The PSLC Theoretical Framework

June 15, 2006
Kurt VanLehn (editor) and the PSLC

This working paper describes both the current version of the PSLC theoretical framework and its application to most of the current PSLC empirical studies. It is intended to facilitate collaboration and communication within the PSLC and beyond. Because it is revised every 6 months, it should not be cited or quoted. Citable publications of the content appear in the References section.

1 The intended scope and granularity of the framework

The PSLC theoretical framework is intended to cover a critical, focused portion of the many kinds of learning studied by learning scientists. This section defines the kinds of learning that the framework is intended to address.

Our studies are intended to produce results that are easily and reliably applied by teachers and instructional designers. Thus, the theoretical framework covers only learning by humans as opposed to animals, neurons or institutions. In particular, it focuses on

- academic learning, that is, the learning of students in their courses, where
- the learning gains are statistically reliable and large when compared to the learning gains of the control instruction, and
- the learning is sustainable, in that the experimental instruction does not become ineffective after a few minutes or hours.

In these respects, the scope of our theoretical framework is somewhat narrower than the scope of existing educational theories, including Bloom's taxonomies (1956), Sweller's cognitive load theory (e.g., Sweller, van Merrienboer, & Pass, 1998), Vygotsky's activity theory (Vygotsky, 1978), Mayer's multimedia learning theory (e.g., Clark & Mayer, 2003), instructional design theory (Reigeluth, 1999) and many others.

Following Newell's *Unified Theories of Cognition* (1990), which classified cognitive processes by their duration, we study learning events that take place from about 1 second up to about an hour and accumulate over weeks and months. In Newell's terminology, this spans the upper part of the "cognitive band" where phenomena are strongly constrained by properties of the human cognitive architecture, and all of the "rational band," where phenomena are constrained mostly by knowledge or information content independent of what kind of information processing architecture (i.e., human vs. computer) is utilizing the knowledge. Thus, our theoretical framework is at a higher level of granularity than existing models of the human cognitive architecture, such as Anderson's ACTR (Anderson, 2002) or Kieras and Meyers' EPIC (1997) which in turn are at a higher level of granularity than cognitive neuroscience theories. As Bruer (1997) has well argued, transferring results from the finer granularity of cognitive neuroscience to the higher granularity of classroom learning is "a bridge too far".

The PSLC theoretical framework focuses on explaining and predicting *robust learning*. Learning is robust if the acquired knowledge or skill meets at least one of the following three criteria:

- *Retention*: It is retained for long periods of time, at least for days and even for years.
- *Transfer*: It transfers, that is, it can be used in situations that differ significantly from the situations present during instruction.

- *Future Learning*: It accelerates future learning. That is, when related instruction is presented in the future, this knowledge allows them to learn more quickly and effectively learn from it.

2 The micro and macro levels

The theoretical framework has two levels: a micro level and a macro level. The micro level describes student's behavior and thinking in a highly detailed way, where the unit of analysis is an event that lasts only a few seconds. The macro level describes learning taxonomically, as different types with different properties.

In many respects, these two levels act like the atomic and molar levels in elementary chemistry. At the atomic level, chemical explanations are built in terms of three-dimensional assemblies of atoms, valencies, bond energies, etc. At the molar level, chemical explanations are built in terms of classes of compounds, such as acidic vs. basic, soluable vs. insoluable, organic vs. inorganic, etc. Both levels are important and serve different purposes. Similarly, the two levels of our theoretical framework are both important and serve different purposes.

2.1 The macro level

The macro level of the theoretical frame is a classification of *robust learning processes.* The term *robust learning process* denotes both an instructional process that is intended to cause robust learning outcomes and the learning/cognitive process that it is intended to promote. There are many robust learning processes, and it is the macro level's job is to formulate a useful taxonomy for them. The current taxonomy is described below:

- *Sense-making*. These are robust learning processes wherein students try to understand the instruction or engage in higher-level thinking to create knowledge independent of instruction. PSLC research focuses on two types of sense-making processing.
    - *Coordinative learning*: When students go beyond direct instructional feedback to learn on their own by integrating results from multiple input sources, representations, or reasoning strategies, we say they are engaged in coordinative learning. Coordinative learning includes co-training (Blum & Mitchell, 1998), a theoretically sound technique in machine learning for using multiple input sources to perform unsupervised learning from data that does not include correct responses or feedback. Coordinative learning also includes other ways of learning from "multiples" including multimedia, multiple representations and multiple strategies.
    - *Interactive communication*: When two agents takes turns with each other, share initiative during the instruction, and may explore an idea at arbitrary depth, then we say they are engaged in interactive communication. We mean to include natural language dialogues between a student and a peer

or a tutor as well as other non-verbal (e.g., computer interface mediated) forms of dialogue.

