Towards Learner Models based on Learning Progressions in DeepTutor

Vasile Rus, William Baggett, Elizabeth Gire, Don Franceschetti, Mark Conley, Arthur Graesser
The University of Memphis
Contact: vrus@memphis.edu

INTRODUCTION

Learner modeling is a central component in any tutoring system that also claims to be intelligent. A learner model is a set of learner characteristics that impact learning and therefore is important for adapting instruction to each individual learner and thereby maximize learning. Indeed, research has shown that learners' diverse backgrounds require tailoring instruction to each individual learner or, more realistically in classroom settings, to homogeneous groups of learners (Corcoran, Mosher, & Rogat, 2009). It should be noted that a major advantage of computer-based instruction, including Intelligent Tutoring Systems (ITS), over classroom instruction is the potential of offering tailored instruction to each and every learner. These systems presumably can be scaled up in the sense that once developed for one user they could be easily replicated or modified to handle many users. Learner modeling is a particular case of user modeling, which constitutes a central component in any user adaptive system. User adaptive systems tailor their behavior to the individual user's characteristics in order to optimize the value of their function with respect to the user (Conati, Gertner, & VanLehn, 2002; VanLehn, 2006). We focus here on user models in ITSs, i.e. on learner models.

The learner model plays a role in both the outer and inner loop of ITSs (VanLehn, 2006). The outer loop handles macro-adaptivity in the sense that it selects tasks and other instructional materials to present to the learner. The outer loop also selects a mode for the task, e.g., demonstrating a step-by-step solution to the task or just providing hints. The inner loop manages the tutor-tutee interaction within a task by monitoring the learners' actions while she is working on an instructional task. The inner loop handles micro-adaptivity.

We present in this chapter a novel approach to learner modeling that enables improved macro- and micro-adaptivity. We emphasize here the impact of the new approach on macro-adaptivity. The novel approach relies on a research framework, called Learning Progressions (LPs), developed recently by the science education research community as a way to increase adaptivity in traditional instruction. Indeed, "assessment for learning" (Black & William, 1998) has been a focus in this community for more than a decade (National Research Council, 2001, 2005, 2007). This effort led to the emergence of the framework of LPs, defined as "descriptions of the successively more sophisticated ways of thinking about an idea that follow one another as students learn" (NRC, 2005, 2007). Corcoran, Mosher, and Rogat (2009) state "progressions can play a central role in supporting the needed shift toward adaptive instruction as the norm of practice in American schools." Importantly, LPs provide a promising means to organize and align content, instruction, and assessment strategies to give students the opportunity to develop deep and integrated understanding of science ideas. The time is ripe for the ITS community to integrate in computer tutors such advances in assessment and instruction proposed by the assessment and science education research communities in order to increase ITS's adaptivity and in turn their effectiveness at inducing learning gains in learners.

Our dialogue-based ITS DeepTutor (Rus et al., to appear) incorporates the framework of Learning Progressions (National Research Council, 2007) as a way to improve assessment and better tailor instruction to each individual learner. DeepTutor is under development as of this writing which explains the preliminary flavor of the ideas presented in this chapter. Because DeepTutor is a dialogue-based ITS, the main form of interaction in DeepTutor is tutorial dialogue that mimics interaction between human tutors and learners. Deep natural language processing technologies are also needed to accurately assess students' level of understanding while interacting with DeepTutor. In fact, the quality of the algorithms for dialogue and language processing has a direct impact on other core ITS tasks such as feedback generation. An authoring tool that allows us to explore and design algorithms for deep natural language processing has been developed (more information about the SEMantic simILARity, or SEMILAR, toolkit is available at www.semanticsimilarity.org).

Based on (1) LPs that promote deep and integrated understanding of science and (2) deep natural language processing algorithms (as well as several other novel aspects of DeepTutor such as advanced tutorial strategies which we do not discuss here), DeepTutor is expected to provide better assessment and in turn better adapt instruction to each individual learner. This will lead to learning gains beyond the interaction plateau, i.e., the hypothesis that as interactivity of tutors increases, their effectiveness plateaus (VanLehn, 2011). VanLehn's interactivity plateau finding calls for ITS researchers to propose qualitative advances that can increase the learning effectiveness beyond the interactivity plateau. The proposed advances in DeepTutor are meant to address this challenge by proposed qualitative shifts for core ITS components such as the domain and student modeling components through the use of LPs.

