HAL archives-ouvertes.fr

Computer-Based Large-Scale Assessments in Germany

Andreas Breiter, Lea Groß, Emese Stauke

To cite this version:

Andreas Breiter, Lea Groß, Emese Stauke. Computer-Based Large-Scale Assessments in Germany. Don Passey; Andreas Breiter; Adrie Visscher. 10th Next Generation of Information Technology in Educational Management (ITEM), Aug 2012, Bremen, Germany. Springer, IFIP Advances in Information and Communication Technology, AICT-400, pp.41-54, 2013, Next Generation of Information Technology in Educational Management. <10.1007/978-3-642-38411-0_4>. <hal-01468478>

HAL Id: hal-01468478

https://hal.inria.fr/hal-01468478

Submitted on 15 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution| 4.0 International License

Computer-based Large-scale Assessments in Germany

Andreas Breiter, Lea Maria Groß and Emese Stauke

Institute for Information Management, University of Bremen, Germany
abreiter@ifib.de

Abstract. In nearly all countries, technological changes in the field of digital media have raised great interest in procedures for testing students' performance by means of computer-based assessment. For extensive international tests (large-scale assessments) such as the Programme for International Student Assessment (PISA) or Trends in International Mathematics and Science Study (TIMSS), pilot studies with computer-based assessment were first carried out. In the context of national education systems, the question of feasibility plays a crucial role. Considering the framework conditions in Germany, scenarios for a potential nation-wide roll-out have been developed. Based on this, we evaluated the feasibility of technical and organizational factors.

Keywords. Computer-based assessment; large-scale assessment; feasibility study.

1 Introduction

With the emergence of international school performance studies (e.g. PISA, Progress in International Reading Literacy Study (PIRLS)), large-scale assessments (LSA) are frequently used as instruments for monitoring school quality and accountability. The tests can be used for different purposes, e.g. for generating data for educational policy-making, system monitoring, quality management for schools and indicators for measuring the performance of students. Thus they are strategic assets of empirical educational research and policy.

Referring to LSA, computer-based assessments are computer-based procedures that can be used for large-scale school performance studies, learning level surveys or the examination of educational standards [1 - 3]. They include software solutions as well as solutions with corresponding hardware platforms (e.g. mobile terminals such as a personal digital assistant (PDA), smart phones, or digital pens).

According to Hartig, Kröhne and Jurecka [4], computer-based assessments (CBA) run on a computer, but can also be performed on a local area network or based on Internet protocols. Many empirical studies have been conducted to analyse the difference between paper-and-pencil and computer-based testing – with varying results (e.g. [5 - 10]). From Hartig, Kröhne and Jurecka [4] we can conclude that computer-based and paper-and-pencil tests are comparable if the same conditions concerning the items and the test parameters are given. What is missing are empirical studies on the technical and organizational feasibility of transferring computer-based assessments to large-scale studies.

Economical advantages of testing have to be contrasted with the development of a suitable infrastructure. If adequate equipment already exists in the schools, the costs of test distribution and implementation, evaluation and feedback of results will be kept within reasonable limits. The question of testing fairness is brought up as another disadvantage of CBA. Discrimination, due to cultural, ethnical or gender reasons connected with the level of computer skills, is feared, which would question the validity of the test.

However, many advantages speak in favor of computer-based testing. Standardization of tests (using identical parameters, instruction and evaluation routines) improves quality criteria (objectivity, reliability and validity). Due to the information technology base, quick evaluation processes and an adaptive test procedure become possible.

Since the last PISA study, several experiments for representing the existing assessment process with computer systems have been carried out. Hence, any feasibility study has to be designed as a multiple cost-benefit analysis including suitability for the task, requirements of the information technology (IT) infrastructure concerning information security and data protection as well as operations, user support, usability and accessibility of software. For our study design, the leading question was to identify socio-technical scenarios, which are suitable for analysing the feasibility of computer-based tests in large-scale studies.

2 Scenarios for computer-based testing in large-scale assessments

For assessing the implementation potential of different international approaches in Germany, five scenarios were developed. The timeline was defined as the next ten years. The evaluation of the scenarios focussed on organizational implementation as a part of the social subsystem.

