Chapter 10

Technology and Assessment

Heiko Sibberns

Abstract The International Association for the Evaluation of Educational Achievement (IEA) is moving toward computer-based assessment (CBA) in all its internationally conducted large-scale assessment studies. This is motivated by multiple factors, such as the inclusion of more comprehensive measures of the overall construct, increased use of online tools for instruction and assessment in IEA member countries, the need to develop assessments that are more engaging for students, and the hope that CBA will lead to increased efficiency and yield cost savings for participating countries. The transition to CBA has an impact on the design and procedures in both the international study centers and the national study centers. The very high standards in paper-based testing must be preserved throughout the transition steps. This relates not only to the layout, translation, and translation verification steps but also concerns the test administration and all associated procedures. The consistency, accuracy, and completeness of all collected data have to be achieve the same standards as paper-based testing, and the maintenance of the trend measures is of critical importance. The challenges arising from the transition to e-assessment are presented and discussed, including procedures developed to correct for any potential deviations that can be attributed to changes in the delivery mode. CBA provides potential new avenues for data analysis. Log-file and process data are of particular interest to researchers because such data provide an opportunity to increase the validity of the scales and minimize influences irrelevant to the construct being measured.

Keywords Computer-based assessment · Log-file data · Mode-effect · Process data · Transition from paper to computer assessment

H. Sibberns (✉)

International Association for the Evaluation of Educational Achievement (IEA),
Hamburg, Germany

e-mail: heiko.sibberns@iea-hamburg.de

10.1 Introduction

Innovation and change has been an ongoing feature of the development of the International Association for the Evaluation of Educational Achievement's (IEA's) large-scale assessments. Nowhere has this been more evident than in the use and development of computer-based data capture and assessments. For the most part, this change has been triggered by changes in the environment. Internationally, there is greater use of technology for both learning and instruction. This coincides with the increased availability and use of computer technology outside of school. The transition to greater use of technology in assessment is based on a number of perceived benefits and initial successes and reflects the changing learning environment. However, there are still inequalities in information technology (IT) infrastructure between and within countries that need to be addressed if all envisaged benefits are to be achieved.

10.2 Technology in Education

In the majority of countries participating in IEA studies, instruction and assessment have gradually moved toward increasing use of technology in recent years (see, e.g., Bundesministerium für Bildung und Forschung 2019; UK Department for Education 2018). In many countries, ambitious goals for the provision of IT infrastructure in schools are not only formulated but large amounts of money have been made available to achieve them (see e.g., Bundesministerium für Bildung und Forschung 2019). As a result, there is a growing education technology industry. The central idea behind this is “that digital devices, software and learning platforms offer a once-unimaginable array of options for tailoring education to each individual student's academic strengths and weaknesses, interests and motivations, personal preferences, and optimal learning pace” (Herold 2016).

In addition to the provision of hardware and educational software that can be adapted to the needs of the teachers and students, a greater emphasis on the preparation of teachers in the effective use of technology is evident and demanded if the stated goals are to be achieved (again, see e.g., Bundesministerium für Bildung und Forschung 2019). Approaches that support the integration of private student computers (Herold 2016) have also advanced the use of the use of technology for learning and instruction.

IEA and its member countries recognized the need to make its studies available as computer-based assessments (CBAs). Today, whenever a new cycle of a study is presented at IEA's decision-making body, a computer-based version is immediately requested. Key arguments are based on major investment in IT infrastructure in schools, fears that a paper-based test would lack acceptance as the use of technology in schools advances, and a realization that students are increasingly using computers for

everyday tasks. However, among the diverse set of countries in which IEA operates, there are still a considerable number of countries that do not yet have the requisite infrastructure to implement CBA, and hence paper-and-pencil assessment is still important.

10.3 Promises and Successes of Technology to Reform Assessment Data Collection

Most programs for the introduction of CBA are accompanied by the promise to increase the efficiency of the survey. Other arguments are that the reliability and validity of the tests will be increased and new content domains can be explored.

10.3.1 Efficiency

CBA may lead to a decrease in costs. It is assumed that no printing and shipping of material to schools is required because the school can use their available resources for testing. Data entry is no longer required because student answers are directly entered into the computer. Finally, the scoring of constructed response items could be supported by machine scoring, provided the stored responses are comparatively simple, like numbers or fractions. However, gains in efficiency must also take into account changes in the overall cost structure and the redistribution of costs between national study centers and the international study center (ISC). When using paper-and-pencil testing, the participating country funds the necessary infrastructure and organization for many of the steps involved in the survey operations. For example, the layout of the national test instruments is undertaken within the country with the help of a desktop publishing program to eliminate deviations from the internationally-specified layout. The scoring of the constructed response items is also organized and carried out nationally. However, with CBA, the ISC must instead provide the technical infrastructure to revise the layout, as this requires programming knowledge and knowledge of the system used. The ISC must also provide the entire technical environment for scoring; this includes the display of all student responses, a tool for training the scorer, the assignment of the scorer to student responses, and a tool for monitoring the scoring process.

