ProNIFA: A Tool Supporting Semi-Automatized Formative Assessment in STEM Teaching

Michael D. Kickmeier-Rust and Dietrich Albert
Cognitive Science Section, Knowledge Management Institute
Graz University of Technology, Graz, Austria

Abstract - There is no doubt that modern technologies significantly changed the educational needs and practice; the impact of the continuously evolving technology will increase the demands of teachers and learners. NEXT-TELL is a multidisciplinary European project that attempts to develop a European infrastructure to facilitate and broaden the use of modern information and communication technologies in order to tackle the educational challenges of today and tomorrow. A significant aspect in this context is a strong focus on modern, formative, non-numerical approaches to the assessment of learning in particular in STEM areas. In this paper we outline the challenges we see in the modern classrooms and we provide an introduction to a formal, cognitive theory that supports modern ideas of assessment and student appraisal. In addition, we present a software tool, named ProNIFA, that enables teachers and learners to utilize the strength of the formative assessment in an effective and semi-automatized way.

Keywords: 21st century education; technology use in classrooms; formative, non-numerical, probabilistic assessment; Competence-based Knowledge Space Theory

1 21st Century Education

21st century education clearly is a big buzzword in today’s media. The new millennium is accompanied by substantial technological evolutions; we became a highly diverse, globalized, complex, real-time media- knowledge-information- and learning society. Since the 1990s, the progress of media and technology was breath taking; during these one or two decades, we were facing the rise of a serious and broad use of computers at home (although the development started earlier, of course), the rise of the internet and how it revolutionized our society. We faced the spread of mobile phones and their evolution from telephones to omnipresent computer and communication devices; we see spread of mp3, twitch speed computer games and TV shows. We saw how our world got closer by changing the bridges over continents and oceans from 56k wires to hyper speed fiber glass networks.

But what does this mean for educational systems and the way our children learn and what they learn? Today’s kindergarten kids will retire in 2070. Facing the pace of technological and societal changes and demands, we cannot predict what knowledge will be required in such a “far” future. But we are in charge to equip our children with the abilities and backgrounds to survive in that world. Our students are also facing many important emerging issues such as global warming, famine, poverty, health issues, a global population explosion and other environmental and social issues. These issues lead to a need for students to be able to communicate, function and create change personally, socially, economically and politically on local, national and global levels.

Formative assessment, defined as a bidirectional process between teacher and student to enhance, recognize, and respond to the learning, is one vital aspect in addressing those challenges Formative assessment is considered a promising approach to enable 21st century teaching since it potentially promotes self-reflection and self-directed learning processes and, more importantly, it facilitates the integration of new subject-specific knowledge into the student’s existing knowledge network. It also helps adapting the teacher the educational processes to the individual needs and, therefore, making formal education more effective and also more enjoyable. At the moment, however, formative assessment is a time consuming interactive teacher-student communication process. The teacher must carefully monitor problem/task solution processes of each individual student, the teacher must understand how individual students think and learn, and the must help them to overcome conceptual difficulties, again on an individual basis. Considering today’s situation and the daily routines in schools this desirable feedback process oftentimes drowns in limited time and resource context conditions. In this paper we want to introduce NEXT-TELL (www.next-tell.eu), a European initiative addressing the challenges of future school education. The major aim of this project is to support teachers in using modern technologies in the classes, to link various technologies with each other, and to benefit from synergies. NEXT-TELL also aim at supporting formative assessment processes and the aligned teaching with smart planning, assessment, and teaching technologies. Secondly, we want to introduce a formal, learning theoretic approach to non-numerical probabilistic assessment to support a formative appraisal of students and ideas of evidence-centered appraisal.
2 Numerical vs Non-Numerical Test Theories

Presently, western school systems are largely characterized by described the broad and rich spectrum of students’ abilities and knowledge by single (numerical) value, the grade. This holds true on a small scale (e.g., on the level of tests or single examinations) and this holds true on a large scale (e.g., on the level of a final grade for a course). This approach, however, cannot express what students really and exactly can/know and what they cannot do/do not know satisfyingly. A good example for the weakness of the approach is the I.Q. (intelligence quotient), which attempts to characterize all the various abilities, strength and weaknesses of a person in many categories and disciplines (math, language, cognition, memory, etc.) with a single numerical value, lately.

