COMMUNITY-LEGITIMATED e-TESTING
A Basis for a Novel, Self-Organized and Sustainable (e)Learning Culture?

Fritz Nestle

Institute of Mathematics/Informatics, University of Education, Reuteallee 46, D 71634 Ludwigsburg, Germany

Nikolaus Nestle

Institute of Condensed Matter Physics, TU Darmstadt, Hochschulstraße 7, D-64289 Darmstadt, Germany
BASF SE Ludwigshafen, D-67056 Ludwigshafen, Germany

Abstract: Based on the assumption that educational standards can be operationally defined by pools of specific testing items properties of such item pools are discussed. The main suggestion of the paper is that pools of testing items defining a standard should be free accessible in internet, that they provide immediate feedback in form of scores and that certified results should be equivalent to results of classroom work. For the development of the item pools, web-2.0-type methods can be much more effective than closed expert groups and item evaluation by statistic methods.. Finally the consequences of such transparent community-legitimated standards for the future role of teachers and future forms of learning environments are discussed.

1 e-LEARNING AND e-TESTING - THE PRESENT SITUATION

Present attention of both teaching professionals and the general public towards e-learning and e-testing seems somewhat unbalanced: On google, about 200 times more hits are found for e-learning than for e-testing. This may lead to the conclusion that there is much more interest into tools to support learning than into ways to test the success of learning and the actual topics that should be learned. Norms and standards in education are even harshly criticized as "teaching to the test" while autonomous "learning to the test" may be one of the most important "soft skills" in the future.

The established way of learning in secondary education is still the classical teacher-in-classroom paradigm in which both the presentation of knowledge and the control of learning progress are done as a more or less not standardized "batch process" with a groop of 10 to 50 learners. Even if there is a well-thought syllabus behind this learning model, both learning and testing items and the respective assessment often seem arbitrary and planless to the learners – and there's quite a bit of

empirical evidence that this impression is not even wrong (European Commission, 2008).

In many countries such as Germany, sucess in this kind of arbitrary learning and learning control environment has a great impact on the learners' future development such as admission to university studies. The school system's monopoly on this is almost complete and unchallenged – hardly an egalitarian and democratic access to educational opportunities. Recently, the school system's absolute power in Germany was even increased by introducing a number of required classroom hours in secondary school for university access (EQF, 2008; KMK, 2006) – instead of introducing a catalog of required qualifications for university access. In other countries such as Finland, the US or Brasil, a common standard of competences for university access is either secured by central tests for graduation from secondary schools or by University entrance exams that my be run nationally (such as the SAT in the US) or by the individual universities. In such exams, competences are tested instead of the way the candidate has had to learn these competences. If the testing criteria are transparent, this approach offers much better chances for equal opportunities than the school-based model as

preparation for the exams may also be done by self-organized learning and does not completely depend on the success in a school system with arbitrary learning and testing processes. (The existence of such external exams nevertheless may even lead to the establishment of special schools that aim at a good preparation for the exams. As such schools tend to be costly private schools the failure of the public school system to prepare for the entrance exam may again lead to seriously unequal opportunities; this is for example the case in Brazil.)

2 THE SUCCESS OF NORMS AND STANDARDS IN INDUSTRY

In the first half of the last century, norms and standards have been developed in many fields of industrial goods (e.g. 1917 the DIN norms in Germany, international norms: ISA since 1926 and 1947 ISO norms supported by the United Nations). In addition to basic engineering norms for material properties or threads of screws also more abstract issues such as the classification of books and other printed products (ISO 2108 covering the International Standard Book Numbering ISBN), layout of documents (ISO 2145 Numbering of divisions and subdivisions in written documents) and a multitude of other even more complex testing and certification problems are internationally standardized. The procedures for the creation of ISO norms consist of many different steps involving experts and committees to ensure norms and test methods based on sound expertise and with broad acceptance in the communities affected.