10.3.2 Increased Reliability: Direct Data Capture

When students enter their responses directly on the computer, this not only leads to a reduction in the costs of data entry but also leads to the elimination of expected

input errors. Errors that arise from illegible handwriting are also avoided, especially for constructed response items, since illegible items cannot be scored correctly or at all.

10.3.3 Inclusion of More Comprehensive Measures of the Overall Construct

A major advantage of CBA is the inclusion of more comprehensive measures of the overall construct. Aspects of the constructs that cannot be assessed via the traditional paper-and-pencil assessment are now accessible. Three examples illustrate the potential for enhancing measurement:

- (1) The 2016 cycle of IEA's Progress in International Reading Literacy Study (PIRLS) included an electronic assessment (ePIRLS) of the reading competencies needed in the informational age, namely the non-linear reading skills needed to navigate webpages
- (2) The 2019 cycle of IEA's Trends in Mathematics and Science Study (TIMSS) also included an electronic assessment (eTIMSS) that included innovative problem-solving and inquiry tasks (PSIs), extending the student assessment into areas that could not have been assessed by traditional paper-and pencil testing.
- (3) The 2018 cycle of IEA's International Computer and Information Literacy Study (ICILS) included novel items measuring students' abilities in the domain of computational thinking.

10.3.4 Reading: Additional Competencies Needed in the Information Age

For ePIRLS 2016, CBA provided an engaging, simulated internet environment that presented grade 4 students with authentic school-like assignments involving science and social studies topics (see Mullis et al. 2017). An internet browser window provides students with a website containing information about their assignments, and students navigate through pages with a variety of features, such as graphics, multiple tabs, links, pop-up windows, and animation. In an assessment window, a teacher avatar guides students through the ePIRLS assignments, prompting the students with questions about the online information. Through this environment, non-linear reading of texts is introduced into the assessment.

The ePIRLS 2016 assessment consisted of five tasks, with each task lasting up to 40 min. Each student was asked to complete two of the tasks according to a specific rotation plan. The assessments were administered via computer (typically MS Windows-based computers) and students entered their answers by clicking on options or typing words. All input material for ePIRLS was tailor-made for the purposes of the study. The websites were hard-coded and therefore no templates had to be made available (Mullis and Martin 2015).

10.3.5 Mathematics and Science: Inclusion of Innovative Problem-Solving Strategies

To extend coverage of the mathematics and science frameworks, eTIMSS 2019 included additional innovative PSIs that simulated real-world and laboratory situations where students could integrate and apply process skills and content knowledge to solve mathematics problems and conduct scientific experiments or investigations. PSI tasks, such as designing a building or studying plant growing conditions, involve visually attractive, interactive scenarios that present students with adaptive and responsive ways to follow a series of steps toward a solution. Early pilot efforts indicated that students found the PSIs engaging and motivating (Mullis and Martin 2017). These PSIs provide an opportunity to digitally track students' problem solving or inquiry paths, and analysis of the process data, which reveals which student approaches are successful or unsuccessful in solving problems, may provide information to help improve teaching (see IEA 2020a).

It should be emphasized that the demanding criteria for PSIs made their development very challenging and resource intensive. Special teams of consultants collaborated both virtually and in person to develop tasks that: (1) specifically assessed mathematics and science ability (and not reading or perseverance); (2) took advantage of the electronic ("e") environment; and (3) were engaging and motivating for students (Mullis and Martin 2017).

10.3.6 Computational Thinking: Developing Algorithmic Solutions

Computational thinking refers to an individual's ability to recognize aspects of real-world problems which are appropriate for computational formulation and to evaluate and develop algorithmic solutions to those problems so that the solutions could be operationalized with a computer (Fraillon et al. 2019, p. 27).

Following this definition, the computational thinking test modules in ICILS 2018 contained tasks developed to make use of elements originating in visual programming languages. Students were able to arrange blocks of code to solve real-world problems without knowing a particular programming language. More broadly, ICILS was developed to specifically measure computer skills. So CBA was essential in this case, and the process and sequencing data are of special interest. ICILS measures international differences in students' computer and information literacy (CIL): their ability to use computers to investigate, create, participate, and communicate at home, at school, in the workplace, and in the community. As mentioned, participating countries also had an option for their students to complete an assessment of their computational thinking (CT) ability, and approaches to writing software programs and applications. ICILS 2018 was administered by USB (universal serial bus) stick and local server mode (Fraillon et al. 2019).

