
# Supporting Formative Assessment and Appraisal by Smart, Competence-Based, Probabilistic Systems

**Michael D. Kickmeier-Rust, Dietrich Albert**
Knowledge Technologies Institute, Graz University of Technology (Austria)
michael.kickmeier-rust@tugraz.at

## Abstract

*Individuals have individual needs; this is in particular true for learners. Although there is a movement towards a more holistic and competence-oriented approach to appraisal of student achievement and subsequent individual tutoring, most common practice is still evaluating learning performance by a single numerical value – the school grades. And most often, these summaries are based on superficial, one-dimensional test items. This "over-simple" approach cannot express what learners really know or can do. Thus, over the past years and increasing shift from summative to formative assessment practices is noticeable. European school curricula are heading towards competence-centered approaches and feedback policies beyond mere grades or achievement degrees. What it takes, however, to realize such ideas and visions is sound theories and smart technologies unburden teachers and support them focusing on individual performance and needs in a formative sense. In this paper we introduce an approach based on so-called Competence-based Knowledge Space Theory. In the context of the European Next-Tell project, we are developing smart software solutions to support teachers in collecting information about the learners' achievements and progress, to aggregate them in, and to profit from the bunch of information in form of smart multiple analyses and visualizations. Due to the CbKST-based logic, the system allows clear statements about individual learning progress and enables suggestions about next educational steps. Other scenarios focus on working collaboratively with e-Portfolios or cloud tools. Presently we are conducting field studies with the prototype software in schools in Austria, Denmark, Norway, and the UK.*

## 1. Introduction

Abilities, strength, weaknesses of learners, their knowledge and misconceptions, their needs and goals are extremely broad and rich; in terms of evaluating learning performance, most often all achievements are summarized by a single numerical value – the school grades. And most often, these summaries are based on superficial, one-dimensional test items. This approach, however, cannot express what learners really know or are able to; characterizing proficiency by a single variable at best suffice basic fail/pass decisions, as argued by Mislevy and colleagues [1]. 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.
The origin of this popular test theoretical approach lies in 19th century physics and the occurrence of disciplines like "anthropometry", the "art of measuring the physical and mental faculties of human beings". Prominent proponents were Francis Galton, William Kelvin, or Carl Pearson. 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" (cf. [2]).
When aiming at an evidence-centred approach to evaluating achievements and proficiency in school settings, when focussing on a formative approach to appraisal with the idea of supporting learners in a meaningful way and on an individual level, a deeper and more precise understanding is necessary. Such attempt, however, is not trivial. It is complex, demanding, and costly. On the one hand it takes a profound theoretical approach to evaluation which includes all the various dimensions and on the other hand, evaluators (teacher, in the first instance) are required to develop a fair and comprehensive image of learners, their origin, learning performance and present ability/proficiency/knowledge for each individual learner. Non-numerical test theories provide ideas for such broadened evaluation. Basically 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 learner along multiple dimensions appear more feasible in practice

A more formative and competence-centred assessment and appraisal must be based on two fundamental concepts; that of the so-called substantive features, which concern the characteristics of the learning domain and the learning process, and the evidentiary-reasoning aspect, which concerns the information we can draw from the learners' behaviours. It takes a formal framework that links both, the substantive and the evidentiary-reasoning aspects of a sound, reliable, and, in a way, formative assessment. Most likely, such frameworks are based on probability values, for example, Item Response Theory [3], Latent Class Models [4], or Bayesian inference networks [5].

## 2. Competence-based Knowledge Space Theory for formative assessment

*Knowledge Space Theory* (KST), founded by [6, 7] and extensions such as *Competence-based Knowledge Space Theory* (CbKST) that melds both, the substantive and the evidentiary-reasoning, coming from the genre of autonomous intelligent and adaptive tutoring systems. The idea was to broaden the ideas of the linear Item Response Theory (IRT) scaling, where a number of items are arranged on a single, linear dimension of "difficulty". In essence, KST provided a basis for structuring a domain of knowledge and for representing the knowledge based on prerequisite relations. More recent advancements of the theory accounted for a probabilistic view of test results and they introduced a separation of observable performance and the actually underlying abilities and knowledge of a person. Such developments lead to a variety of theoretical, competence-based approaches (see [8] for an overview). An empirically well-validated approach to CbKST was introduced by Korossy [9]; basically, the idea was to assume a finite set of more or less atomic competencies (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.