As summarized by Falmagne, Cosyn, Doignon, and Thiéry [1], the origin of this popular test theoretical approach lies in the developments in 19th century physics. The frontiers between natural sciences, physics, psychology, and medicine we much more blurred as they are today and disciplines like “anthropometry” occurred, the “art of measuring the physical and mental faculties of human beings”. Prominent proponents were Francis Galton, William Kelvin, or Carl Pearson (Figure 1). The predominant tenor was, if you cannot measure it, it is not science. Kelvin, for example, said “If you can’t assign an exact numerical value, express it in numbers, your knowledge is of a meager and unsatisfactory kind”.

The problem of this “charming” approach gets very clear by the following example (taken from [1]). Imagine an athletic quotient (A.Q.) and imagine 3 distinct sportsmen, the basketball player Michael Jordan, the tennis player Roger Federer, and the golf player Tiger Woods. All of them are, apparently, excellent athletes and deserve a high A.Q. but they are very different in their individual abilities. Depending on which scale one would take to measure the A.Q. (maybe the jumping height, maybe stamina?), the three sportsmen would receive very different appraisals. Now, an obvious response would be to compute a quotient of all the three men’s main strength. But – what if another sportsmen is added, e.g., a long distance runner, a downhill racer, a chess player? For each distinct sportsman another measuring dimension could be added. And in the end, the resulting quotient would be able to precisely describe the real characteristics and strength of each individual sportsman.

An empirically well-validated approach to CbKST was introduced by Klaus Korossy (1997, 1999); basically, the idea of the Competence-Performance Approach (CPA) is to

The same problem occurs for formative assessment and the ECAAD approach, specifically in a limited resource classroom scenario. An approach for a fairer and more precise appraisal comes with ideas of non-numerical test theories. With the rise of mathematical combinatorics and with the rise of powerful computer technologies the more demanding approaches to precisely describe the various abilities of a student along multiple dimensions got in the focus for researchers and educators. In parts, the basic idea found its way into class rooms, for example, when students are described in narrative text instead of a single grade in the report card.

The approach we want to utilize for NEXT-TELL – in form of smart software services – is based on the probabilistic, non-numerical method of Competence-based Knowledge Space Theory (CbKST).

2.1 The structure of Knowledge

CbKST originates from KST established by Jean-Paul Doignon and Jean-Claude Falmagne [2, 3], which is a well-elaborated set-theoretic framework for addressing the relations among problems (e.g., test items). It provides a basis for structuring a domain of knowledge and for representing the knowledge based on prerequisite relations (see Figure 2 for an example). While KST focuses only on performance (the behavior; for example, solving a test item), CbKST introduces a separation of observable performance and latent, unobservable competences, which determine the performance.

An empirically well-validated approach to CbKST was introduced by Klaus Korossy (1997, 1999); basically, the idea of the Competence-Performance Approach (CPA) is to

The same problem occurs for formative assessment and the ECAAD approach, specifically in a limited resource classroom scenario. An approach for a fairer and more precise appraisal comes with ideas of non-numerical test theories. With the rise of mathematical combinatorics and with the rise of powerful computer technologies the more demanding approaches to precisely describe the various abilities of a student along multiple dimensions got in the focus for researchers and educators. In parts, the basic idea found its way into class rooms, for example, when students are described in narrative text instead of a single grade in the report card.

The approach we want to utilize for NEXT-TELL – in form of smart software services – is based on the probabilistic, non-numerical method of Competence-based Knowledge Space Theory (CbKST).