10.3.7 Increased Reliability: Use of Log-File Data

The extended potential of computer-based testing is that the data record not only the student answers but also capture additional information about *how* students achieve those answers, such as their navigation behaviors, information about the time taken to reach an answer, or which tools they used (ruler, compass, or screen magnifier). This “log-file” data can be analyzed to reveal more about test-taking behaviors and strategies.

In an overview article, von Davier et al. (2019) outlined many of the options available for the additional use of log-file data. Timing information can not only be used to increase the reliability and validity of scales but also to discover rapid guessing or disengagement. When dealing with complex problem-solving tasks, the analysis of the processing steps and sequences makes it possible to distinguish successful processing strategies from less successful ones.

Ramalingdam and Adams (2018) showed that the reliability and validity of the scale in the field of digital reading can be increased by using log-file data. In the context of their study, only multiple choice items were considered. Log-file data helped to distinguish students who read the clues that were necessary to answer the item correctly from students who only guessed.

10.3.8 Development of More Engaging and Better Matching Assessments

More interactive test material and new item types that cannot be used on paper are perceived as an advantage of a CBA system because of the likelihood of increased student engagement. However, in cases where students are being assessed using test material that is also administered on paper, care must be taken to ensure that the CBA environment does not have a different impact on student performance; this is known as a mode effect (Walker 2017). It is assumed that, ultimately, the implementation of “adaptive testing” will enable student ability to be better matched to item difficulty, increasing student engagement, reducing student frustration, and, consequently, reducing the amount of missing data. Overall, this should result in more accurate assessment of student ability. However, like other parallel forms of test administration, mode effects need to be carefully monitored to ensure measurement invariance. Another potential advantage of CBA lies in the option to increase the accessibility of the tests by making supporting tools available for students with special needs. Many such opportunities are under investigation for implementation, such as applications for magnifying texts or graphics, or options to increase contrast between text and background. Text-to-speech solutions are another option; however, such initiatives have consequences for the test administration as earphones must be available if this option is used and all text elements in all languages would have to be available as voice recordings, which would be a considerable additional effort.

10.4 The Transition

10.4.1 *Delivering Questionnaires Online*

Before developing computer-based applications for the assessment of student achievement, IEA tackled the web-based administration of the background questionnaires. Since the content of the background questionnaires differs from the student assessment, the protocols that have to be followed for their administration are not as strict because the security requirements for the protection of questionnaire items are less critical. There is also no need for a controlled and standardized computer environment, thus the first step was to develop a comparatively soft system for surveying teachers, school leaders, and parents. The first system used for background questionnaires was the IEA Survey System, which was developed in 2004 and first used as part of the Second Information Technology in Education Study (SITES) in 2006 (Carstens et al. 2007). This was later renamed the Online Survey System (OSS). To be applicable in an international environment, the OSS had to provide adaptable questionnaires that enabled the necessary cultural adaptations (e.g., use of different metrics), and accommodated the addition or deletion of answering options. When, for example, asking about school types, certain countries may have a limited amount of school types while others, like federal systems, may have a larger variety. Equally, in other cases, questions may not be applicable and can therefore be deleted. National research programs may also wish to add optional national questions to the international questionnaire.

As questions need to be translated and the translations need to be verified, a tool to allow for adaptations, translations, and translation verification had to form part of the survey system, or, at the very least, the necessary interfaces had to be available. The system also had to work in a variety of cultures and for right-to-left languages, as well as languages with additional character sets.

Interfaces had to be programmed to guarantee the seamless integration of respondents' answers. Metadata, like codebook information from the IEA's existing data processing environment, also had to be included. Finally, the content had to be a standard hypertext markup language (HTML) format so there were no limitations on browser displays.

The OSS was always a purely online solution and did not support offline options like USB-delivery or local server mode. For this reason, from the very beginning, the amount of data that needed to be transmitted, in both directions, was kept to a minimum. Another prerequisite was that the OSS had to be easy to use, both in setting up the system and in enabling the adaptations and translations required by countries. OSS is still used for IEA studies, but ongoing maintenance and modifications to permit proper content display on current browsers means that new solutions are always under investigation. Over the years, OSS has proved to be a reliable and robust tool for carrying out surveys; slow internet connections or outdated browsers have not presented a problems. However, as technology as evolved the design has become increasingly dated, and complex surveys with many filter questions and branches require more modern solutions.