- *Fluency and refinement.* Some knowledge and skill must be mastered in order to provide a foundation for subsequent learning. The learning processes involved in obtaining such mastery appear to be somewhat different than those involved in obtaining an initial understanding of the knowledge or skill. They seem to involve *refinement*, which is making modifications to the knowledge itself and more importantly, to the conditions under which the knowledge should be applied. However, even after the knowledge becomes well-understood and its information content is not changing much, fluency-building processes continue to act, leading not only to correct performance, but eventually to fast and effortless performance. Part of the study of fluency is determining how it accelerates further learning. Although it is widely believed that mastery of the parts facilitates more efficient sense making of the whole, it is not clear exactly how this acceleration works. Such open questions that address knowledge-based influences on learning are just the kind that the PSLC is uniquely positioned to address in ways that cognitive neuroscience cannot.

Cognitive processes are controlled by the student. No matter what kind of instruction is given to students, they can usually find a way to get through it without learning. Thus, part of understanding what is going on during instruction is understanding the student's self-regulative or *metacognitive strategies*. Thus, all these taxonomic categories involve studying processes that are a mixture of cognitive and metacognitive processing. We tend to think of the sense-making processes as having a higher degree of metacognitive involvement than the foundational skill building processes, but that remains to be seen.

2.2 The micro level

The micro level of our framework takes a much more detailed view of the student's cognition. We assume that learning results from the acquisition of *knowledge* (and skills, which we consider a subclass of knowledge), and that knowledge is decomposable into a large number of small *knowledge components*. Many knowledge components are abstract, in that they can be applied by the student in many situations. For instance, a word's pronunciation is a knowledge component, and knowing it allows the student to say the word on many occasions.

A knowledge component is defined by its content. That is, the definition is at Newell's rational level—information. How that information is represented in the cognitive architecture of the student may change over time. For instance, in ACT-R terms (Anderson, 2002), a knowledge component might initially be presented to the student as text, be converted by a reading process to a declarative encoding, and later be re-encoded procedurally as the student applies it over and over. That is, even though the encoding may change, if the different mental representations all have the same information and produce the same ultimate decision, then they represent the same knowledge component. We do not mean to say that changes in encoding guarantee equivalence in knowledge components and, indeed, learning produces many related knowledge components that are more or less accurate approximations of a target correct knowledge component.

A *learning event* is a time interval in the life of the student, usually lasting a few seconds or a minute, wherein the student applies or constructs a knowledge component. For instance, the following are all learning events involving the same target geometric knowledge component (to be sure, the knowledge component(s) learners acquire from these learning events may be not be identical, for instance, including irrelevant features and missing relevant ones):

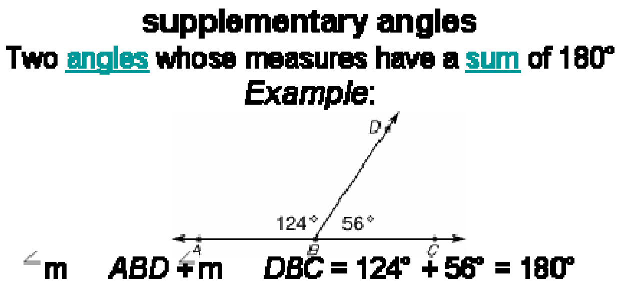

Figure 1 Supplementary angles defintion including an example.

- reading a definition of supplementary angles, such as the one at the top of Figure 1.
- studying an example of supplementary angles, such as the one at the bottom of Figure 1.
- using supplementary angles in a proof
- explaining supplementary angles to another student
- incorrectly selecting a supplementary angle pair from set of angle pairs, getting feedback from a tutor, and then correctly selecting the supplementary pair.
- Starting to write a definition of supplementary angles, realizing that one is uncertain of the distinction between supplementary angles and complementary angles, looking supplementary angles up in the textbook, and writing a correct definition for supplementary angles.