2.1 The structure of Knowledge

CbKST originates from KST established by Jean-Paul Doignon and Jean-Claude Falmagne [2, 3], which is a well-elaborated set-theoretic framework for addressing the relations among problems (e.g., test items). It provides a basis for structuring a domain of knowledge and for representing the knowledge based on prerequisite relations (see Figure 2 for an example). While KST focuses only on performance (the behavior; for example, solving a test item), CbKST introduces a separation of observable performance and latent, unobservable competences, which determine the performance.

An empirically well-validated approach to CbKST was introduced by Klaus Korossy (1997, 1999); basically, the idea of the Competence-Performance Approach (CPA) is to

The same problem occurs for formative assessment and the ECAAD approach, specifically in a limited resource classroom scenario. An approach for a fairer and more precise appraisal comes with ideas of non-numerical test theories. With the rise of mathematical combinatorics and with the rise of powerful computer technologies the more demanding approaches to precisely describe the various abilities of a student along multiple dimensions got in the focus for researchers and educators. In parts, the basic idea found its way into class rooms, for example, when students are described in narrative text instead of a single grade in the report card.

The approach we want to utilize for NEXT-TELL – in form of smart software services – is based on the probabilistic, non-numerical method of Competence-based Knowledge Space Theory (CbKST).

Figure 1. Proponents of “Anthropometry”. From left to right: Sir Francis Galton, William Kelvin, and Carl Pearson.

Figure 2. Panel (a) illustrates a prerequisite relation among a set of atomic competencies; panel (b) illustrates the prerequisite relation among related test items, induced by the competence structure. The example stems from learning the relations about time, distance, and speed (cf. [4]).
assume a finite set of more or less atomic competences \( C = \{a, b, c, d, \ldots\} \) (in the sense of some well-defined, small scale descriptions of some sort of aptitude, ability, knowledge, or skill) and a prerequisite relation between those competences. A prerequisite relation states that competence \( a \) (e.g., to multiply two positive integers) is a prerequisite to acquire another competence \( b \) (e.g., to divide two positive integers). If a person possesses competence \( b \), we can assume that the person also possesses competence \( a \). To account for the fact that more than one set of competences can be a prerequisite for another competence (e.g., competence \( a \) or competence \( b \) are a prerequisite for acquiring competence \( c \) ), prerequisite functions have been introduced, relying on and/or-type relations. A person’s competence state is described by a subset of competences of \( C \), for example \( \{a, b\} \). Due to the prerequisite relations between the competences, not all subsets of competences are possible competence states. To give an example, imagine four competences from the domain of basic algebra, the abilities to add, subtract, multiply, and divide numbers. Given four competences, the set of all possible knowledge states is 24. If we assume that the competences to add, subtract, and multiply numbers are prerequisites for the competence to divide numbers, not all of the 16 competence states are plausible. For example, it is highly unlikely that a child has the competence to divide numbers but not to add numbers. The collection of possible competence states corresponding to a prerequisite relation is called competence structure. Such competence structure also singles out different learning paths for moving from the naïve state \( \{\} \) (having no competences of a domain) to the state of possessing all of a domain’s competences \( C \).

So far, the structural model focuses on latent, unobservable competencies. By utilizing interpretation and representation functions the latent competences are mapped to a set of tasks (or test items) \( Q = \{p, q, r, s, \ldots\} \) relevant for a given domain. Through the aforementioned example, \( Q \) might include a number of addition, subtraction, multiplication, and division tasks. The interpretation function assigns a set of competences required to solve a task to each of the problems in \( Q \). Vice versa, by utilizing a representation function, a set of problems is assigned to each competence state, which can be mastered in this state. This assignment induces a performance structure, which is the collection of all possible performance states. Due to these functions, latent competences and observable performance can be separated and no one-to-one mapping is required. Moreover, learning or development is not seen as a linear course; equal for all children, rather, development follows one of a set of individual learning or developmental paths. More in-depth introductions to ChKST can be found, for example, in Doignon & Falmagne (1999).