10.4.2 Computer-Based Assessment

IEA's move toward CBA has been gradual and careful. The OSS was initially developed for background information only, followed by the development of an assessment system supporting the assessment of web-based reading in 2014 for PIRLS 2016 and, shortly afterward, a full electronic assessment (eAssessment) system for TIMSS 2019. ICILS was based on a national study conducted in Australia. Consequently, ICILS international implementations in 2013 and 2018 made use of the CBA system that was developed for the Australian national study.

ePIRLS. Because internet reading is increasingly becoming one of the central sources that students use to acquire information, in 2016, PIRLS was extended to include ePIRLS, an innovative assessment of online reading (Mullis and Martin 2015). The main characteristics are non-continuous texts, such as those found on websites on which content is linked via hyperlink. These types of text were presented to the students in a simulated internet environment.

eTIMSS. TIMSS 2019 continued the transition to conducting the assessments in a digital format. eTIMSS provided enhanced measurement of the TIMSS mathematics and science frameworks and took advantage of efficiencies provided by the IEA eAssessment systems. About half the countries participating in TIMSS 2019 transitioned to administering the assessment via computer. The rest of the countries administered TIMSS in a traditional paper-and-pencil format, as in previous assessments.

To support the transition to eTIMSS, IEA developed eAssessment systems to increase operational efficiency in item development, translation and translation verification, assessment delivery, data entry, and scoring. The eTIMSS infrastructure included the eTIMSS Item Builder to enter the achievement items, an online translation system to support translation and verification, the eTIMSS Player to deliver the assessment and record students' responses, an online Data Monitor to track data collection, and an online scoring system to facilitate the national study centers' work in managing and implementing scoring of students' constructed responses.

eTIMSS also included new digital ways for students to respond to specific constructed response items, which enabled student responses to many items to be scored by computer rather than by human scorers. In particular, a number keypad enabled students to enter the answers to many constructed response mathematics items so that the answers can be computer scored. Other types of constructed response items that can be computer scored use drag-and-drop or sorting functions to answer questions about classifications or measurements.

eTIMSS was administered via USB delivery and through local server mode. Originally, the plan envisaged delivery via tablet devices, and it was expected that students could write down their answers with styluses. However, first trials with grade 4 and grade 8 students revealed that the interaction with the tablet device proved overly complicated and the stylus approach was dropped. On tablets, students could instead activate built-in keyboards and, since a significant number of countries had already invested in laptop computers for assessment programs, the

tablet approach was extended and delivery using personal computers (typically MS Windows-based computers) was implemented where students could make use of the physical keyboard and the mouse. Drawings were supported by a line drawing tool, and calculations through a built-in calculator (Mullis and Martin 2017).

ICILS. ICILS 2013 evaluated students' understanding of computers and their ability to use them via an authentic CBA of grade 8 students. ICILS 2018 was linked directly to ICILS 2013, allowing countries that participated in the previous cycle to monitor changes over time in computer and information literacy (CIL) achievement and its teaching and learning contexts (Fraillon et al. 2019). A third cycle is planned for 2023 (IEA 2020b).

10.5 Challenges

Regardless of the potential and possibilities of computer-based testing, there are many challenges and problems at all levels that need to be solved and which demand careful handling. Some of the main challenges are briefly considered here.

Most international large-scale assessments (ILSAs) generate trend measures enabling assessment results over time to be compared. Scores reported over time are established using the same metrics and represent valid and reliable measures of the underlying constructs; new scaling approaches may provide additional information. Differences in the properties of the items must be carefully considered. Items that show different behavior in the paper and the computer modes may have to be treated as different items, even if they are apparently identical (von Davier et al. 2019). The mode effect challenge remains an ongoing issue as the various assessments transition to become computer based. IEA continues to make considerable investments to ensure these risks are mitigated in all phases of the study implementation.

At the design stage, before significant resources are spent in the implementation of a CBA, one single small group of students is asked to work on practice versions of the items and “think aloud” when working through them. These qualitative student reactions are reported and analyzed. They inform the graphical design of the entire assessment system, as well as the arrangement of graphical and text elements in individual items. Equivalence studies, such as those undertaken in the development of eTIMSS 2019 are critical (see Fishbein et al. 2018). In the year prior to the field trial for eTIMSS 2019, participating countries were invited to administer certain trend items in both paper and electronic format. Students received two test forms, one was computer based, and the second was paper based. It was assumed that both tests measured the same construct and the differences in item behavior were only marginal. However, results showed that, while the same item administered in the computer environment and on paper showed similar behavior, students found items administered on computers slightly harder than the same items on paper. The effect was more prominent for mathematics items than for science items, and, although small, could not be dismissed. Thus, in order to safeguard the trend measure, each participating country had to take what was termed a paper bridging study. In each

eTIMSS country, a smaller sample of 1500 students took a paper test based on the trend items to establish the trend differences for each country. These paper tests were analyzed with the standard calibration methods established for previous cycles of TIMSS (see chap. 13 of Martin et al. 2016). Finally, the computer-based test scores were adjusted (so that they could be aligned with the trend results established by the paper bridging study).

