays an important role in ensuring the overall validity and reliability of IEA data across studies. While few issues have emerged over the years, it is still important to continue to consider ways in which monitoring of data collection procedures can be streamlined or improved. These procedures can have a large impact on overall data quality and comparability. It is important that studies continue this type of monitoring to maintain confidence in the quality of IEA data.

References

- Binkley, M., & Rust, K. (Eds.). (1994). *Reading literacy in the United States: Technical report of the U.S. Component of the IEA reading literacy study (NCES 94-259)*. Washington, DC: US Department of Education, National Center for Education Statistics, US Government Printing Office.
- Bloom, B. S. (1969). *Cross-national study of educational attainment: Stage I of the I.E.A. investigation in six subject areas. Final report: Volume I*. US Department of Health, Education, and Welfare: Office of Education, Washington, DC: US Government Printing Office.
- Cresswell, J. C. (2017). Quality assurance. In P. Lietz, J. C. Cresswell, K. F. Rust, & R. J. Adams (Eds.), *Implementation of large-scale education assessments* (pp. 193–204). Chichester, UK: Wiley.
- Foshay, A. W., Thorndike, R. L., Hoyt, F., Pidgeon, D. A., & Walker, D. A. (1962). *Educational achievements of thirteen-year-olds in twelve countries: Results of an international research project, 1959-1961*. UNESCO Institute for Education, Hamburg, Germany. <https://unesdoc.unesco.org/ark:/48223/pf0000131437>.
- Garden, R. A. (1990). *Second IEA mathematics study: Sampling report*. US Department of Education, Center for Education Statistics, Washington, DC: US Government Printing Office.
- IEA. (2020a). Early IEA studies [webpage]. Amsterdam, The Netherlands: IEA. <https://www.iea.nl/studies/iea/earlier>.
- IEA. (2020b). Other IEA studies [webpage]. Amsterdam, The Netherlands: IEA. <https://www.iea.nl/studies/iea/other>.
- Johansone, I., & Wry, E. (2016). Quality assurance program for TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 8.1–8.13). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/publications/timss/2015-methods/chapter-9.html>.
- Johansone, I., & Wry, E. (2017). Quality assurance program for PIRLS 2016. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016* (pp. 9.1–9.14). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-8.html>.