Norms and standards improve or even enable intersubjective comparability of industrial goods. Like that, they also provide a basis for exchange of manufactured goods over large distances as customers can expect well-defined product quality and properties. This holds true both for simple goods as shoes as well as complicated manufacturing equipment. In both cases, standardization doesn't mean uniformity: rather, the customer has a wide choice between different products and is not forced to accept a dedicated one-off.

In the educational sector, generally accepted norms are still an exception. Examples are tests like TOEFL or SAT which are offered by ETS in the United States and even frequently used by universities outside the US in order to assess e.g. a candidate's proficiency in English.

In the absence of transparent standards learning progress and even more so learning results cannot really be objectively judged. Therefore, the learner is left with very limited possibilities to assess his or her learning progress. This, in turn, makes self-organized learning almost impossible. Nevertheless, self-organized learning is often more effective than school-based learning which is also often not very sustainable. A strong focus on exams obtained by school-based learning is therefore especially difficult for mature students and other learners with unconventional educational biographies who might find themselves shut-out from educational opportunities despite better qualifications than the graduates of the conventional school system.

In countries like Germany where a poorly standardized school graduation exam is the main requirement for access to university education, universities often organize pre-term courses to catch up with elementary deficits from school education – regardless whether the required skills were not taught at all or not taught sustainably. For school graduates who go for vocational training instead of college education the regional Chambers of Commerce and Industry (IHK) organize entry exams which are not standardized even on the state level and where the test items are usually not made publicly transparent.

Like this, there is the paradox situation that the skill levels of new employees are much less standardized than more or less everything else in industrial production. The development of community-legitimated sets of subject-specific transparent tests may offer a way out of this paradox situation and encourage self organized (life long) learning (EQF, 2008).

2.1 Definition of Educational Standards by Subject-specific Test Modules

Almost all cognitive qualifications which are taught and learned on the secondary school level can be tested by appropriate test items. It even may be much more appropriate to define educational standards by means of classes of test items than by abstract verbal descriptions such as those presently used in Germany's so-called national educational standards (KMK, 2003).

Classical test psychology is based on tests which consist of items which are chosen from an item pool according to a statistical test model (e.g. the Rasch model). The validity of tests and test items is assessed on the basis of the test model. The strict

observance of this statistical model may lead to didactically problematic results as „too difficult" or „too simple" items may never make it into the item pool as they are eliminated during the pilot runs of the test even if they should be included from the didactical point of view. For example, the multiplication with 0 was eliminated from a mastery-learning-oriented elementary multiplication test as it was wrongly answered by too many learners who obviously were not taught it along with the remaining multiplication rules. Like that, didactical failures dictate the item pool instead of using a didactically sound item pool to identify possible deficits in elementary multiplication completeness.

Another problem with the item pool in standard test psychology is the practice to keep the item pool itself secret in order to prevent candidates from "learning the right answers" to the item pool. While this may be a good practice in actual psychological tests such as IQ tests, this secrecy of the item pool is also uncritically taken over for tests in secondary education where learning is the actual goal. Keeping learning objectives secret is therefore highly contraproductive. This is especially true in those fields in which test items with numerical answers can be created automatically with random parameters so that it is impossible to "learn" the results for individual test items (Nestle et al. 2007). We therefore suggest an alternative use of item pools in secondary education and entry level college education: Open, community-legitimated item pools that serve both as a operational definition of an educational standard and as a transparent pool of test items.

The production and maintenance of open item pools can be organized in a similar way as open source software development or web 2.0 projects such as wikipedia: These projects are not run and organized by a closed small panel of experts but by an open community that produces, discusses and evaluates the contributions to the project.

Based on this approach, wikipedia has developed within a few years into a comprehensive and multilingual encyclopedia of a quality that meets or even exceeds the standard set by established encyclopedias such as the Encyclopedia Britannica or the German Brockhaus. The content in wikipedia is not only freely available to all users but all users can modify the entries or initiate new entries when they consider the available information incorrect or incomplete. All changes are documented and other users can control, discuss or reject them. For controversial cases, the community has developed moderation and mediation procedures. This system has proven to be remarkable stable against biased manipulations even by powerful players such as big companies. Wikipedia therefore provides an example for the successful production, maintenance, evaluation and legitimation of content and knowledge by an open community. The processes of Wikipedia may serve as a model for the generation of a comprehensive item pool.