A single learning event often mixes application, construction and/or refinement of the knowledge component.

Descriptions of learning research often discuss the design of the instruction and its intended effects on students. However, such descriptions often do not get to the level of specific learning events and how these lead to new or modified knowledge components. Facilitating such description is a key goal of PSLC theory development. In such descriptions of intended instruction, the term "learning events" refers to events that are intended to occur.

In instructional settings, learning events are often evoked by instruction in one or more of three forms 1) a verbal *description* (e.g., definition, rule, or principle) of a knowledge component, 2) an *example* of a knowledge component's applications, or 3) an opportunity to *practice* a knowledge component's application. This three-part distinction is an important refinement of the well-known explicit vs. implicit learning distinction (Dienes & Perner, 1999; Reber, 1967). The sole source of explicit learning is verbal description, whereas implicit learning can occur during either an example or a practice opportunity.

Although a knowledge component is a piece of information, and can exist anywhere information does (e.g., on a page; in a computer; in a human brain; in a conversation), when it is stored in a human brain, the probability of successfully recalling it is roughly proportional to the *strength* of the knowledge component's encoding in memory, and the *feature validity* of the way in which it is encoded. The feature validity of a knowledge component measures how well the features associated with the mental representation of the knowledge component match the features present during all situations where the component should be recalled.[1] Strength is roughly proportional to a knowledge component's frequency (more times accessed is better), recency (more recent access is better), spacing (longer times between access is better), and utility (more times employed successfully is better).

Learners will acquire partially correct knowledge components with features that are different from those in a target correct knowledge component. A learner may acquire any of a cluster of knowledge components that all have the same mental or physical response as the target, but may be missing relevant features or including irrelevant features. The theory thus can model common student errors (a knowledge component with over-generalized features) or failures to transfer (over-specialized features) in a uniform way that unpacks commonly used, but vaguely defined, notions like "misconceptions" or "inert knowledge" (cf., Bransford, Brown, & Cocking, 1999). Further, learners have often acquired more than one such knowledge component in a cluster and these compete with each other with their strengths determining access and changing over time. This explains why student errors may disappear for a while but then reappear—the correct knowledge component may lose strength and an incorrect component may become relatively stronger (Siegler, 1996).

These two concepts, strength and feature validity, are meant to be broad and yet consistent with many lower-level theories of human memory, attention, reasoning, categorization, problem solving. Although details of these lower-level theories are left out, this broad-strokes theory provides adequate and cost-effective leverage for explaining and predicting classroom learning phenomena and for theoretically generating and evaluating instructional innovations. Just as chemistry theory can be built on physics theory, we may want to link robust learning theory to such more detailed cognitive and cognitive neuroscience theories. However, just as chemistry theory is necessary in its own right, so too is robust learning theory. Lower level computational and neuro-biological theories *alone* are not sufficient to yield understanding of higher-level learning and instruction.

2.3 Connections to robust learning

Both sense-making and fluency building can lead to robust learning. That is, they can increase retention, transfer and acceleration of future learning. In order to exercise the

1 Feature validity is a generalization of the standard concept of cue validity. Cues are usually understood to be perceptual or at least rapidly computed (MacWhinney & Bates, 1989). The term "features" includes cues as well as higher level properties, such as those used by experts but not novices (Chi, Glaser & Feltovitch, 1981).

terminology and to preview some of the explanations that accompany specific studies, let us consider first how each process can cause robust learning.

Sense-making can yield robust learning: 1) by allowing students to apply domain fundamentals to re-derive partially forgotten knowledge at long retention intervals, 2) by providing conceptual understanding that can be adapted or transferred to novel situations, 3) by allowing students to more quickly learn new material by having well-practiced meta-cognitive strategies (like systematic discovery strategies) to make sense of it and learn on their own. Although all three pathways to robust learning—*rederivation*, *adaptation* and *self-supervised learning*—can be used by individual learners, we hypothesize that they can be significantly enhanced through certain collaborative processes, including instructional dialogues and peer collaboration.