In a first step, CbKST attempts to develop a model of the learning domain, e.g. algebra. Examples for such competencies might be the knowledge what an integer is or the ability to add two positive integers and so on. The level of granularity to which a domain is broken down depends on the envisaged application and might range from a very course-grained level on the basis of lessons (for example to plan a school term) to a very fine-grained level of atomic entities of knowledge/ability (for example as the basis of an intelligent problem solving support application). In a second step, CbKST looks into a natural course of learning and development and into logical/natural prerequisites between competencies. Usually, learning and the development of new abilities as well as the stabilization of skills occurs along developmental trajectories. On the basis of a set of competencies and a set of prerequisite relationships between them, we can formally derive a collection of so-called competence states (Figure 1). Due to the prerequisite relations between the competencies, not all subsets of competencies (the power set) are possible competence states.

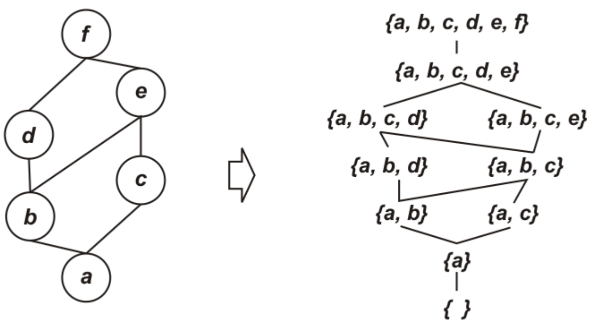


Fig. 1. On the left a prerequisite relation for 6 competencies is shown (reading *a* is the prerequisite or predecessor of *b* and *c*, etc.); on the right the derived competence structure is displayed.

So far, the structural model focuses on latent, unobservable competencies; loosely speaking the model makes hypotheses about the *brain's black box*. By utilizing interpretation and representation functions the latent competencies are mapped to evidence or indicators relevant for a given domain. Such indicators might be test items but might refer to all sorts of performance or behaviour (e.g., the concrete steps when working with a spread sheet application). Due to these functions, latent competencies and observable performance can be linked in a broad form where no one-to-one

correspondence is required. This means that an entire series of indicators can be linked to underlying competencies / competence states. CbKST, of course, accounts for the fact that indictors such as test items cannot be perfect evidence for the latent knowledge or ability. There is always the possibility that a person makes a lucky guess or exhibits a correct behaviour/activity just by chance. In turn, a person might fail in a test item although the necessary knowledge/ability is actually available, for example, by being inattentive or careless. Thus, CbKST considers indicators on a probability-based level, this means that, as an example, mastering a test item suggest having the underlying competency/competencies, with a certain probability. Conceptually, this view constitutes a probability distribution over the competence structure. A further significant advantage of such approach is that learning is not only considered a one dimensional course on a linear trajectory, equal for all learners. Learning and development rather occur along one of an entire range of possible learning paths.
Recent advancements of CbKST primarily concern the integration of theories of human problem solving (given that most indictors can be interpreted as solving some sort of problem). This work was essentially driven in the genre of smart, educationally adaptive computer games for learning – loosely speaking for developing an educational AI support the players of the game [10, 11].

## 3. Smart solutions for the classroom

While the origin CbKST lies in the field of intelligent tutorial systems, in recent work, we aim at bringing the advantages in the classrooms. With the European *Next-Tell* (www.next-tell.eu) project, we are developing smart software solutions to support teachers in collecting information about the learners' achievements and progress, to aggregate them in *Open Learner Models*, and to profit from the bunch of information in form of multiple analyses and visualizations. One scenario developed in the context of Next-Tell concerns teaching English in virtual environments such as *OpenSim*. Without smart software support, a teacher is not able to monitor and interpret all the activities of an entire class in an open virtual environment. It is also hard to evaluate log files manually. The Next-Tell system enables an automatic log file analyses and, subsequently, the rule-based linking of activities in the virtual world with certain competencies and learning progress. Finally, the tool provides the teacher with in-depth analyses on the basis of individuals as well as the entire class and it offers multiple visualizations of results. Due to the CbKST-based logic, the system allows clear statements about individual learning progress and enables suggestions about next educational steps. Other scenarios focus on working collaboratively with e-Portfolios or cloud tools.