- Koršňáková, P., & Ebbs, D. (2015). Quality assurance of the ICILS data collection. In J. Fraillon, W. Schultz, T. Friedman, J. Ainley & E. Gebhardt (Eds.), *ICILS 2013 technical report* (pp. 127–142). Amsterdam, The Netherlands: The International Association for the Evaluation of Educational Achievement (IEA). <https://www.iea.nl/publications/technical-reports/icils-2013-technical-report>.
- LaRoche, S., & Foy, P. (2016). Sample implementation in PIRLS 2016. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016* (pp. 5.1–5.126). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. <https://timssandpirls.bc.edu/publications/pirls/2016-methods/chapter-5.html>.
- Leimu, K. (2011). 50 years of IEA: A personal account. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Eds.), *IEA 1958–2008: 50 years of experiences and memories* (pp. 591–626). Nicosia, Cyprus: Cultural Center of the Kykkos Monastery. <https://www.iea.nl/publications/iea-reference/iea-1958-2008>.
- Martin, M. O., Hoyle, C. D., & Gregory, K. D. (1996a). Monitoring the TIMSS data collection. In M. O. Martin & I. V. S. Mullis (Eds.), *Third international mathematics and science study: Quality assurance in data collection*, (pp. 3.1–3.12). Chestnut Hill, MA: Boston College. <https://timssandpirls.bc.edu/timss1995i/TIMSSPDF/QACHP3.PDF>.
- Martin, M. O., Mullis, I. V. S., & Kelly, D. L. (1996b). Quality assurance procedures. In M. O. Martin & D. L. Kelly (Eds.), *Third international mathematics and science study (TIMSS) technical report volume I: Design and development* (pp. 11.1–11.12). Chestnut Hill, MA: Boston College. <https://timssandpirls.bc.edu/timss1995i/TIMSSPDF/TRCHP11.PDF>.
- Martin, M. O., Rust, K., & Adams, R. J. (1999). *Technical standards for IEA studies*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA). <https://www.iea.nl/publications/iea-reference/technical-standards-iea-studies>.
- Mullis, I. V. S., & Martin, M. O. (2018). 25 years of TIMSS and PIRLS. In S. Finlay (Ed.), *60 Years of IEA (1958–2018)* (p. 24). Amsterdam, The Netherlands: IEA. <https://www.iea.nl/publications/60-years-iea-1958-2018>.
- Noveanu, G. N., Kobelt, J., & Köhler, H. (2018). Quality assurance procedures for the ICCS data collection. In W. Schultz, R. Carstens, B. Losito & J. Fraillon (Eds.), *ICCS 2016 technical report* (pp. 67–86). <https://www.iea.nl/publications/technical-reports/iccs-2016-technical-report>.
- Postlethwaite, N. (1967). *School organization and student achievement: A study based on achievement in mathematics in twelve countries*. Stockholm, Sweden: Almqvist and Wiksell. https://ips.gu.se/english/Research/research_databases/compeat/Before_1995/FIMS/FIMS_Postlethwaite.
- Purves, A. C. (2011). The evolution of IEA: A memoir. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Eds.), *IEA 1958-2008: 50 years of experiences and memories* (pp. 531–556). Nicosia, Cyprus: Cultural Center of the Kykkos Monastery. <https://www.iea.nl/publications/iea-reference/iea-1958-2008>.
- Schwille, J. (2011). Experiencing innovation and capacity building in IEA research: A memoir, 1963–2008. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Eds.) *IEA 1958-2008: 50 years of experiences and memories* (pp. 627–708). Nicosia, Cyprus: Cultural Center of the Kykkos Monastery. <https://www.iea.nl/publications/iea-reference/iea-1958-2008>.
- Statistics Canada. (2010). *Survey methods and practices*. Catalog number 12-587-X. Ottawa, Canada: Statistics Canada. <http://www.statcan.gc.ca/pub/12-587-x/12-587-x2003001-eng.pdf>.
- Wagemaker, H. (2011). IEA: International studies, impact, and transition. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Eds.), *IEA 1958-2008: 50 years of experiences and memories* (pp. 253–272). Nicosia, Cyprus: Cultural Center of the Kykkos Monastery. <https://www.iea.nl/publications/iea-reference/iea-1958-2008>.
- Weber, S. (2018). Sampling design and implementation. In W. Schultz, R. Carstens, B. Losito & J. Fraillon (Eds.), *ICCS 2016 technical report: IEA International Civic and Citizenship Education Study 2016* (pp. 43–52). Amsterdam, the Netherlands: IEA. <https://www.iea.nl/publications/technical-reports/iccs-2016-technical-report>.

Lauren Musu is a Senior Research Officer at IEA and has a background in educational psychology and statistics. She coordinates the International Quality Control Program for various IEA studies, which includes overseeing the program and developing materials for International Quality Control Monitors. She is also involved in data analysis and reporting for both internal and external purposes.

Sandra Dohr is a Research Officer at IEA and has a background in sociology and educational science. She is involved in managing International Quality Control Programs for various IEA studies, which includes recruiting and supporting International Quality Control Monitors and overseeing the implementation on a local level. In addition, she is involved in the coordination of Translation Verification procedures for IEA studies.

Andrea Netten is the Director of IEA Amsterdam. She has a background in child psychology and was the PIRLS NRC for the Netherlands for PIRLS 2006 until PIRLS 2016. She holds managerial responsibilities for the IEA headquarters and helps ensure that the responsibilities of IEA Amsterdam in international projects are met. She also maintains contact with IEA partners and responds to membership requests.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