In the context of NEXT-TELL, this approach to the modeling of the knowledge domain in the first instance and of meaningful learning sequencing in a second instance provides some key advantages:

- One the one hand, we have a formal and computable method to analyze and define domains (i.e., the subject matter) on the level of (more or less) atomic entities of knowledge or ability – what term competency here. This formal nature in combination with a set of available tools and procedures enables to teacher to develop an individual structure of the subject matter, an individual curriculum, that involves the information and task from multiple sources (e.g., textbooks, books, websites, multimedia sources, peer activities, etc.). The flexible nature of the approach allow an easy mapping to other instances of NEXT-TELL, for example, the KSA modeling, the ECAAD methodology, software tools, and so forth. The prerequisite relations / functions induce competence structures which may serve as a means of planning, evaluating, and visualizing learning on the basis of admissible learning paths (cf. Figure 3). The knots of this Hasse diagram indicate meaningful competence states of a student while the edges indicate admissible transitions from one competence state to another by acquiring another competency.

- In terms of learning activity planning, a teacher may identify the starting point of an educational episode (e.g., a semester) by selecting the corresponding competence states in the figure and likewise she may identify the learning goal in a very precise manner.

- In terms of assessment planning, a teacher may select or construct items the perfectly suit the learning progress and competence states of the students.

- The Hasse diagram and associated graphs may be used to visualize the knowledge of students in a non-numerical and precise manner. This opens (a) a link to the use of open learner models (OLM) and it may serve (b) the ideas of negotiating and communicating appraisal outcomes (towards students and parents).

- The notions of inner and outer fringes enable very clear statements of what a student knows/can do at this very moment and it defines very clear indications of what a student can/should learn or practice next.

The ChKST approach originates from the field of intelligent and adaptive tutorial systems, that is, autonomous computer algorithms that attempt to interpret learner needs, learning progress, and problems within learning episodes and tailor educational measures accordingly. The aim is, essentially, to provide individual students with personalized and psychologically optimal learning experiences and environments. The challenge within NEXT-TELL is, to translate this “computer-centered” approach to the 21st century classrooms and the use of human teachers. In addition, the approach has to be mapped and linked to the other important aspects and facets of NEXT-TELL such as OLM, ECAAD, or TISL.
2.2 Practical Issues

The outlined approach includes three core components: the domain/learning/learner modeling and the obtained planning of sequences of learning activities and assessment activities. The approach facilitates an evidence-centered approach to evaluating students’ performance by allowing for a mapping of various activities (the evidences) to concrete competencies (the level of proficiency). Initially a conceptual analysis of the domain/subject matter is performed, subsequently the domain is modeled according the CbKST methodology (as briefly introduced above), followed by mapping the latent competencies to concrete learning and assessment activities. In the first instance, learning and assessment activities must be identified (by a teacher or learning designer) these activities are associated with the latent competencies in terms of required competencies (to be able to understand/perform a learning activity or to be able to master an assessment problem) and taught competencies (those that are taught or practiced with an activity). On the basis of the initially developed competence structure an activity structure is induced automatically. Subsequently, these modeling work leads to a well-defined learning sequence and it allows an assessment from multiple sources. The ECAAD planner serves as a means of the concrete analysis and planning work while the results may feed into an (open) learner model.

An important issue is the concrete application of such ideas in the classrooms. Establishing formative assessment scenarios including the necessary demands to learning analytics and the efforts required for visualization and communicating the information is highly resource demanding. Smart software is required that supports teachers. In the context of the NEXT-TELL project we developed a tool named ProNIFA which is supposed to promote and facilitate formative assessments.