Similarly, open source software projects such as Linux or OpenOffice provide successful alternatives to commercial software packages and even offer better stability that those products. A major reason for this is the large community that contributes in identifying and fixing poorly functioning parts of the software.

A common feature in open source software and web 2.0 projects such as wikipedia is the free accessibility of the software and/or he content under the GNU public licence. Nevertheless, this public accessibility does not necessarily mean that all activities in the field are done in a non-for-profit parallel world. Donations to the developer communities allow the establishment of professional management structures for most of the larger projects and services around the programs also might be profit-oriented (e.g. installation support or customization of the programs or user trainings). Similarly, certified qualifications based on community-based standards may be offered for profit. This could provide a sound basis for funding the activities and infrastructure needed to support the creation of such communities and standards.

Furthermore, courses to support learners in their studies for a specific test module may be offered on a commercial basis. However, the existence of such courses would be an indicator of the failure of public schools to support the community legitimated standards and an adaptation of the public school system to the standards would be a more desirable development. Nevertheless, the experience with university entrance tests in countries like France and Brasil seems to indicate that the existence of such tests also opens up good business opportunities for preparatory schools.

2.2 The Vision: Bringing Community-legitimated Standards to Work

In order to make community-legitimated standards really work, some requirements concerning the nature of the item pools, the accessibility and the

documentation of work with the item pool must be fulfilled.

2.2.1 Test Items

Main requirements for test item pools accessible via the internet should be the following:

- Sufficiently large item pools: If test items are randomly chosen from a large item pool, learning individual test item results by heart is no longer realistic
- Items must allow automatic evaluation of user input. In most cases this doesn't actually restrict the formulation of test items. Possible forms of test items may be cloze tests, multiple choice answers or restricted free answers (e.g. numerical values). The correct answers may be static or dependent on random generated numerical or string values. The latter case is preferable. In addition to the correctness of the user's results, also the time needed for working through the test items is recorded and a score is calculated from the time and correctness. Various individual and general score lists for each set of items are kept and published in order to provide the user with feedback both on his or her own learning progress and the comparison to the general learners' community.
- Commenting and rating possibilities for each item. The user rating for the items (along with a teachers' and graduates' rating) will be used to evaluate and legitimate the items. Compared to the statistical model, this approach is more transparent and flexible.

Authoring systems that support all these requirements are available. An example is eExercise (Nestle et al., 2007).

2.2.2 Access

The access to the item pool may either be free (if sponsoring/advertisements or public funding is available), or communities may charge a small annual fee for the access. Various types of access to the item pool must be provided:

- Anonymous access with the possibility to view and work through test items, receive feedback, comment test items and suggest new test items. Under this kind of access, learners can train and test their learning level, and the general public can gain an insight into the respective standard.

- Access for certified testing under registration. Certified tests are to be done in an environment where the identity of the candidate and the independent work through the test can be appropriate checked and documented. The certified success in the test must be considered equivalent to the respective qualification from classical schooling.
- Registered access to modify test items, suggest new test items or test topics or delete or regroup test items. This kind of access must require a proof of qualification before registration. Modifications and deletions of test items are provisional for a fixed period in which other users with the same rights can contradict. If no contradiction occurs, the deletion or modification is permanent.

All accesses to the item pool (anonymous, registered candidate for certified exam or qualified user modifying the item pool) are documented and archived. This is a standard procedure in web 2.0 projects such as wikipedia, too. This offers on the one hand novel possibilities for learning research (Nestle et al., 2007) and provides a possibility to identify destructive accesses to the item pool that may need action such as retraction of access rights.