The rest of this chapter is organized as follows. After a brief general discussion of learner modeling in ITSs, we describe the framework of LPs. Then, we describe general ideas about the development and validation of LPs and illustrate how this process is implemented in DeepTutor, the first ITS based on LPs. The role of LPs for macro-adaptation in our DeepTutor project is then highlighted. We conclude with a discussion about how the Generalized Intelligent Framework for Tutoring can accommodate LP-driven ITSs.

LEARNER MODELING IN ITSs

ITSs researchers have investigated and integrated learner models in their systems to increase the effectiveness of these systems (Conati, Gertner, & VanLehn, 2002; Corcoran, Mosher, & Rogat, 2009; Lintean, Rus, & Azevedo, 2012; Sottilare et al., 2012; VanLehn, 2006). Although other learner characteristics such as emotions have been considered recently, researchers and developers of ITSs have focused primarily on learners' performance with respect to the target domain, i.e. knowledge assessment, to guide adaptivity.

Different ITSs employ different approaches to the problem of learning modeling, including knowledge tracing in the Cognitive Tutor (Anderson et al., 1995), constraint-based modeling (Mitrovic, Martin, & Suraweera, 2007), and the Expectation-Misconception approach used in AutoTutor (Graesser, Rus, D'Mello, & Jackson, 2008). In all of these models, the emphasis is on supporting micro-adaptivity. Attempts to handle macro-adaptivity have been modest. The SMART student model (Shute, 1995) uses regression equations to estimate students' mastery of each curriculum element. Conati, Gertner, & VanLehn (2002) describe the learner model used in their ANDES system that tutors students on Newtonian Physics. The model is based on Bayesian networks in which explicit domain-general and context-specific rules are encoded. Conati and colleagues describe the use of the model for micro-adaptivity during example studying and problem-solving, but they do not report any use for macro-adaptivity. The mathematics tutor ALEKS (Doignon & Falmagne, 1999) relies on the knowledge space theory (KST) for domain modeling. KST relies on the precedence relation that is evident for some domains like mathematics where a student must learn a concept A before learning another that relies on A. A student's knowledge state is the set of items she has mastered. Not all subsets of items are feasible knowledge states due to the precedence relations among items. Based on assessment, a student's knowledge state is inferred, which guides instruction. Learning occurs on the outer fringe of the knowledge state that is the immediate successor state in the knowledge structure of the domain. The knowledge structure is the collection of all possible states. ALEKS offers macro-adaptivity but the organization of the domain is based on a logical organization provided by experts according to intrinsic dependencies among concepts in a domain. There is no emphasis on the developmental milestones students pass while learning a topic. KST may be suitable for domains where precedence relations are well defined such as mathematics but for domains where there is no such clear structure KST is less useful. An interesting research question is about how KST for well defined domains can be transformed into or combined with the LP framework.

There are three general challenges with deriving knowledge assessment learner models for each individual student based on her actions during learning: (1) determining how much activity from an individual learner the system must monitor and analyze to reliably estimate the parameters of the learner model, (2) keeping the parameters up-to-date while the student's level of understanding of a domain evolves, and (3) representing and handling misconceptions. The knowledge assessment learner models encode primarily what the students are supposed to learn, i.e. true knowledge, while not embedding explicitly alternative conceptions. These alternative conceptions, or misconceptions, are handled separately. Novel solutions based on LPs can better address these issues as shown next in the context of our DeepTutor ITS.

We use the framework of Learning Progressions as a paradigm shift in learner modeling for ITSs. LPs model and organize domain knowledge based on what is known about how learners actually progress through that content as opposed to a logical decomposition of the content by a domain expert. LPs explicitly consider the alternative conceptions, mapping how they developmentally emerge. This learner-centered organization of content enables improved macro-adaptation and micro-adaptation. We focus here on the role of LPs in supporting better macro-adaptation. Coupled with appropriate assessment instruments and instructional strategies, LPs offer the possibility of qualitative shifts in learner modeling and learner experiences with ITSs.