3 Probabilistic, Unobtrusive, Formative-Assessment

ProNIFA stands for probabilistic non-invasive formative assessment and is developed in the context of the Next-Tell project. The tool, in essence establishes a handy user interface for CbKST-related services and functionalities. The services are running on a service and cover, broadly speaking, CbKST-related computation tasks, such as updating of probability distributions over competence structures. In addition to that, ProNIFA provides several authoring, analysis, and visualization features. The tool is a Windows application that utilizes various interfaces and links to online-based contents (Figure 4).

A distinct feature in the context of formative assessment is the multi-source approach. ProNIFA allows connecting the analysis features to a broad range of sources of evidence. This refers to direct interfaces (for example to Google Docs) and it refers to connecting, automatically or manually, to certain log files. The only requisite is the availability of log files on the local computer, through HTTP access or via FTP. Using this level of connectivity, multiple sources can be merged and can contribute to a holistic analysis of learners’ achievements and activity levels. As an example, ProNIFA enables a teacher to use the results of a Moodle test, exercises done in Google Spreadsheets, and the commitment displayed in a virtual meeting in a chat, to conduct a semi-automated appraisal of students.

The interpretation of the sources of evidence occurs depending on a-priori specified and defined conditions, heuristics, and rules which associate sets of available and lacking competencies to achievements exhibited in the sources of evidence. Very basically, the idea is to define certain conditions or states in a given environment (no matter if a Moodle test or a status of a problem solving process in a learning game). Examples for such conditions may be the direction, pace, and altitude a learner is flying with a spaceship in an adventure game or a combination of correctly and incorrectly ticked multiple choice tasks in a regular online
school test. The specification of such state can occur in multiple forms, ranging from simply listing test items and the correctness of the items to complex heuristics such as the degree to which an activity reduced the ‘distance’ to the solution in a problem solving process (technically this can be achieved by pseudo code scripting). The next step of this kind of planning/authoring is to assign a set of competencies that can be assumed being available and also lacking when a certain state occurs. This assumption can be weighted with the strength of the probability updates. In essence, this approach equals the conceptual framework of micro adaptivity as, for example, described by [5].

3.1 Innovative visualizations

Summarized, ProNIFA collects information about learning-related activities from various sources and analyses and interprets these information in terms of learning progress and abilities. This can be done on the level of individual learners as well as on the level of groups and entire classes. The analyses are running in the background and, in essence, deliver probabilities sets and probability distributions. To support formative assessment, the results need to be worked up and visualized in an easy to understand, intuitive way. Presently ProNIFA provides several techniques to visualize assessment results. On the one hand there is a set of charting methods such as bar charts or line graphs that can visualize abilities in a rather conventional form. For example, the distribution of grades can be illustrated as a pie chart (cf. Figure 4). In addition, ProNIFA offers more innovative and distinct ways of visualization information. Three of them are presented and introduced in the following section.

3.1.1 Hasse diagrams

A Hasse diagramm is a strict mathematical representation of a so-called semi-order. Invented in the 60s of the last century by Helmut Hasse, entities (the knots) are connected by relationships (indicated by edges), establishing a directed graph. The properties of a semi-order are (i) reflexivity, (ii) anti symmetry, and (iii) transitivity. In principle, the direction of the graph is given by arrows of the edges; per convention however, the representation is simplified by avoiding the arrow heads, whereby the direction reads from bottom to top. In addition, the arrows from one element to itself (reflexivity property) as well as all arrows indicating transitivity are not shown. The following image illustrates a Hasse diagram (Figure 5).