As a conceptual frame, we restricted the possible scenarios to the current activities in Germany for large-scale assessments. We considered VERA 3 and VERA 8, which are comparative assessments of student performance in all public schools in Germany in grades 3 and 8 in the main subjects of mathematics and reading. The tests are taken at the same time [11, 12]. Additionally, we took into account the assessment of educational standards identified by the Standing Conference of the Ministers of Education and Cultural Affairs of Germany (Kultusministerkonferenz). This is an assessment of a representative sample of students in grade 10 in the main subjects [13, 14]. Due to the sample size, the amount of simultaneous tests can be limited to 10,000 whereas the test formats VERA 3 and VERA 8 involve all students in these grade levels. The test setup always requires test administrators. The responsibility is divided between the organization of the test process and the management of the technological infrastructure, which can both be provided internally or externally.

In order to compare the demand for the IT network infrastructure such as bandwidth and server architecture, the potential number of participants (in parallel) needs to be calculated. As an assumption, we took the average number of classes per grade level and students per class. For VERA 3 in German primary schools, the average is three classes per grade level, with pupils evenly distributed across the grades and years. Per year, three or four classes of a grade are taught in parallel. The average class size is 22. As there is no central registry for IT infrastructure, we had to base the calculations on available numbers from 2008. Sixty-two percent of German primary schools (grades 1 to 4) were equipped with computer laboratories, and 82 percent with a stationary computer in each classroom. This can be summed up as nine students per computer on average [15].

For VERA 8, the assumptions are extended to secondary schools. In 2008, 98 percent of these schools were equipped with computer laboratories. There was one computer available for ten students on average [15]. Based on the assumption of four classes per grade level with an average size of 25 students, schools with a total number of students of 1,000 have the technical equipment for the implementation of the tests. The calculated numbers for the national tests on grade level 10 are identical (although it should be noted that values for all school types has been simplified; there certainly are smaller schools, which provide the required infrastructure to carry out computer-based tests, whereas some bigger schools do not).

The more students participating in the test, the higher is the bandwidth of the local area network needed as well as a faster connection to the Internet. On average it can be assumed that a download bandwidth of 100kb/s has to be provided per test person. The upload bandwidth can be disregarded inasmuch as a low amount of data has to be transferred back. Thus, in the case of 100 simultaneous tests per school, a connection of 16Mb/s should be provided at the location.

The scenarios developed were based on extensive literature reviews of existing approaches in other countries. We conducted expert interviews with representatives of the test administration and included questions on transferability to Germany. Additionally, several German experts were consulted.

2.1 Evaluation Criteria

For the evaluation of the scenarios, we developed a range of criteria, which were selected on the basis of an extensive literature study of international cases. We investigated the computer-based assessment methods of 16 different countries including the USA, Canada, the Netherlands and Denmark (e.g. [16 - 25]).

The first criterion was control over the process. This included the need for test centres to have an overview of the procedure from production to implementation to the return of the results (end-to-end) and the effect of this on security precautions and the assignment of responsibility in the case of problems. The second criterion was capacity and availability. This related to the amount of tests that could be taken simultaneously at one place and limitations, which arose concerning terminals, rooms, bandwidth, etc. The next aspect was standardization versus heterogeneity. How great is the reliance on hardware (memory, processor, graphic card)? Which are the (uniform) configurations at the location? For the analysis of the fourth criterion it was important to consider all legal requirements concerning information security (e.g. integrity, authenticity, availability, non-repudiation) and data privacy (e.g. active acceptance by users, control of access, admittance, availability and the principle of separation).

Furthermore, logistical aspects concerning preparation and implementation had to be taken into account. Questions like how are the devices ("virtual test booklets") transported to the test place and how are the devices securely placed, who ensures availability, who provides technical and administrative support at the location, who collects the devices and how are they transferred; these all arose in connection with this criterion.