Fluency-building activities involve practice with attention to appropriate deep features and to the point of fluent execution. This can yield robust learning by allowing students 1) to develop strong, well-practiced mental connections that survive long retention intervals, 2) to encode and represent instructional situations in terms of well-learned and general features that support transfer to novel future situations, 3) to have acquired connections to a level of automaticity that reduces cognitive load during new learning, leaving "headroom" to apply sense-making strategies that accelerate new learning. In contrast to the three pathways of the sense-making path, which are primarily meta-cognitive or social, these three mechanisms, *strengthening*, *deep feature perception* and *headroom*, here are primarily cognitive.

3 Analyzing learning events

Our analytical technique has its roots in the early history of cognitive science. Before 1950, psychologists treated the time interval during which a problem was being solved as a black box. They varied the types of problems (the inputs to the black box) and measured the success rates of the solvers (outputs), but they did not look at what the solvers actually did during the problem solving. Newell and Simon (1972) pioneered a new method of analysis, where they observed solvers at work on problems, sometimes collecting new kinds of data (e.g., verbal protocols) in order to elucidate what the solvers were thinking about as they worked. Newell and Simon analyzed an individual's struggle to solve a problem as a chronological path through a so-called *problem space*. The problem itself dictates the topology of the problem space, which in turn determines the paths *available* to solvers. The problem solver determines the choice of a path (or paths). Some paths lead to solutions, and some do not.

The PSLC has begun to apply the same analytic technique to learning events. That is, a learning event is viewed as analogous to the "problem" that the student is invited to solve. There are many possible paths. Some lead to robust learning, and some do not. The learning event itself dictates what paths are *available*, but the student's decisions determine the particular path that the student follows. We use a *learning event space* to define the set of paths that students have available for a particular learning event. The first section introduces the concept in more detail, and the second section shows how we have used it to integrate results across studies.

3.1 *Learning event spaces*

When researchers design instruction, they usually have a fairly clear idea of what they would like the student to do during the learning events. When their idea is written out in detail for a learning event, it comprises one or more paths through the learning event. However, students can almost always find other ways to make their way through the learning event. These may involve guessing, or copying, or playing dumb in order to extract hints. A learning event space should include all possible paths, including both the paths that the designer intended for students and the paths that students invent for themselves.

As an example of a learning event space, suppose physics students are analyzing the motion of an elevator that is descending and slowing down, and the tutor asks them to draw the elevator's acceleration. This is learning event for the knowledge component, "If an object is moving in a straight line and slowing down, then it has an acceleration that is parallel to its velocity and in the opposite direction." A typical tutoring system would give the student positive feedback if they drew the correct acceleration (i.e., pointing straight up) and negative feedback and hints otherwise. One learning event space is:

Start

1. Student applies a shallow strategy, e.g., guessing or copying from a hint.
   - 1.1. The entry (an acceleration vector) is correct → Exit, with little learning
   - 1.2. The entry is incorrect and the tutor gives a hint → Start
2. Student recalls some knowledge and applies it
   - 2.1. The entry is correct → Exit, with learning via strengthening and refinement
   - 2.2. The entry is incorrect and the tutor gives a hint → Start
3. Student reads over the hints and any other information sources available, conjectures some knowledge and applies it.
   - 3.1. The entry is correct → Exit, with learning via knowledge construction
   - 3.2. The entry is incorrect and the tutor gives a hint → Start

This learning event space says that the students have a choice of 3 tactics. Paths 2 and 3 are ones that the designer intended students to take, because they can yield useful learning, whereas path 1 is a collection of shallow methods for getting through the learning event. Each leaf in this branching tree of paths ends with an arrow showing where the student goes next. The notation "→ Start" means that the student goes to the top and starts over, whereas "→ Exit" means that the student is finished with this learning event. For instance, this learning event space indicates that whenever a student gives an incorrect response, the tutor gives the student a hint and asks the student to try again.

Just as problem solving can be analyzed at different levels of detail, so too can learning events. Different degrees of abstraction are useful for different analytical purposes. For instance, the learning event space above includes two paths, 2 and 3, that differentiate between learning via strengthening and refinement versus learning via knowledge construction. One would use this analysis in order to articulate a hypothesis that contrasts these two cognitive processes (as do Booth, Siegler, Koedinger and Rittle-Johnson; see below). If, on the other hand, one only cared that useful learning occurred and did not care to distinguish what kind of learning it was, then the following learning event space would be more appropriate:

Start

1. Student applies a shallow strategy, e.g., guessing or copying from a hint.
   1.1. The entry (an acceleration vector) is correct → Exit, with little learning
   1.2. The entry is incorrect and the tutor gives a hint → Start
2. Student applied a deep strategy, such as recall or construction of knowledge.
   2.1. The entry is correct → Exit, with learning
   2.2. The entry is incorrect and the tutor gives a hint → Start

This learning event space is more abstract (less detailed) than the earlier one, but it is intended to describe the same events.