THE FRAMEWORK OF LEARNING PROGRESSIONS

The NRC (2001) report called for better descriptions of how students learn based on models of cognition and learning. Based on such descriptions of how students learn, "assessments can be designed to identify current student thinking, likely antecedent understandings, and next steps to move the student toward more sophisticated understandings" (NRC, 2001, p. 182). This was basically a call for developing Learning Progressions (Corcoran, Mosher, & Rogat, 2009). The term Learning Progressions was first used in a subsequent NRC report (2005). According to Corcoran, Mosher, & Rogat (2009), learning progressions in science are “empirically grounded and testable hypotheses about how students’ understanding of, and ability to use, core scientific concepts and explanations and related scientific practices grow and become more sophisticated over time, with appropriate instruction “ (p.8).

Corcoran, Mosher, and Rogat (2009) also mention as major conceptual contributions (although they did not use the term of LPs): (1) the work of Lehrer and Shauble (2000) on model-based reasoning in science; and (2) research by Valverde and Schmidt’s (1997) on the relationship between content and structure of curricula and student achievement. It is also worth mentioning the work of Mark Wilson and colleagues (e.g., Roberts,Wilson, & Draney, 1997) on “construct maps” and work on progress maps by Masters and Fosters (1996). Wilson and colleagues and Masters and Fosters pushed for a concentrated effort on improving assessment to guide instruction. Work on learning trajectories in mathematics (Driver, Leach, Scott, & Wood-Robinson, 1994) as well as work on cognitive development maps (Baxter & Junker, 2001) further contributed to the development of LPs.

LPs can be viewed as incrementally more sophisticated ways to think about an idea that emerge naturally while students move toward expert-level understanding of the idea (Duschl et al., 2007). LPs define qualitatively different levels of understanding of big ideas. The levels can be sequentially related or the relation could be more complex. For instance, topic A may develop the ideas from a less sophisticated topic B but also connect to other topics. LPs model how students develop deep and integrated understanding of complex science topics. The emphasis on increased sophistication (i.e., depth) and integrated understanding of ideas sets LPs apart from classroom instruction and assessment, which are typically based on curriculum materials that follow local, state, and national standards. These standards tend to support compartmentalized understanding and shallow coverage of a broad range of topics instead of a deep, integrated understanding of few key ideas (Schmidt, Wang, & McKnight, 2005; Stevens, Delgado, & Krajcik, 2009). The bottom line is that LPs adopt a *learner-centric* view of a topic, modeling students’ successful paths towards mastery as opposed to paths prescribed by domain experts following a logical decomposition of big ideas. The logical decomposition provided by experts could be useful as a starting point but then needs be re-organized based on evidence of how students actually develop mastery of the big ideas. These actual paths must be documented and guide instruction. While some paths towards mastery defined by domain experts may be successfully followed by some students, some other students may follow other paths, i.e. might be more responsive to different topic sequences as well as instructional tasks that help their conceptual change.

LPs are organized in levels of understandings which reflect major milestones in learners’ journey towards mastery. The lower level, called the Lower Anchor, represents naïve thinking that usually novices hold. The top level, called the Upper Anchor, represents the mastery/expert level of understanding. If targeting a particular population, e.g. high-school students, the Lower Anchor would coincide with the Upper Anchor of grade 8 students. Such an assumption is not always practical, as we learned during the development of our DeepTutor system, because not all students entering high-school are at the Upper Anchor of grade 8. Furthermore, if someone develops an instructional intervention such as DeepTutor to be used by learners of all ages then the Lower Anchor should be specified accordingly, i.e. to reflect the lowest possible level of understanding. The anchoring of the Upper Level in an LP raises the issue of whether standards should indicate what the best and brightest students can achieve in a particular grade or whether standards should consider what average students can do (Corcoran et al., 2009). It is beyond the scope of this chapter to discuss this issue. In our work, we use standards to specify the Upper Anchor for a particular grade level. Our true Upper Anchor specifies the true and strongest scientific conceptions.

Open Issues

While there are many commonalities and a general understanding of what LPs are, some aspects of LPs are open for debate and interpretation and vary from one developer to another. For instance, the span of an LP can vary from one instructional unit that is covered over several weeks during a semester in a particular grade tocovering multiple grades, e.g. 6-8 years of instruction. Alonzo and Steedle (2009) differentiate between broad LPs and small LPs that cover in more detail a particular idea or construct in the sense proposed by Wilson in his Structured Construct Model (SCM; Wilson, 2009). That is, broad LPs may cover the development of a set of big ideas with each big idea constituting a small LP. The broad LP has the advantage of providing the big picture: the big ideas and how they develop across grades as well as the inter-dependencies among these big ideas. The smaller LPs provide sufficient detail such that instructors can use them to track students' progress over instructional units and guide their instruction.