One distinctive aspect between the organisational and technological scenarios addressed the qualification requirements for test takers as well as administrators and technical staff. This included programs for teachers and the test administrators as well as the provision of training materials for students. The last criterion was the cost aspect, including different cost categories of the scenarios, like direct costs (procurement, infrastructure, etc.) and indirect costs (support, insurance, etc.).

2.2 Scenario 1: Local IT-Infrastructure

In this scenario, the existing technical infrastructure of the respective schools was used, i.e. devices in computer laboratories or libraries. The tests were carried out in these rooms or in larger rooms such as the assembly hall or the gym. The software was provided either as a web-based application via the Internet or via a local mobile server using the existing local area network. Test administrators were needed for organization as well as IT experts for the provision, maintenance, setting up and restructuring of the infrastructure. The effort for the teachers stayed the same as in the case of paper-based tests. The identity check on students who took the test was carried out by local staff (teachers). Normally special measures (e.g. a check of an identity card) were not required.

Existing IT infrastructure was assumed to be heterogeneous. Schools worked with different client-server systems, the software differed and the quality of the existing hardware varied from school site to school site. Ninety-nine per cent of schools had some sort of Internet access. However, the availability of local area networks was significantly lower: in primary schools approximately 60 per cent, in secondary schools about 84 per cent. Currently, there were no reliable data about the bandwidth of the internal or the external connections. The majority of schools in rural regions were connected to the Internet via a sponsorship program of Deutsche Telekom AG with DSL quality (1-10 Mbit/s). However, many larger cities served as network providers to their schools and offered higher bandwidth (10-100 Mbit/s per school).

Concerning the organisation, the main focus (analogous to the paper-based test procedure) was on the training of the test participants with decreasing effort in the higher-grade levels. The qualification of teachers concerning the handling of digital media differed very much depending on school type. Technical administration by teachers could be found in most secondary schools; in primary schools there was no such support. Depending on the local school authority, additional technical support was offered. Hence, this scenario needed to rely on high-skilled technicians at the local level.

Another challenge was the preparation and qualification of the test administrators. The existing paper-based procedures for the three selected tests in Germany were not administered centrally. Hence, support structures had to be provided, which implied additional resources. Furthermore there was a need for qualified IT administrators in the schools or at the school district to ensure the availability of the IT infrastructure during the test phase.

Table 1. Evaluation of Scenario 1.

Criteria	**Strength (+) or Weakness (-)**
Control of the Complete Process	- Control of implementation remains the responsibility of the school - Depending on the local school authority, there are fixed infrastructure standards which have to be considered
Capacities and Availability	- Limited capacities, depending on school equipment - Not all rooms sufficiently equipped (especially in primary schools) - Bandwidth varying (local city networks/sponsoring)
Standardization vs. Heterogeneity	- Very heterogeneous infrastructure (operating systems, software, hardware) - Technical standardization not enforceable - Different configuration at every location
Information Security and Data Protection	- Realization requires much effort - Strongly dependant on local conditions - Know-how can hardly be expected on site (perhaps even opposes the security policy of the school authority)
Logistics – Preparation and Implementation	+ Easy organization if existing rooms can be used - Technical support has to be ensured (different situations in the schools) - Lacking technical know-how when devices are combined to an internal test-centre (e.g. assembly hall)
Qualification Requirements	- Comprehensive qualification required: Test administrators (technically) and teachers
Qualification Requirements for Students	+ Familiar environment, devices + Small exercise efforts
Costs	+ No procurement costs, no hire charges for the devices - Additional equipment for badly equipped schools required - Technician required when devices are combined to an internal test-centre (e.g. assembly hall)

2.3 Scenario 2: Test Centre

Universities provide test management for computer-based assessments in test centres. For example, the Brigham Young University in Utah, USA, provides its test center for any interested party. Forty computers are available for online testing.

This scenario deals with the use of adequately equipped rooms located in public or commercial institutions. Generally this scenario is conceivable in four different forms, involving computer laboratories:

- Of educational institutions (schools, universities).
- Of public institutions (centres for adult education, libraries).
- Of commercial service providers.
- Under school management.