Layout adjustments that are necessary after the tests have been translated are a particular challenge. They result when the translated text is longer than the original English text or when it contains different alphabets.

Graphics often contain text elements such as labels in coordinate systems or frequency tables, or descriptions in maps. All text elements must be translated and then indicated so that the graphic will continue to be correctly labelled. Since some languages require significantly more characters in translation than the English original, so dynamic adaptation of the text fields has to be supported. In extreme cases, manual adjustment is essential.

Often, text in the translated language is not displayed correctly because the space provided for it in the assessment system was too small and had to be expanded afterwards. In principle, this is also a problem in paper-based testing, but is relatively easier to solve when the commercial layout software used in paper-and-pencil design provides tools that enable quick and uncomplicated adjustments

Frequently, text boxes not only contain plain text but also graphics in the form of special mathematical symbols, currency symbols, or special symbols from the natural sciences. In part, these symbols must also be “translated” when used in another language, for example, in the English original the “/” is used as a sign for the division, while in German “÷” is used.

In particular, right-to-left (Hebrew and Arabic) written language systems, as well as some Asian languages, present a major challenge. Discussion of all possible special cases would go beyond the scope of this chapter. Suffice to say, a mathematical expression in a sentence is written in Arabic from left to right, while the rest of the sentence is written from right to left. Thus when writing “What is the value of x in the equation $x + 12 = 15$?” in Arabic it will appear as:

ما هي قيمة x في المعادلة $x + 12 = 15$ ؟

Irrespective of the language used, geometry items pose a particular challenge. In the paper version, geometrical figures must be constructed using a ruler or compass. A ruler is also needed to determine distances. When such items are transferred to a CBA system, technical tools must also be provided to allow the required operations to be performed. The use of the aids must be explained and practiced in an introductory section prior to the test so that students taking the test understand exactly how to activate the ruler and how it has to be handled to take the required length measurement.

A fundamental decision must be made as to which devices and which associated operating systems should be allowed to run the assessment. The planned assessment material and the expected interactions determine the type of devices. Screen size, screen resolution, the decision for or against touch screen, the type of keyboard (pop

up or physical keyboard) and the use of mouse or touchpad are set after the decisions regarding the assessment are taken.

For eTIMSS 2019, supporting tablet, laptop, and personal computer delivery turned out to be extremely demanding. Different technologies had to be used for Android®-based tablet devices and Windows®-based computers. As a consequence, digitalPIRLS was programmed only for personal computers. When using tablets, frequent updates of the operating system might pose a risk for the data collection. It may happen that accesses of the test software to certain system components of the operating system no longer work after an update and the assessment software stops functioning. The problem can be mitigated by resetting updates or by adapting the test software.

A challenge that is pertinent to tablet delivery is the interaction with the assessment system. In the initial phase of moving to CBA in eTIMSS, it was envisaged that student might use a stylus to record their answers on the tablet screen. However, early trials indicated that students were not used to this form of electronic interaction and the idea was quickly abandoned. Pop-up keyboards and number pads were implemented instead. For digitalPIRLS, physical keyboards are mandatory.

In a best-case scenario, the necessary (number and kind) resources are available in all schools and the national research coordinator (NRC) conducting a study has permission to use those resources. Even in this situation, schools need to be contacted prior to the testing in addition to the contacts necessary for the organization of students' testing. A diagnostic tool needs to be run on all the computers that the school plans to use for the test. The diagnostic tool checks screen resolution, available memory, processor speed, and other parameters that need to be met in order to run the assessment program smoothly. This diagnostic tool also has to reflect the delivery mode that is planned for a specific school. The assessment can only be conducted in the school if all machines pass the checks, and it may be that individual computers fail the requirements; in this situation the NRC has to devise a backup solution, like providing carry-in laptops or splitting test groups in schools.

Often, however, there may be no school resources available or access to them might be limited or denied because of school security policies. In this situation, laptops need to be delivered to the school purely to conduct the assessment, and, in most cases, this would impose a financial burden on national study centers because they would have to be purchased or rented for the assessment. This is generally only a viable approach if the computers are to be used regularly, for example for other international assessments or national testing programs. Such machines need to be especially robust because they have to be shipped to schools, and should be equipped with up-to-date hardware that allows a reasonable duration of use. For laptop delivery, a reliable and detailed logistic system has to be established in order to have a minimum number of laptops distributed among test administrators in a well-planned order.