Hasse diagrams enable a complete view to, partially, huge structures. Insofar, they are ideal for capturing the large competence spaces occurring in the context of CbKST-based assessment and recommendations. Very briefly, a Hasse diagram shows all possible (admissible) competence or knowledge states. By the logic of CbKST, each learner is, with a certain likelihood, in one of the competence states. This allows coding the state likelihoods for example by colors and thereby visualizing areas and set of states with high (or vice versa low) probabilities. The simplest approach would by highlighting the competence state for a specific learner with the highest probability. The same coding principle can be used for multiple learners. This allows for identifying various subgroups in a class, outliers, the best learners, and so on. A second aspect comes from the edges of the graph. Since the diagram reads from bottom to top, the edges indicate very clearly the “learning path” of a learner. Depending on the domain, we can monitor and represent each learning step from a first initial competence state to the current state. In the context of formative assessment, such information elucidates efforts of the learners, learning strategies, perhaps used learning materials, but also the efficacy of the teachers. Finally, a Hasse diagram offers the visualization of two very distinct concepts, the inner and out fringes. The inner fringe indicates what a learner can do / know at the moment. This is a clear hypothesis of which test/assessment items this learner can master with a certain probability. Such information may be used to generate effective and individualized tests. The concept of the out fringe indicates what competency should or can be reasonably taught to a specific learner as a next step. This provides a teacher with clear recommendation about future teaching on an individualized basis.

3.1.2 Pixel Clouds

Pixel clouds are a similar concept of representing ability on an individual or group level. In principle, the pixel clouds depicts each competence state (or skill/competency) as a single pixel. Each of the competence states is assigned a probability value which is color coded. The brighter a pixel is the higher is the corresponding probability, vice versa, the darker a pixel is the lower is the corresponding probability. The difficulty (or in other terms the structural location) of a competence or competence state in given by the position in the Euclidean space, ordered from left to right. This type of visualization has the great advantage that huge competency spaces can be grasped with a single sight. Despite maybe huge spaces, important information for teachers can be displayed on a single screen without the need for zooming. As show in the Figure 6, by this means also temporal information can be illustrated easily and quickly.
3.1.3 Problem Spaces
A problem space is a formal and complete description of all possible solutions steps for a specific problem, represented from the starting state to the final (desired) state. All steps are represented on the basis of their admissibility according to the corresponding set of rules. The figure below illustrates the famous problem (game) the “Tower of Hanoi”. In the context of formative assessment, this type of visualization can be used to illustrate the progress of students in problem solving situations. This might be on a small scale (within a short term problem, e.g., a test item) or on a large scale (e.g., in terms of a medium to long term project). It shall be highlighted that the term “problem” refers to a broad notion of task, a problem might well be a mathematical test item which is based on a set of calculation rules. The type of illustration (cf. Figure 7) is highly intuitive and, maybe more importantly, allows not only a snapshot of the present state but also indicates chosen solution paths.

Figure 6. Pixel cloud illustrating an assessment process. With each piece of evidence (in an ECAAD sense), the picture of a learner’s competencies is getting clearer and more stable.

Figure 7. The image shows the problem space for the famous Tower of Hanoi problem/game.

4 Conclusions
It is evident that the future of teaching and learning necessarily needs to pursue innovative and novel paths. In particular in the STEM area, a clear and important trend is formative assessment, ideally on the basis of non-numerical information. This, unfortunately, is not a trivial attempt. It requires a highly amount of resources from a teacher to support the individuals in such a “formative” supportive and progressive way. The problem is that such resources are sparse and, facing today’s classroom situations, formative assessment and appraisal of learner’s cannot be mastered by teachers with a comprehensive and above all smart technology. In the context of the NEXT-TELL project we try to realize the necessary innovations and technical developments to empower teachers and to support teachers in their daily routines in order to realize better teaching and assessing. ProNIFA and CbKST as the conceptual backbone, are a promising approach. Presently, the ideas are investigated in European schools. First feedback shows that teachers highly appreciate this kind of assessment support. Of course, ProNIFA, as presented here, is just a small snapshot of the developments in the entire project. The fundamental aim is to cover the entire teaching cycle from planning, to conducting technology-supported teaching with various tools, to gathering evidence, analyzing and visualizing the student achievements, and finally to communicating and negotiating results and decisions.

5 Acknowledgements
The research and development introduced in this work is funded by the European Commission under the seventh framework programme in the ICT research priority, contract number 258114 (NEXT-TELL; http://www.next-tell.eu).

6 References