The provision of the test environment can also take place in different ways. This includes the use of a web-based application on a central server via the Internet and the provision of this application on a mobile server using the local area network. Concerning staff, this scenario requires test administrators to secure the organizational process at the premises. Technical support can be provided by local IT administrators. Admission control is checked by mechanisms of authentication such as identity card control or knowledge control (username and password).

The implementation of the technical part of the scenario is the same for all test examples. The provision of sufficient workplaces in computer laboratories is a prerequisite for the simultaneous test of all participants. Based on the assumed average values, this means that a sufficient infrastructure has to be provided for 75 students in primary schools and 120 students in secondary schools. In metropolitan areas such as Berlin, 25,000 pupils would have to be brought to appointed locations to be examined at the same time. This appears to be an enormous expenditure. Thus the necessity of simultaneity is decisive. If time-shifted test formats are possible, this scenario would appear to be more realistic. Therefore we have to differentiate between the ways of testing; for a sample-based survey (such as the survey of national educational standards) it would be possible to find adequate test centres due to the limited number of test participants. As the tests are not taken simultaneously, a time-shift would be possible. Thus, not only the number of test administrators could be reduced, but also the search for test centres would be easier.

In the context of this scenario (except the provision of special rooms in a school) an internal network and broadband Internet can be expected as well as the protection against failure and the adequate provision of spare devices. More expenditure is generated by the organisation of the tests. Logistics especially (acquirement of capacities, allocation of resources, transport and supervision of the students) should be considered. Central training of test administrators in test centres could reduce the workload for teachers.

In summary, the logistic expenditure in primary schools (VERA 3) is highest with regard to transport and supervision of the participants. It decreases with increasing age and grade level of students (VERA 8).

Table 2. Evaluation of Scenario 2.

Criteria	**Strength (+) or Weakness (-)**
Control of the Complete Process	+ High degree of control of test production, implementation and return of results
Capacities and Availability	+ High availability + Reliable IT- and network infrastructure - Limited capacity (urban-rural divide)
Standardization vs. Heterogeneity	+ Homogeneous systems + Standardized infrastructure - Differences between locations
Information Security and Data Protection	+ Easy control (e.g. authentication) + Use of existing know-how + Established procedures
Logistics – Preparation and Implementation	- A lot of organizational effort: Reservation and allocation; transport of students (long distance) - Urban-rural divide
Qualification Requirements: for test administrators and teachers	+ Technical administrators on location + Probably organizational qualification requirements for teachers (only attendance)
Qualification Requirements for Students	- Unfamiliar environment and devices - No opportunity for exercise
Costs	- Rent - Transport of students - Staff costs on location

2.4 Scenario 3: Mobile test devices

The use of mobile test devices by test participants is logistically a digital version of paper-based test booklets. Test administrators have to transport the necessary infrastructure to the test locations and these are distributed to the test participants and collected after completion. Laptops, netbooks or other mobile devices with touch screen technology like tablets could be used as client systems. The test application is provided either by proprietary installation on the devices or by provision via the Internet or the use of mobile servers. When provision via the Internet is chosen, the local network has to be accessible. This requires corresponding network bandwidths. For this scenario it is necessary that test administrators provide technical and organisational assistance. The identity check of participants can be carried out by local staff.

From a technical point of view, the provision and acquisition of the required infrastructure is the critical success factor. If we refer to the numbers of 2008, on average 75 devices per primary school and 120 devices per secondary school have to be provided. Logistical issues especially concerning the delivery and collection as well as the subsequent use of the devices have to be considered. Additionally, the procedures for organizing the return of the test results have to be considered. When the mobile devices are used within the school, the test data have to be collected, saved and analysed centrally. This produces high technical and organisational expenditure.

The local technicians are responsible for the provision of the basic infrastructure. This includes the availability of enough electrical connections for recharging the client systems and the possible use of the existing network infrastructure including Internet access. In this case, Internet access and the local area network (LAN) in the schools have to be regarded as a second critical success factor. A wireless LAN especially has to be provided for the mobile devices. Currently, there are no data available on the diffusion of wireless LAN in German schools, but it can be assumed that wireless LAN is not available in all locations – in primary schools even less. To use this scenario for comparative studies, the local infrastructure and the Internet connections in primary schools have to be expanded.