### 3.1. ProNIFA

ProNIFA is a tool to support teachers in the assessment process. The name stands for probabilistic non-invasive formative assessment and, in essence, establishes a handy user interface for related data aggregation and analysis services and functions. ProNIFA retrieves performance data (e.g., the results of a test or the activities in a virtual environment) and updates the probabilities of the competencies and competence states in a domain. When a task is mastered, all associated competencies are increased in their probability, vice versa, failing in a task decreases the probabilities of the associated competencies. A distinct feature in the context of formative assessment is the multi-source approach. ProNIFA allows for 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. 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. The idea is to define certain conditions or states in a given environment (regardless if it is 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 space ship 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 to be available and also lacking when a certain state occurs. This assumption can be weighted with the strength of the probability updates. Figure 2 is a screenshot of ProNIFA analysed data from a *Second Life* activity. The resulting domain model, build around atomic competencies, as well as the related probability distribution is passed to an open learner model platform as a next step in supporting the teacher's evaluation and appraisal efforts.

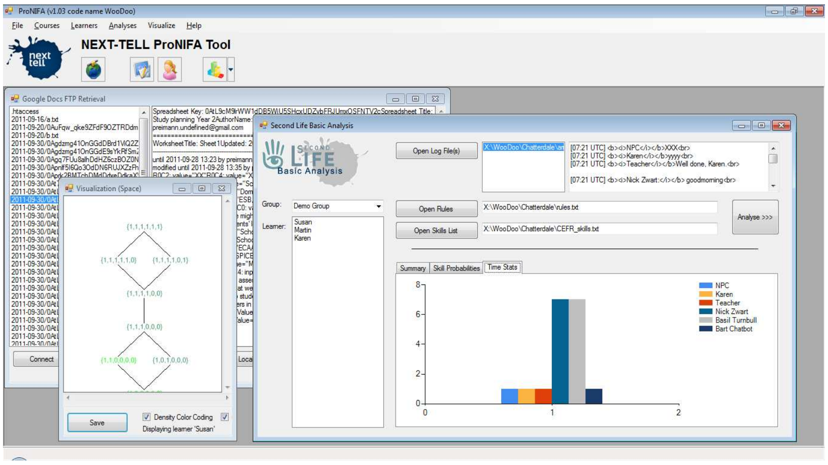


Fig. 2. Screenshot of ProNIFA.

## 4. Summary

This paper has introduced a way to help teachers take the wide range of data about students, and transform it into a form that can be used for formative ideas of assessment and support of learners on an individual basis. It is important to build modern educational approaches for the 21st century classroom on a sound psycho-pedagogical foundation in order to achieve a broader meaning then mere data mining over activities. CbKST may serve as such foundation. The software we are realized in the context of Next-Tell has the practical use by European teachers in mind. Presently we are conduct evaluation studies in schools in Austria, Denmark, Norway, and the UK. The preliminary results give hoe that the introduced approach really advances today's and tomorrow's educational practice.

## Acknowledgements

This project is supported by the European Community (EC) under the Information Society Technology priority of the 7th Framework Programme for R&D under contract no 258114 NEXT-TELL. This document does not represent the opinion of the EC and the EC is not responsible for any use that might be made of its content.

## References

[1] Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives, 1, 3-67.
[2] Falmagne, J.-C., Cosyn, E., Doignon, J.-P., & Thiery, N. (2003). The assessment of knowledge, in theory and in practice. In R. Missaoui & J. Schmid (Eds.), Lecture Notes in Computer Science, (pp. 61–79). Berlin: Springer.
[3] Van der Linden, W.J., & Hambleton, R.K. (Eds.) (1997). Handbook of modern item response theory. New York: Springer.
[4] Collins, L.M., & Lanza, S.T. (2010). Latent class and latent transition analysis for the social, behavioral, and health sciences. New York: Wiley.
[5] Jensen, F.V. (1996). An Introduction to Bayesian Networks. Berlin: Springer.


[6] Doignon, J.-P., & Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. International Journal of Man-Machine Studies, 23, 175-196.
[7] Doignon, J.-P., & Falmagne, J-C. (1999). Knowledge spaces. Berlin: Springer.
[8] Albert, D., & Lukas, J. (Eds.). (1999). Knowledge spaces: Theories, empirical research, and applications. Mahwah, NJ: Lawrence Erlbaum Associates.
[9] Korossy, K. (1999). Modelling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), Knowledge spaces: Theories, empirical research, and applications (pp. 103–132). Mahwah, NJ: Lawrence Erlbaum Associates.
[10[ Kickmeier-Rust, M.D., Mattheiss, E., Steiner, C.M., & Albert, D. (2011). A Psycho-pedagogical framework for multi-adaptive educational games. International Journal of Game-Based Learning, 1 (1). 45-58.
[11] Kickmeier-Rust, M. D., & Albert, D. (2010). Micro adaptivity: Protecting immersion in didactically adaptive digital educational games. Journal of Computer Assisted Learning, 26, 95-105.