2.3 Consequences of Community-legitimated Standards

The existence of a transparent data base of test items has consequences for all actors in this educational system as well as for the general public.

2.3.1 Consequences for Teachers

Presently, teachers are faced with a Herculean cognitive and emotional challenge: Preparing lessons and keeping their own skills in the subject up to date by continuous education, permanent self evaluation und evaluation of student's learning performance and providing education in soft skills to the students.

Evaluation their students' learning progress takes often more than 20 % of teachers' working time. Nevertheless, this work is done with a giant uncertainty. For example different teachers may judge student's work according to enormously different standards. Therefore, the present teacher-and-examiner classroom model allows no equal opportunities for learners. Rather, there is a lot of randomness in the assessment of students' progress that should not be accepted by a democratic and egalitarian society.

If a community legitimated standard for qualifications in a certain field is available, this will make the teacher's role in this field much more simple. First of all, the teacher won't need to invest a lot of time into the development and evaluation of written exams any more. Furthermore, the teacher's work will become independent from parents' pressure to give better marks to the students as the standard for rating students' performance will become external and transparent. Instead of the teacher-examiner role conflict, the teacher can concentrate on the original art of teaching: helping students learn and advising and challenging them in the respective subject. Students may chose to accept this help from the teacher or opt for self-organized learning. Their performance in the test will be dependent on their learning progress and not on their relationship to the teacher or the teacher's subjective standard for the exams.

2.3.2 Consequences for Students

Students can check their learning progress without observation by the teacher or their peers. Like this, feedback on their performance will no longer be a possibly fearful and embarrassing classroom experience. Furthermore, students are free to choose their own learning pace instead of being forced into the rhythm of their class. Such self-organized learning is much closer to most learning situations in professional or other non-school contexts. Therefore, open standards are much better to develop self-organized learning as a key soft skill than usual classroom learning.

Similarly, students can also choose more freely the focus of their learning efforts than in classroom learning bound to a teacher's (maybe outdated) interpretation of a (maybe outdated) syllabus. The possibility to obtain certified proofs of qualifications outside standard school subjects will be a great incentive to students to learn special skills outside traditional curricula. The existence of such certification possibilities may even trigger the demand for schools to adapt their curricula to qualifications that many students or employers are interested in. Like that, the selection of subjects at schools will become much closer coupled to real life than it is today.

2.3.3 Consequences for Society as a Whole

Certified exams based on community-legitimated standards will provide less biased and more transparent information on a persons' skills and qualifications than grades from traditional schooling.

Item pools on specialized subjects provide a new flexibility for the establishment of qualification profiles. Each social or economic group may provide new educational objectives without fighting them tiring processes in the administration of the public school system. And new subjects can be chosen by students who are interested in them instead of forcing all students through curricula with very limited possibilities for choice.

There is a lot of matters that are regularly recommended for introduction as novel subjects at school: economics, health education, psychology, … Introducing all those subjects compulsory for all students would lead to an unacceptable swelling of classroom hours. Offering those subjects via transparent item pools, allows students to decide whether they are interested in the respective qualifications or not. If there's a real need for them from universities' or employers' point of view, large numbers of students may choose them and maybe even trigger the demand for a support of those subjects in the public school system.

3 ONLINE GAMES AS A MODEL FOR THE FORMATION OF ONLINE COMMUNITIES

Children up to age 12 are naturally eager to learn. To conserve this motivation in classroom learning is a still unsolved problem. By contrast, other contexts that allow self-organized and active participation result in the formation of large communities. In addition to somewhat nerdish activities as open source programming, online computer games have formed large communities in which many people around the globe invest a lot of time and creative energy.

Such games often win the permanent battle for attention and "learning" over other activities such as studies, work or physical exercise. Like that, they also pose a serious hazard to physical, mental and social health of persons involved too heavily into gaming. Nevertheless, they provide an interesting example to study the formation and dynamics of online communities and their certification systems.