Even assuming that equipment in most schools is ready for the type of CBA programs used in ILSAs, indications from NRCs are that schools tend to be more restrictive due to security policies. A further obstacle to the use of school resources is the paucity of dedicated information technology administrators or coordinators

within schools; these tasks are frequently outsourced to external contractors. Thus, changes to school contracts to grant permission to run an external assessment software often come with additional costs for schools and are therefore rejected. But the situation is constantly evolving, and there is evidence that countries are investing more in computer technologies in schools, including additional staffing (see for example, Bundesministerium für Bildung und Forschung 2019; UK Department for Education 2018).

Whatever situation is found, manuals have to be adapted to reflect the changes due to CBA. Unlike paper-based assessment, where paper instruments are distributed according to a strict protocol as manifested in the test administrator manual, computers have to be set up and prepared. The need for diagnostic tests of all computers to ensure proper functioning prior to the assessment has already been mentioned. On the test day, schools need to ensure that the testing software is either installed or accessible from all machines. When laptops are brought into the school, sufficient and properly fused power lines have to be available, and all computers have to start. Depending on the quality of computers, this could last for some time and may justify the presence of a second test administrator responsible for the proper functioning of the technology. This person could also be responsible for remedying problems. Students have to be provided with login and password information, and test administrators have to make sure that all students apply the information correctly.

During the first assessments making use of CBA, IEA noted shifts in the study cost structures. All processes that were needed in the paper-based assessment and that were, to a large extent, executed on paper or in word processors or layout programs, needed to be reproduced in a computer environment (see Sect. 10.3.1). The modules had to be developed, maintained, and updated or replaced if changes in computer technology required these changes. In addition, a move towards web-based delivery has significant demands on the system in use to guarantee confidentiality, integrity, and availability. This requires redundant systems. Identical server solutions around the world are needed to cope with delays due to long distances between the test taker's device and the server on which the data is stored and the system also needs to cope if significant numbers of respondents are accessing the system simultaneously when CBA is set up on a large scale and makes use of web-based testing. A shift of costs towards the international operations need to be properly accounted for in a revised fee structure. Conversely, significant savings can be expected at the country level because the central infrastructure is kept and maintained under the supervision and responsibility of the ISC.

10.6 The Future: Guiding Principles for the Design of an EAssessment Software

Based on IEA's experience with a variety of programs, including IEA's propriety software, the following principles can serve as a guideline for developments that secure high quality software components in the future.

10.6.1 Adaptive Testing

In adaptive testing, test tasks or blocks of tasks are put together during the test session and then presented to the student so that the maximum information about the student's performance is obtained. This provides a very precise and reliable estimate of student performance. In summary, it can be said that the difficulty of the test obtained in this way corresponds to the ability of the student and that the boundary conditions for the content of the test are also met. The possibility of adaptive testing must also be considered in future developments of ILSA (Wainer 2000). Different methods are available, starting with group adaptive testing where tests of different difficulty are created and groups of pupils of different ability are identified from available data before the test administration; student groups then focus on the test most appropriate to their abilities.

In a second form of multi-stage testing, each student initially receives a test of medium difficulty. The test is evaluated immediately in the CBA system. Depending on how well they managed the first test section, the student then subsequently receives a test section that may be more closely tailored to their ability; less challenging if the student had difficulty, or more challenging if the student achieved good results. This is also again evaluated immediately and another test section assigned depending on the previous result.

Finally, in the case of a fully adaptive design, the presentation of test items is a function of their success or failure on previously presented items. After each item has been processed, student skill is estimated and a new item of appropriate difficulty assigned.

All these test methods have in common that a lot of item material has to be developed and calibrated. In particular, many boundary conditions, such as a given distribution of the cognitive domains, the content domains, or the item types, must be taken into account when assembling a test form on-the-fly for particular students during the test session.

10.6.2 Translation

The translation system makes it possible to reliably translate the international survey instruments. Since many translators work with professional translation software that provides functionalities beyond what can be provided in an assessment system, it is necessary to implement a common interface to other systems. The XLIFF (XML localization interchange file format) is an XML-based format created as a common exchange format for translations and is also supported in IEA's propriety software (OASIS 2008). Further information about translations and the challenges created by a CBA system can be found in Chap. 6.

10.6.3 Printing

Very desirable is a functionality that allows the creation of paper instruments from the CBA system because items in common for both assessment modes are then maintained in only one system and an item need only be changed once avoiding deviances between items delivered on paper and on computer. Previous attempts for a paper export have not been satisfactory, although, as a rule, all text elements could be displayed correctly. Graphics, however, were often out of focus. Then, when the resolution of the graphic was increased, too much data load was generated for the computer-based delivery. Nonetheless an export function remains an important goal because the same assessment may be carried out both on paper and delivered by computer until all countries are able to make the step towards computer-based testing. With such an export function, the test development could take place in one environment only and, as well as avoiding inconsistencies between the paper and the computer environment, the additional costs for a double input with associated quality control would not be incurred.