Another varying aspect of LPs is their granularity. Fewer levels describing in summary core ideas means a more general LP and also more reliable diagnostic. However, finer-grained LPs offer a more sensitive instrument to measure progress, have more explanatory power, and can be more useful in guiding instruction. A related contentious aspect of LPs is their relationship with curriculum and instruction. Some developers do consider instruction as an integral part of LPs (Songer, Kelcey, & Gotwals, 2009) while others develop LPs independent of instruction (Mohan, Chen, & Anderson, 2009). The approach adopted affects the LP validation process.

We conclude this section by noting that research on LPs is thriving with many LPs being developed (Alonzo & Steedle, 2009; Songer, Kelcey, & Gotwals, 2009; Jin & Anderson, 2012; Neumann, Viering, Boone, & Fischer, 2012; Johnson & Tymms, 2011), many LPs conferences and other events being organized, and LPs being adopted by states (see the LPs adopted by the state of Massachusetts).

THE DEVELOPMENT AND VALIDATION OF LPs IN DEEPTUTOR

We now describe the structure and process of developing and validating LPs. LP development is driven by assessment, i.e. what gets measured and what counts as evidence for learning, and should include the following:

1. Content and conceptions (student thinking) – what gets measured. That is, LPs must specify what students need to know and the various alternative conceptions they may have.
2. Learning performances which are the "operational definitions of what learners' understanding and skills would look like at each of these stages of progress" (Corcoran et al., 2009) – what counts as evidence of level of understanding.
3. The Upper Anchor describes the set of knowledge and skills students are expected to have at the end of the progression. These expectations can be found in state or national science standards and recent learning research on the subject matter.
4. Progress levels or steps of achievement (intermediate levels) include both correct and incorrect/weak conceptions that learners consistently use to reason about phenomena in a given domain. Some of the correct conceptions are held early on. Some weak conceptions (or misconceptions) can be extremely persistent across many progress levels.
5. The Lower Anchor describes student conceptions at entry level, i.e. when they start learning about something.
6. Assessments that measure student understanding of the key concepts or practices.
7. Purposeful curriculum and instruction that mediates targeted student outcomes (Duschl, 2011; Corcoran et al., 2009).

We describe next the process we followed in developing the Force and Motion LP in our DeepTutor project. We adopted a design-based research iterative process that allowed us to develop and validate the progressions (Mohan, Chen, & Anderson, 2009; Stevens et al., 2009). The process first conjectured a hypothetical learning progression (HLP) and then derived an empirical progression (EP) which was used to refine the HLP. After several iterations of refinements and empirical validations, the HLP becomes an empirically tested LP. We used the following three criteria to guide our validation process: conceptual coherence, compatibility with current research, and empirical support from real student data (Mohan et al., 2009). The development of the initial HLP presupposes defining the ideas and concepts to learn at an appropriate level of detail while considering recent cognitive and education research that could provide insights about potential challenges students face or typical prior knowledge students may be expected to have.

The result of our LP design effort in the DeepTutor project is a broad Newtonian Physics LP structured in seven strands, or smaller LPs in the sense of Alonzo & Steedle (2009). We have one small LP for each of seven themes or big ideas in Newtonian Physics: Kinematics, Force and Motion (linear motion), Mass and Motion, Free-Fall new Earth, Newton's Third Law, Vectors and Motion (motion in two dimensions), and Circular Motion. Each strand is organized in a number of levels. The number of levels in the broad LP varies from strand to strand, e.g. the Mass and Motion strand has 3 levels while the Free-Fall near Earth strand has 7 levels. Levels are not equivalent across strands, i.e. level 2 in the Mass and Motion strand is not equivalent to level 2 in the Free-Fall strand. The strands in turn are ordered in terms of their complexity and prerequisite requirements. For instance, understanding the basic concepts of position, velocity, and acceleration covered in Kinematics is needed before studying Newton's second law in the Force and Motion strand. That is, we have a two dimensional broad LP in which one dimension illustrates students' level of understanding and the other the complexity and interdependencies among the big themes (see the left side of Figure 1). The broad LP can also be regarded as a one-