From an organizational point of view, sufficient training of test administrators has to be ensured. Due to the variety of the client systems, training on the use of the software, the basic handling of the device and trouble-shooting has to take place. The training of test participants is similar to that in previous scenarios.

Table 3. Evaluation of Scenario 3.

Criteria	**Strength (+) or Weakness (-)**
Control of the Complete Process	+ High degree of control of test production, implementation and return of results - For network access, coordination with IT officer in charge required
Capacities and Availability	- Capacity depends on number of terminals + High degree of availability of proprietary software - Availability depends on local infrastructure in case of web-based software
Standardization vs. Heterogeneity	+ Homogeneous systems + Standardization largely possible
Information Security and Data Protection	+ Easy control (e.g. authentication) + Established procedure (closed systems)
Logistics – Preparation and Implementation	- High efforts for delivery and collection of the devices (especially with simultaneous tests) - Installation of devices requires technical know-how
Qualification Requirements: for test administrators and teachers	- Technical qualification of test administrators mandatorily required - Qualification of teachers required as regards content
Qualification Requirements for Students	- Unfamiliar handling and devices - Lacking opportunity for exercise
Costs	- Procurement (incl. spare devices) - Insurance - Transport - Technician and qualified test administrators on location

2.5 Scenario 4: Use-your-own-device

The concept “use-your-own-device” addresses the fact that more and more students have access to individual mobile computers. According to Medienpädagogischer Forschungsverbund Südwest [26], 80 per cent of young people in Germany aged 12 to 16 years have a computer or laptop of their own. Hence, test participants could use their own hardware. It would be essential to ensure that hardware is operational, i.e. a definition of minimal requirements is needed. This has to be done before tests are conducted, e.g., by a previous certification of the hardware by test administrators at the location. Apart from these requirements, this scenario is similar to Scenario 1, except that the school computers are replaced by private hardware.

As a second, more visionary sub-scenario, private computers could be used for testing in a familiar environment at home. In this case the test environment is provided with a web-based application via the Internet. In order to ensure equal access and use, test administrators are needed to train the participants. Technical support can be provided through a central point of contact (hotline, or service desk). The key success factor in this sub-scenario is authentication: it could be effected by knowledge control (password) and, if applicable, by biometrical features (fingerprints, or iris scan). Even using the new German electronic identity card could be considered. Furthermore, it is necessary to ensure that the participants work on the test individually and without help, according to the test instructions. This is possible by visual control via the network infrastructure and reliable applications (e.g. camera), though this will directly influence minimum requirements for the hardware. Additionally, the intrusion into private homes may have implications for privacy.

From a technical point of view, the question is if and how far the test participants have a suitable infrastructure available in their private environment. Although in 2010, 81 per cent of German households had a stationary or a mobile computer, this does not show whether this equipment meets the requirements for electronic testing. Furthermore, it was recorded that 73 per cent of the households had Internet access. The bandwidth varies between German regions (Statistisches Bundesamt Deutschland 2010) [27]. Additionally, access to computers and the Internet varies in relation to the income of the households. In a survey in 2008, only half of low-income families had a computer with Internet access (Statistisches Bundesamt Deutschland 2009) [28]. Given this, it is a legal question as to whether a test can be mandatory when privately owned devices are a prerequisite.

Another challenge is the standardization of the test conditions in terms of hardware and software in order to provide a fair test. It is hardly possible to standardize the types of privately owned devices as well as the bandwidth of the Internet connection.