10.6.4 Web-Based Delivery

A web-based test would have the great advantage that all information can be available in real time for both the national study centers and the respective ISC. No programs need to be installed and, given the webpages created are not too demanding, any web browser would work. If problems occurred during test application, the change would be made only once, and a time-consuming and logistically difficult new rollout on USB sticks or server notebooks would be superfluous. If a suitable monitoring tool is available, the course of the study can be closely monitored and, if necessary any irregularities addressed, for example, if participation rates were lower than expected. It would also be possible to react to delays or irregularities in the sequence, such as an accumulation of tests in a certain period of time in the test window; the testing

sessions could be rearranged and spread more evenly across the testing window. A connection to a logistics system is also conceivable, with which the use of test administrators could be planned, controlled, monitored and, if necessary, accounted for.

This approach is only viable, if sufficient bandwidth for data transmission is available. Considerable efforts may be necessary to secure the availability, integrity, and confidentiality of a web-based delivery system: the test must be available at the required time during the testing session in the school, confidential information must be protected from unauthorized access, and the data must be complete and unchanged when transmitted or during data processing. Unlike a USB delivery, a system failure would cause massive damage because a malfunction could potentially endanger the entire assessment; a corrupted USB stick also leads to loss of data, however, only a single student is affected.

10.6.5 General Considerations

In all IEA's ILSAs, constructed response items are used. A significant number of these items have to be scored by people, as only short answers, such as intact numbers or fractions, can be machine-scored. For the scorers, it is important that they see not only the isolated response of the student but also the response embedded in the page the student has answered. The displayed page must be based on the translated version, as it is the only way to detect inaccuracies in the translation that influence the student's response.

In any future developments, the following factors should be considered. CBA systems and workflows should work under secure conditions, include a reasonable backup strategy, and fulfil data protection requirements. Countries with special data security concerns should be handled appropriately (e.g., through data encryption, as implemented already in various systems). What can be accessed within the different eAssessment system modules and what actions a user can and cannot undertake depends on the different user roles. Participating countries should only be able to access the version that was generated for the specific country and their own data; they should not be able to see data from other countries. CBA environments need to show only the specific study-related parts if countries and ISCs access the different modules and they should not provide access to data from other studies.

The designer tool should ideally be able to integrate third-party tools like GeoGebra (a free online mathematics tool; <https://www.geogebra.org/>) or import items that were generated using third-party programs and stored according to agreed industry standards (such as the IMS Global Learning Consortium's question and test interoperability (QTI) format, which defines a standard format for the representation of assessment content and results, supporting the exchange of this material between authoring and delivery systems; <http://www.imsglobal.org/>). The designer tool should also be used as an item bank, tracking items that were used in different studies. Items should be transferable between study cycles, and between one study

and another, as needed, especially questionnaire and released assessment items. The item designer tool should be able to track the item development workflow.

IEA has used three delivery modes for CBA in their studies. It is advisable to support at least two modes in a CBAs because conditions in countries vary considerably in terms of IT-infrastructure like hardware, connectivity to the internet, bandwidth, or access rights. Web-based delivery was described in Sect 10.6.4. Two other modes were USB delivery and local server mode:

In USB mode, the entire testing system, including the delivery and storage system, and any necessary browsers are stored on a USB stick, and the assessment is run from the stick. This technology is advisable when computers in schools do not have a solid internet connection or when there is only restricted bandwidth capacity for data transmission. Computers need permission to access USB devices and run programs from USB sticks. One disadvantage is the logistics necessary to implement the USB approach. USB sticks need to be duplicated and shipped to test administrators. After testing, the data need to be uploaded to a server.

In local server mode, the testing program is delivered via a local server that is located in the school. Computers need to be connected to the server, either through a wired network or through a wireless local area network (Wi-Fi), and a version of the test program supporting the local server mode needs to be installed. The test itself is run on the client server via a web browser. For this approach, access rights need to be granted and the browser that has to display the assessment needs to meet the standards defined by the study. Upon completion of the test, the data are uploaded by the test administrators.

10.7 Conclusions

The change from paper testing to computer-based assessment has far reaching implications for test design, implementation, and analysis, including test and item development, downstream procedures such as translation and layout control, test execution, scoring, data processing, and analysis. They are accompanied by changes in the workflow and the cost structure. The constant that remains is the importance to maintain the very high quality standards achieved in the paper-based test, and to improve the technology as far as possible. In particular, the preservation of trend metrics and the security of assessment material are of paramount interest. With the systems already in use, IEA has succeeded in keeping the path to computer-based testing on track without sacrificing quality standards, by introducing changes gradually and deliberately with appropriate monitoring of mode effects.