3.2 Four common learning event spaces

Although a learning event space can have any topology, the learning event spaces for most studies in the PSLC seem to have one of four topologies. This section briefly describes them.

The first common structure is for studies where students are given some instruction (e.g., a diagram) and have a choice of how they will study it (e.g., to self-explain it or not):

Start

1. Student applies a shallow strategy → Exit, with little learning
2. Student applies a deep strategy → Exit, with learning

The second common topology occurs when a tutoring system gives a student immediate feedback and remedial instruction, but it does not ask the student to try again. This is common, for instance, in vocabulary drills. Such learning spaces look like this:

Start

1. Student applies a shallow strategy
   1.1. The entry is correct → Exit, with little learning
   1.2. The entry is incorrect and tutor gives remedial instruction
      1.2.1. Student processes the remediation shallowly → Exit, with little learning
      1.2.2. Student processes the remediation deeply → Exit, with learning
2. Student applies a deep strategy

2.1. The entry is correct → Exit, with learning
2.2. The entry is incorrect and tutor gives remedial instruction
2.2.1. Student processes the remediation shallowly → Exit, with little learning
2.2.2. Student processes the remediation deeply → Exit, with learning

The third common topology occurs when a tutoring system gives the student immediate feedback, and when the student's entry is incorrect, it gives an unsolicited hint and asks the student to try again. A typical analysis does not distinguish how the student processes the hints at the point where the hint was given (as in the learning event space above), but instead allocates that cognitive processing to the student's next attempt. Thus, the learning event space is:

Start

1. Student applies a shallow strategy, e.g., guessing or copying from a hint.
   1.1. The entry is correct → Exit, with little learning
   1.2. The entry is incorrect and the tutor gives a hint → Start
2. Student applied a deep strategy, such as recall or construction of knowledge.
   2.1. The entry is correct → Exit, with learning
   2.2. The entry is incorrect and the tutor gives a hint → Start

The fourth common topology occurs when a tutoring system give immediate feedback after attempts, but only gives hints when asked. The learning event space is:

Start

1. Student applies a shallow strategy, e.g., guessing or copying from a hint.
   1.1. The entry is correct → Exit, with little learning
   1.2. The entry is incorrect → Start
2. Student applied a deep strategy, such as recall or construction of knowledge.
   2.1. The entry is correct → Exit, with learning
   2.2. The entry is incorrect → Start
3. The student asks for and receives a hint → Start

Distinguishing between the third and fourth topologies may not be particularly useful. In real tutoring systems, such as the Cognitive Tutors or Andes, hints are given automatically after some errors and not after other errors, and students can always ask for hints. Thus, these tutors are a blend of the third and fourth topologies. Although one could in principle write out a learning event space that captured the tutoring system's hint-giving policies completely accurately, that often does not make the analysis of the student's learning any clearer. As a general rule, distinguishing between topologies 3 and 4 is inadvisable if the hypothesis being investigated does not depend on whether a student who makes a mistake gets a hint without asking or has to ask for a hint (which usually requires only one mouse click).

3.3 Path choices versus path effects

Learning event spaces are constructed by researchers in order to decompose a difficult theoretical question—what learning occurred here?—into two simpler questions:

- *Path choice*: Which paths are students taking in the normal or control instruction? Which paths are they taking in the experimental instruction or treatment? Does this tell us anything in general about how to get students to take paths that they should take?
- *Path effects*: Given that a student has gone down a particular path, what are the effects on the student's knowledge? Does this path tend to cause immediately detectable gains? Robust learning? Does this tell us anything in general about what kinds of paths are effective?

In order to illustrate this decomposition, suppose an experimenter wished to test the hypothesis that abstract problems are sometimes more effective than concrete problems (e.g., Ohlsson, 1993). An abstract version of the elevator problem is, "A helicopter is moving in a straight line and slowing down. What direction is its acceleration?" This version of the problem does not specify the direction that the helicopter is moving, so a correct answer must be abstract, that is, it must mention the relationship between its motion and the acceleration. In particular, suppose the tutoring system were modified to accept typed descriptions of the acceleration's direction e.g., "in the opposite direction of its motion." Would this version of the tutoring system with its typed, abstract directions be more effective than the standard version wherein students draw concrete vectors with the mouse?