strand LP with seven levels of understanding, each level corresponding to one of the big Physics themes/topics (see Figure 1, right side). This HLP has been validated based on data collected from high-school students. The HLP is now an Empirical Progression (EP).
Our LP works in close relationship with a set of assessment items that allows the instructor, in our case DeepTutor, to place the student somewhere in the LP. Based on where the student is placed in the LP, instructional tasks and materials that are appropriate for that level of understanding are assigned to her. Because the LP levels are so designed to encode the most successful paths to mastery followed by students, the LPs greatly facilitate the selection of tasks to be given to a particular student, i.e. LPs enable macro-adaptation of instruction. Tasks and materials associated with a level in the LP are so designed to help students improve their understanding and move up the LP hierarchy.

It should be noted that tasks based on an LP can be sequenced in various ways, leading students on different learning trajectories. For instance, drilling tasks which offer training on one big theme modeled by one LP strand will more likely help students move up the level of understanding within that strand while not making progress on other big ideas. For some strands, e.g. Newton's 3rd law, for which the correct answer to many problems is the same (the tasks are isomorphic at some degree) a repetitive drilling strategy is problematic as students may learn the jingle ("for every action there is an equal and opposite reaction") after seeing the solution to a few problems and just recite the jingle when prompted for solutions to subsequent tasks without actually developing a deep understanding of Newton's 3rd law. Smarter sequencing of problems must be adopted as isomorphic problems lead to copying and therefore shallow learning (VanLehn, 2011; Renkl, 2002).

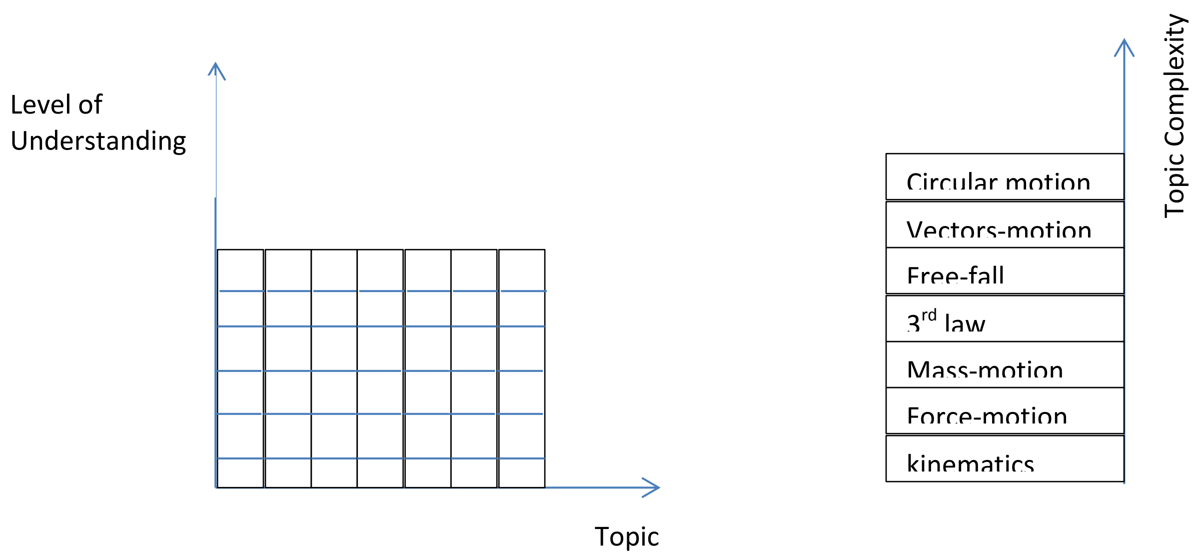

Figure 1. The Newtonian Physics LP: 2-D depiction (left) versus 1-D, broad LP depiction (right).