Table 4 Evaluation of scenario 4

Criteria	**Strength (+) or Weakness (-)**
Control of the Complete Process	- The school and the test participants are responsible for implementation - In case of network access, cooperation with the IT officer is necessary
Capacities and Availability	+ All students can be tested simultaneously - Not all of the students own a device (spare devices are required) - Dependent on the infrastructure of the school (such as in the case of local infrastructure)
Standardization vs. Heterogeneity	- Very heterogeneous infrastructure (operating systems, software, hardware) - Technical standardization not enforceable - Each student has a different configuration.
Information Security and Data Protection	+ Easy to control access (authentication) - Impossible to control access (private devices)
Logistics – Preparation and Implementation	+ Low expenditure (effective time management essential) - Technical support has to be ensured
Qualification Requirements: for test administrators and teachers	+ No test administrators required - Comprehensive qualification required for school staff
Qualification Requirements for Students	+ Familiar environment + Students can practice in their home environment
Costs	+ No procurement costs - Additional costs for spare devices (such as insurance)

2.6 Scenario 5: Digital Pens

Technically, digital pens are a combination of an input device (pen and camera) and the corresponding digital paper. Digital pens look like conventional pens with an integrated digital camera filming the environment of the pen tip and thus registering where the pen is used. By means of corresponding software, forms are generated and printed as a greyish grid on paper. This is possible with a common laser printer. With the help of the grid, the camera and the special software, it is possible to identify the exact location of the pen tip. The user fills in the sheets in the conventional manner; the digital pen records the entries (pictures and text can be recorded as well). The content can be transmitted via Bluetooth or a docking station. The previously defined documents allow the data to be automatically entered into a database. This approach has been used already in higher education for small-scale assessments (e.g. [29, 30].

The advantages of digital pens are obvious: there is no need to train participants, the school staff and the test administrators. The pens are easy to deal with and relatively cheap (less than 100 Euros). Scientific formulas, continuous texts and handwritten commentaries can be recorded as well as audio annotations. The algorithm for character recognition works best with multiple-choice questions. Therefore the pens and the test booklets can easily be handed out and collected.

Up to now, however, the technology is not mature enough to deal with complex applications in real-life environments. The pens have been tested during elections or in health care scenarios [31] but could not meet high security requirements. Storing of changes on a page or returning to previous pages especially cannot be ensured. Although the devices are cheap, costs for licences for pattern generation on the paper are high. Distinctions are made between unique patterns and copied patterns. The former are reproduced for each page and therefore allow a direct assignment between the pen (user) and the page – and have to be paid per page. The latter can be copied as often as necessary, but then the user has to mark in the booklet on which page he/she is working. The first model is expensive; the second model fails in terms of usability.

All in all, despite all obvious advantages of the technology, it is not yet sufficiently developed for the use in computer-based assessment because of technical problems and resulting organisational difficulties.

Table 5. Evaluation of Scenario 5.

Criteria	**Strength (+) or Weakness (-)**
Control of the Complete Process	+ High degree of control of test production, implementation and return of results
Capacities and Availability	+ High degree of availability + No room limitations - Good printer required at each location
Standardization vs. Heterogeneity	+ Uniform system - Software not yet ready for the market
Information Security and Data Protection	+ Easy control (e.g. authentication) + Established procedure (closed systems) + Additional security through paper
Logistics – Preparation and Implementation	+ No difference to paper based tests, if printed at school + Supply and collection of devices
Qualification Requirements: for test administrators and teachers	+ Short technical introduction by test administrator required
Qualification Requirements for Students	+ Familiar handling + Exercises on paper (without DigiPen) possible
Costs	+ Procurement of terminals (incl. spare devices) - License costs

3 Conclusions

Some of the German experts we consulted said spontaneously that in their opinion computer-based large-scale assessments would not be feasible in Germany in the foreseeable future. Although they agreed on the potential it held for adaptive testing for high-stakes needs and for the motivation of students, they were sceptical as to whether the necessary cooperation between the federal authorities and the states and among the local authorities in the states could be effected, considering the federal education system. This reflects the special situation in the German education system with its separation between internal and external school issues. Neither the federal authorities nor the federal states can determine the IT infrastructure and provide technical support. They can only define requirements on the basis of curricula or via legislation, which have then to be implemented by the local school authority. If the states assign new tasks to local authorities, they need to provide the necessary funding. This requires complex negotiations between the state and the local education authorities. In some of the international cases, the implementation of computer-based testing initiated large-scale equipment programs for the schools. Whether such a strategy could be implemented in Germany is doubtful. Direct funding of schools by federal authorities is hardly possible and also requires a coordination process, as the local authorities will then have to pay all running expenses.