An example: In less than four years the game „World of Warcraft" has grown into a worldwide community of more than ten million people. These gamers pay a substantial monthly fee – and most of them spend 20 to 50 hours per week playing the game. This time is lost for more productive occupations as learning, music, other cultural events, sports and so on.

A google search on 10. Januar 2009 produced more than 600 000 hits to 'play world of warcraft', but only 300 000 contributions to 'learn mathematics'. ('learn math' with about 1,2 millions of contributions lies in the same order). Obviously, the attractivity of just this one online game is comparable to a more than 2 000 years old cultural tradition.

The similarity between 'World of Warcraft' (WOW) and our vision of community-bases educational standards is the production and rating of test items by a community. In WOW, the gamer can choose and solve 'quests' to gain status in the WOW community, and they also can develop and suggest new quests themselves. The gamer's status in the community depends on his gained 'level' and his equipment. Solving quests is rewarded with level and equipment. The rewarding system is transparent: Gamers know in advance which quests will get them which rewards. It's evident that WOW satisfies elementary human desires which classic learning at school does not serve. There is no problem with addiction to learning, but online games may cause serious addiction. Communities supporting open educational standards are probably not going to be addictive, nevertheless, they may still be more motivating than classical schooling for some learners.

4 ON THE WAY TO A WIKI-TESTING COMMUNITY?

At the moment community-based educational standards on internet are a mere vision. It may share the fate of most visions and remain an utopian dream. Alternatively, it may also grow into reality. A possible way to achieve this may be the integration of community based open standards for automatic exams into the Wikipedia family (maybe as a workhorse for organizing exams and training opportunities in Wikiversity).

Community-generated online content was not invented by Wikipedia. Already 1993 Richard Stallmann, founder of the GNU project and one of the godfathers of open source programming, wrote:

„...Since we hope that teachers and students at many colleges around the world will join in writing contributions to the free encyclopedia, let's not leave this to chance. There are already scattered examples of what can be done. Let's present these examples systematically to the academic community, show the vision of the free universal encyclopedia, and invite others to join in writing it.

Courses in the learning resource are a generalization to hypertext of the textbooks used for teaching a subject to yourself or to a class. The learning resource should eventually include courses for all academic subjects, from mathematics to art history, and practical subjects such as gardening as well, to the extent this makes sense. (Some practical subjects, such as massage or instrumental ensemble playing, may not be possible to study from a "book" without a human teacher—these are arguably less useful to include.) It should cover these subjects at all the levels that are useful, which might in some cases range from first grade to graduate school."

It lasted only some years until Wikipedia has accumulated a volume and quality that has already surpassed that of venerable traditional encyclopedias filling whole bookshelves. Extending the presentation of knowledge online to training and testing online with immediate feedback is a logical next step building on Stallmann's ideas.

e-testing based on community-legitimated standards can help to advance the idea of e-learning and also help to provide a novel basis for defining educational standards. It may start with a handful of subjects – either supported by a platform like wikipedia or fostered by motivated educators in established educational institutions who are interested in basing exams and training for their students on a transparent pool of test items.

We concede that presently available resources to create e-testing modules still not have reached the usability comfort levels known from contributing to the Wiki world, but this is only a problem of time and cooperation. Who will help?

REFERENCES

European commission communication COM(2008) 425: http://etuce.homestead.com/com425_en_2008.pdf

EQF (2008) European Qualifications Framework http://ec.europa.eu/education/lifelong-learning-policy/doc44_en.htm

KMK 2003 (German National conference of State Ministries of Education) Bildungsstandards für den Mittleren Bildungsabschluss (Jahrgangsstufe 10) *Agreement on 4.12.2003*.

KMK 2006 (German National conference of State Ministries of Education) *314th meeting 1./2.6.2006*.

Nestle, N., Gädke, A., Enders, J., 2007. eExercise: Möglichkeiten zur Beobachtung individueller Lernstrategien von Studierenden in einer freien, webbasierten Aufgabenplattform *Tagungs-CD Didaktik der Physik 2007*. Lehmanns, Berlin.