DeepTutor infers learners' levels of understanding of the target domain using both summative and formative or embedded assessment. Summative assessment consists of pre- and post-tests. The inferred levels of understanding based on the pre-test are continuously updated based on learners' performance while working on various tasks with the system, i.e. through formative assessment. Formative assessment is seamless as learners are assessed as they work on a problem without being aware of this type of assessment. Formative assessment directly and immediately impacts micro-adaptation while indirectly impacting macro-adaptation through the continuous update of the learner model. For instance, a learner who embraces many misconceptions or confuses concepts such as velocity and acceleration or does not understand the fine difference between velocity and speed will have her level of understanding updated accordingly even though she might have done better on the pre-test. We are moving towards a cloud-model of student assessment in which we allow students to simultaneously hold in their minds different models of reasoning or levels of understanding of a target domain (Rus et al., in press). Students will activate one model or level with a certain probability distribution. Aligning assessment and instruction based on LPs is a major feature in DeepTutor which is expected to increase the adaptivity of the system and in turn improve students' learning experience and learning gains.

DISCUSSION WITH RESPECT TO THE GENERALIZED INTELLIGENT FRAMEWORK FOR TUTORING

The Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg, & Holden, 2012) aims at providing a standard that unifies and streamlines ITS development efforts. Modularity, communication among modules, and separation of code from content are the driving principles of GIFT. Furthermore, GIFT supports a service-oriented architecture to facilitate distributed and mobile learning. The GIFT framework includes the following major modules: sensors module, user module, pedagogy module, and domain module. All these major modules are domain independent except the latter. The user module includes the learner model among models for other users such as trainer, expert, and designer. The domain module performs assessment functions such that the only domain-dependent module is the domain module.

As a web service, DeepTutor is accessible by any learner, anytime, anywhere. Indeed, DeepTutor can be accessed through an HTML5-compatible browser from desktop computers and mobile devices such as smartphones and tablets. As such, it ought to be generally compatible with a generic framework such as GIFT. However, integrating DeepTutor into the GIFT framework will require some important modifications. For one thing, although DeepTutor also employs a a modular design, the modules are organized differently than those in GIFT.. For instance, assessment is a separate, independent module from the domain module. More modules means that the development and management of these modules, especially when scaling up, is more feasible. Assessment is a big topic as well as is domain modeling (knowledge acquisition and representation) which may suggest that they have to be separate modules for scalability purposes. Also, because DeepTutor relies on LPs there is an intrinsic relation between the domain knowledge and learner model. The learner's level of understanding is not recorded in the persistent learner model, stored in a database to be updated and used over time, as a set of learner characteristics and their values (as suggested by the GIFT developers) but rather as pointers to the LP. This arrangement makes the learner and domain modules tightly connected. If the two modules should be decoupled as suggested by GIFT, then the learner module becomes domain-dependent. Indeed, the characteristic-value learner model implies that the learner model should include domain-specific information because some of the learner characteristics reflect learner's performance in a target domain, e.g. how well the student understands Newton's 3rd law. That is, the learner module needs to include domain-specific performance measures that need be traced.

A potential solution would be for GIFT to specialize its proposed specifications for various types of ITSs, similar to specialization principles outlined in Pavlik, Mass, Olney, and Rus (2012).

ACKNOWLEDGMENT

This research was supported in part by Institute for Education Sciences under awards R305A100875. Any opinions, findings, and conclusions or recommendations expressed in this material are solely the authors' and do not necessarily reflect the views of the sponsoring agencies.

REFERENCES

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. Science Education, 93, 389-421.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*, 167-207.

Baxter, G. and Junker, B. W. (2001). Designing Developmental Assessments: A Case Study in Proportional Reasoning. Paper presented at the Annual Meeting of the National Council of Measurement in Education, April 2001, Seattle, WA.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5, 7–74.

Conati C., Gertner A. and VanLehn K. (2002). Using Bayesian Networks to Manage Uncertainty in Student Modeling . Journal of User Modeling and User-Adapted Interaction, vol. 12(4), p. 371-417.

Corcoran, T., Mosher, F.A., & Rogat, A. (2009). Learning progressions in science: An evidencebased approach to reform. Consortium for Policy Research in Education Report #RR-63. Philadelphia, PA: Consortium for Policy Research in Education.

Doignon, J.P. & Falmagne, J. C. (1999). *Knowledge Spaces*. Berlin, Germany: Springer.

Driver, R., Leach, J., Scott, P., & Wood-Robinson, C. (1994). Young people's understanding of science concepts: Implications of cross-age studies for curriculum planning. Studies in Science Education, 24, 75–100.

Duschl, R., Maeng, S. & Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis, Studies in Science Education, 47:2, 123-182.