The results of the international comparison as well as the scenario-based assessment of the organisational framework conditions offer a more optimistic perspective. However, the positive effects of computer-based test procedures can only arise if basic questions of data privacy, information security, and accessibility can be clarified by software manufacturers, and if local education authorities responsible for providing the IT infrastructure are involved in the concept at an early stage.

For the use of IT infrastructures in schools (Scenario 1), two preconditions have to be fulfilled: first, schools have to be provided with appropriate equipment; and second, the local school authorities have to organise the required IT support processes. During the coming years, the IT infrastructure will increase and central IT services (e.g. utility computing in the cloud) will be available. These could then be used for CBA.

With test centres (Scenario 2), these problems could be solved, but here logistic requirements concerning transport of pupils and an urban-rural divide will arise. Due to high costs, it will hardly be possible for test organisers to build up test centres of their own.

Mobile test devices (Scenario 3) are also not suitable for large-scale tests, as they depend on the unreliable local infrastructure of the schools and are expensive.

The vision of “use-your-own-device” (Scenario 4) in the school context may be fulfilled within the next ten years. How far students will (want to) really have their own devices used, and if those will be suitable for computer-based tests, cannot be predicted. In any case, the test organisers, the schools and the local school authorities will have to bear any additional costs to make available the appropriate network infrastructure and the necessary technical support.

When the software for digital pens (Scenario 5) has reached a higher level of maturity, this will be a serious alternative, even if the cost saving compared to the paper-based tests would be rather low and adaptive test procedures could not be carried out. Thus, none of the five scenarios are realizable today.

Computer-based testing for large-scale assessments is a complex socio-technical system, which can only be realized by systematic planning and intelligent collaboration of test provider, software producers, local school authorities, state departments of education and schools.

Acknowledgements. We would like to thank Heiko Sibberns (IEA-DPC Hamburg), Peter Harych (ISQ Berlin-Brandenburg), Heiko Rölke (DIPF Frankfurt), Annemarie Klemm (SfBWG Bremen) and Mark Henning Tonne.

References

1. Hricko, M., Howell, S.L.: Online assessment and measurement: foundations and challenges. Kent State University, USA (2005)
2. Jurecka, A.: Introduction to the computer-based assessment of competencies. In Hartig, J. et al. (eds) Assessment of competencies in educational contexts. pp. 193-214, Hogrefe, Göttingen, Germany (2008)
3. Scheuermann, F., Björnsson, J. (eds.): The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-scale Testing. Office for Official Publications of the European Communities, Luxembourg (2009)
4. Hartig, J., Kröhne, U., Jurecka, A.: Anforderungen an computer- und netzwerkbasiertes Assessment. In Hartig, J., Klieme, E. (eds.) Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. pp. 57-67, Bundesministerrium für Bildung und Forschung, Berlin, Germany (2007)
5. Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., Yan, F.: Does it Matter if I Take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP. Journal of Technology, Learning, and Assessment, 6 (9), 283-294, (2009)
6. Clariana, R., Wallace, P.: Paper-based versus computer-based assessment: key factors associated with the test mode effect. British Journal of Educational Technology, 33 (5), 593-602, (2002)
7. Mead, A.D., Drasgow, F.: Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. Psychological Bulletin, 114, 449-458, (1993)
8. Russell, M.: Testing on computers: A follow-up study comparing performance on computer and on paper. Education Policy Analysis Archives, 7 (20), online, no page numbers, (1999)
9. Scott, C.B., Jared, M.C.: The impact of computer versus paper-pencil survey, and individual versus group administration, on self-reports of sensitive behaviors. Computers in Human Behavior, 24 (3), 903-916, (2008)
10. Horkay, N., Bennett, R.E., Allen, N., Kaplan, B., Yan, F.: Does it Matter if I take My Writing Test on Computer? An Empirical Study of Mode Effects in NAEP. Journal of Technology, Learning, and Assessment, 5 (2), 4-48, (2006)
11. Helmke, A., Hosenfeld, I.: Vergleichsarbeiten - Kompetenzmodelle - Standards. In Wosnitza, M., Frey, A., Jäger, R.S. (eds.) Lernprozesse, Lernumgebungen und Lerndiagnostik. Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert. pp. 56-75, Verlag Empirische Pädagogik, Landau, Germany (2004)
12. Hosenfeld, I., Schrader, F.-W., Helmke, T.: Von der Rezeption zur Ergebnisrückmeldung: Leistungsevaluation im Spannungsfeld von System-Monitoring und Schulentwicklung. In Hosenfeld, I., Schrader, F.-W. (eds.) Schulische Leistung. Grundlagen, Bedingungen, Perspektiven. pp. 289-313, Waxmann, Münster, Germany (2006)