References

- Bundesministerium für Bildung und Forschung. (2019). *Mit dem Digitalpakt Schulen zukunftsfähig machen* [webpage]. Berlin and Bonn, Germany: Author. Retrieved from <https://www.bmbf.de/de/mit-dem-digitalpakt-schulen-zukunftsfahig-machen-4272.html>.
- Carstens, R., Brese, F., & Brecko, B. N. (2007). Online data collection in SITES 2006: Design and implementation. In *The second IEA international research conference: Proceedings of the IRC-2006, Volume 2: Civic Education Study (CivEd), Progress in International Reading Literacy Study (PIRLS), Second Information Technology in Education Study (SITES)* (pp. 87–99). Amsterdam, the Netherlands: IEA. Retrieved from <https://www.iea.nl/index.php/publications/conference/irc-2006-proceedings-vol2>.
- Fishbein, B., Martin, M., Mullis, I. V. S., & Foy, P. (2018). The TIMSS 2019 item equivalence study: Examining mode effects for computer-based assessment and implications for measuring trends. *Large-scale Assessments in Education*, 6, 11. Retrieved from <https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/s40536-018-0064-z>.
- Frailon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). *IEA International Computer and Information Literacy Study 2018 assessment framework*. Cham, Switzerland: Springer. Retrieved from <https://www.iea.nl/publications/assessment-framework/icils-2018-assessment-framework>.
- Herold, B. (2016, February 5). Technology in education: An overview. *Education Week*. Retrieved from <http://www.edweek.org/ew/issues/technology-in-education/>.
- IEA. (2020a). *TIMSS 2019: Trends in International Mathematics and Science Study 2019* [webpage]. Amsterdam, the Netherlands: IEA. Retrieved from <https://www.iea.nl/studies/iea/timss/2019>.
- IEA. (2020b). *ICILS 2023 flyer*. Amsterdam, the Netherlands: IEA. Retrieved from <https://www.iea.nl/publications/flyer/icils-2023-flyer>.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2016). *Methods and procedures in TIMSS 2015*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/publications/timss/2015-methods.html>.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 assessment framework* (2nd ed.). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://timssandpirls.bc.edu/pirls2016/framework.html>.
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. <http://timssandpirls.bc.edu/timss2019/frameworks/>.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). Take the ePIRLS assessment. In I. V. S. Mullis, M. O. Martin, P. Foy, & M. Hooper (Eds.), *ePIRLS 2016 International Results in Online Informational Reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://pirls2016.org/epirls/take-the-epirls-assessment/>.
- OASIS. (2008). *XLIFF version 1.2*. Retrieved from <http://docs.oasis-open.org/xliff/xliff-core/xliff-core.pdf>.
- Ramalingdam, D., & Adams, R. J. (2018). How can the use of data from computer-delivered assessments improve the measurement of twenty-first century skills. In: E. Care, P. Griffin & M. Willson (Eds.), *Assessment and teaching 21st century skills. Research and applications* (pp. 225–238). Cham, Switzerland: Springer.
- UK Department for Education. (2018). *New technology to spearhead classroom revolution* [webpage]. London, UK: UK Government. Retrieved from <https://www.gov.uk/government/news/new-technology-to-spearhead-classroom-revolution>.
- von Davier, M., Khorramdel, L., He, Q., Jeong Shin, H., & Chen, H. (2019). Developments in psychometric population models for technology-based large-scale assessments: An overview of challenges and opportunities. *Journal of Educational and Behavioral Statistics*, 44(6). DOI: <https://doi.org/10.3102%2F1076998619881789>.
- Wainer, H. (2000). *Computerized adaptive testing. A primer*. Abingdon, UK: Routledge.

Walker, M. (2017). Computer-based delivery of cognitive assessment and questionnaires. In: P. Lietz, J. C. Cresswell, K. F. Rust & R. J. Adams (Eds.), *Implementation of large-scale education assessments* (pp. 231–252). Wiley Series in Survey Methodology. New York, NY: Wiley.

Heiko Sibberns is a Senior Research Advisor at IEA. He has a teacher's degree for upper secondary education and extensive experience in the field of international comparative studies in education. He worked for many years at the data processing unit of the IEA Reading Literacy Study. In 1994, he became a co-director of the IEA Data Processing Center (later renamed IEA Hamburg), during which time he was responsible for all national projects conducted by IEA. In 2013, Mr. Sibberns became director of IEA Hamburg, a position he held until October 2019. Mr. Sibberns is a core member of the IEA Technical Executive Group, which provides guidance on the technical feasibility and best practice implementation of all IEA studies.

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.