Graesser, A. C., Rus., V., D'Mello, S., K., and Jackson, G. T. (2008). AutoTutor: Learning through natural language dialogue that adapts to the cognitive and affective states of the learner.

In D. H. Robinson & G. Schraw (Eds.), Current perspectives on cognition, learning and instruction: Recent innovations in educational technology that facilitate student learning (pp. 95–125). Information Age Publishing.

Johnson, P. & Tymms, P. (2011). The Emergence of a Learning Progression in Middle School Chemistry. J. Res. Sci. Teach. 48(8): 849-877.

Lintean, M., Rus, V. & Azevedo, R. (2012). Automatic Detection of Student Mental Models during Prior Knowledge Activation in MetaTutor. International Journal of Artificial Intelligence in Education, 21(3), pp. 169-190.

Lehrer, R. & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), Advances in instructional psychology: Education design and cognitive science (vol. 5, pp. 101-169). Mahway, NJ: Lawrence Erlbaum Associates.

Mitrovic, A., Martin, B., & Suraweera, P. (2007). Intelligent tutors for all: The constraint-based approach. *IEEE Intelligent Systems*, *22*(4), 38-45.

Mohan, L., Chen, J., &Anderson,W.A. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. Journal of Research in Science Teaching, 46, 675–698.

National Research Council. (2001). Knowing what students know: The science and design of educational assessment. (J.W. Pellegrino, N. Chudowsky, & R. Glaser, Eds.). Washington: National Academy Press.

National Research Council. (2005). Systems for state science assessment. (M.R. Wilson & M.W. Bertenthal, Eds.). Washington, DC: National Academy Press.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K–8. (R.A. Duschl, H.A. Schweingruber, & A.W. Shouse, Eds.). Washington: The National Academies Press.

Neumann, K., Viering, T., Boone, W.J., & Fischer, H.E. (2012). Towards a learning progression of Energy. Journal of Research in Science Teaching, Nov 2012.

Jin, H. and Anderson, C. W. (2012), A learning progression for energy in socio-ecological systems. J. Res. Sci. Teach., 49: 1149–1180.

Pavlik Jr., P. I., Maass, J., Rus, V., & Olney, A. M. (2012). Facilitating Co-adaptation of Technology and Education through the Creation of an Open-source Repository of Interoperable Code. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K.-K. Panourgia (Eds.), Proceedings of the 11th International Conference on Intelligent Tutoring Systems (pp. 677-678). Berlin: Springer.

Renkl, A., Worked-out examples: Instructional explanations support learning by self-explanations. Learning and Instruction, 2002. 12: p. 529-556.

Roberts, L.,Wilson, M., & Draney, K. (1997, June).The SEPUP assessment system: An overview. BEAR Report Series, SA-97-1. University of California, Berkeley.

Rus, V., D'Mello, S., Hu, X., & Graesser, A. C. (in press). Recent advances in conversational dialogue systems. AI Magazine.

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of US mathematics and science content standards from an international perspective. Journal of Curriculum Studies, 37(5), 525–559.

Shute, V. J. (1993). A macroadaptive approach to tutoring. *Journal of Artificial Intelligence in Education*, 4(1), 61-93.

Songer, N.B., Kelcey, B., & Gotwals, A.W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. Journal of Research in Science Teaching, 46, 610–631.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S., & Holden, H.K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). Downloaded from www.gifttutoring.org on November 30, 2012.

Steedle, J.T., & Shavelson, R.J. (2009). Supporting Valid Interpretations of Learning Progression Level Diagnoses. Journal of Research in Science Teaching, 46(6), 699-715.

Stevens, S. E., Delgado, C., & Krajcik, J. S. (2009). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*.

Valverde, G.A., & Schmidt.W.H. (1997). Refocusing U.S. math and science education. Issues in Science and Technology, 14(2), 60-66.

VanLehn, K. (2006) The behavior of tutoring systems. International Journal of Artificial Intelligence in Education. 16 (3), 227-265.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. Educational Psychologist, 46, 4, 197-221. [489 KB PDF]

Wilson, M. (2005). Constructing Measures: An Item Response Modeling Approach. Mahwah, NJ: Erlbaum.

Wilson, M. (2009). Measuring Progressions: Assessment structures underlying a learning progression. Journal for Research in Science Teaching, 46(6),716-730.