13. Hartig, J., Klieme, E., Leutner, D.: Assessment of competencies in educational contexts. Hogrefe & Huber, Toronto, Canada (2008)
14. Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Vollmer, H.J.: Zur Entwicklung nationaler Bildungsstandards - Eine Expertise. Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt, Germany (2003)
15. KMK.: Dataset - IT-Ausstattung der Schulen. Schuljahr 2007/2008. Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, Bonn, Germany (2008)
16. Bennett, R.: How the Internet Will Help Large-Scale Assessment Reinvent Itself. Education Policy Analysis Archives, 9 (5), online, no page numbers, (2001)
17. Erickson, G.: National assessment of foreign languages in Sweden. University of Gothenburg, Gothenburg, Sweden (2009)
18. Klinger, D.A., DeLuca, C., Miller, T.: The Evolving Culture of Large-scale Assessment in Canadian Education. Canadian Journal of Educational Administration and Policy, 76, 1-34, (2008)
19. Levin, D., Fletcher, G., Chau, Y.: Technology Requirements for Large-Scale Computer-Based and Online Assessment: Current Status and Issues: State Educational Technology Directors Association (SETDA), Washington DC (2011)
20. Sandene, B., Horkay, N., Bennett, R.E., Allen, N., Braswell, J., Kaplan, B., Oranje, A.: The National Assessment of Educational Progress Online Assessment in Mathematics and Writing: Reports From the NAEP Technology-Based Assessment Project, Research and Development Series. US Department of Education, USA (2005)
21. Turmo, A.: Computer-based Assessment of Science (CBAS) - Empirical Results from Norway. University of Oslo, Oslo, Norway (2006)
22. Vernooy-Gerritsen, M.: The CITO Examination System. Paper presented at the International Association for Educational Assessment, Singapore (2006)
23. Virginia Department of Education: Statewide Web-based Standards of Learning Technology Initiative. Virginia Department of Education, Richmond, VA (2009)
24. Wandall, J.: National Tests in Denmark - Cat as a Pedagogic Tool. Journal of Applied Testing Technology, 12 (3), online, no page numbers, (2011)
25. Wiegers, J.: E-Assessment in the Netherlands, Innovations for the 21st Century. Developments Since 2000 Concerning Computer Based and Computer Assisted National Examinations for Dutch Secondary Education. Paper presented at the IAEA 2010, Bangkok, Thailand (2010)
26. Medienpädagogischer Forschungsverbund Südwest JIM 2011: Jugend, Information, (Multi-)Media. Basisstudie zum Medienumgang 12- bis 19-Jähriger in Deutschland. Medienpädagogischer Forschungsverbund Südwest, Stuttgart, Germany (2011)
27. Statistisches Bundesamt Deutschland 2010 (2010)
28. Statistisches Bundesamt Deutschland 2009 (2009)
29. Brown, S., Glasner, A.: Assessment matters in higher education: choosing and using diverse approaches. Open University Press, Buckingham (1999)
30. Koile, K., Chevalier, K., Rbeiz, M., Rogal, A., Singer, D., Sorensen, J., Wu, K.: Supporting Feedback and Assessment of Digital Ink Answers to In-Class Exercises. Paper presented at the AAAI 2007. Vancouver, BC, Canada (2007)
31. Schiavenato, M.: Digital pen and paper: a review of the technology and its potential application in healthcare. Online Journal Of Nursing Informatics, 10 (1), online, no page numbers, (2006)