Let us use the learning event space shown below to decompose this research question. When comparing two manipulations, it is usually helpful to merge their learning event spaces. In this case, we are comparing two ways to enter directions of vectors, and we assume that both have a shallow path and a deep path, as in the learning event space showing most recently. The shallow paths are merged, but the two deep paths remain distinct, yielding this learning space:

Start

1. Student applies a shallow strategy, e.g., guessing or copying from a hint.
   1.1. The entry is correct → Exit, with little learning
   1.2. The entry is incorrect and the tutor gives a hint → Start
2. The student tries to apply a deep strategy (and the tutor requires typed, abstract directions):
   2.1. The entry is correct → Exit, with learning
   2.2. The entry is incorrect and the tutor gives a hint → Start
3. The student tries to apply a deep strategy (and the tutor requires drawn, concrete directions):
   3.1. The entry is correct → Exit, with learning
   3.2. The entry is incorrect and the tutor gives a hint → Start

First, let's consider path choice. Which tutoring system causes more students to choose the deep paths (2 or 3) as opposed to the shallow path (1)? When students must type in abstract directions, it is probably hard to guess a correct answer. Nonetheless, students who are determined to avoid learning (i.e., to exit via 1.1) can deliberately make mistakes, thus eliciting more and more hints from the tutoring system, and eventually getting a hint that tells them exactly what to type in. But then they must type it. This all takes some time and effort, so students may soon decide that it is just easier to try to recall the appropriate knowledge. Thus, it seems likely that typing abstract descriptions decreases exits via 1.1. If this were the only consideration, then typing abstract descriptions should cause more learning than drawing concrete directions.

Now let's consider path effects. The learning that occurs via path 2.1 involves typing the important concept "opposite to the motion/velocity." This may strengthen the student's memory of the knowledge component. On the other hand, the learning that occurs via path 3.1 involves a visual-motor instantiation of the conceptual knowledge, namely the drawing of a concrete vector. This dual coding may strengthen the student's memory of the knowledge component. It is not clear which path produces more robust learning of the knowledge component. It would be interesting to find out.

Without the learning event space analysis, we faced a complex question, "Is typing abstract vector directions more effective than drawing concrete directions?" With the learning event space analysis, we now face two somewhat simpler questions, namely:

- *Path choice*: Compared to drawing concrete directions, does typing abstract directions decrease the number of students of exiting via path 1.1 (shallow learning)?
- *Path effects*: Do students who exit via path 2.1 have more robust learning for the knowledge component than students who exit via path 3.1?

In general, path choice is partly determined by physical affordances (e.g., one tutoring system has typed input; the other has drawn input) and partly determined by the student's motivations and beliefs. Although the designer can influence path choice, students make the final decisions about what paths to take. On the other hand, path effects are often not under the conscious control of the students. They are often determined by the student's cognitive architecture. A comprehensive theory of learning has to deal with many kinds of issues, and the distinction between path choice vs. path effects nicely separates two rather different types of issues: motivational and architectural.

A learning space analysis also suggests measurements that could answer the questions it raises. In particular, log data from tutoring systems can often determine the frequency of certain path choices. Verbal protocols may also be useful for measuring path frequencies.

3.4 Relationship to the macro level

As mentioned earlier, the macro level categorizes robust learning processes into sense making processes and foundational skill-building processes. Sense-making usually refers to the early stages of instruction when the student is trying to understand the instruction and construct a coherent, meaningful, operational knowledge component. For instance, self-explaining a text and self-explaining a diagram are both sense-making processes. On the other hand, foundational skill building often focuses on the later stages of instruction where students are consolidating their knowledge via practicing its application.

The major issue for sense making is whether students actually engage in the intended sense-making process or find some way to avoid it. For instance, when students are asked to explain a text to themselves, they often just paraphrase it instead of trying to really explain it. Thus, most studies in the PSLC sense making clusters focus on path choice, as opposed to path effects. In particular, they attribute the success or failure of instruction to the frequency of good vs. poor paths.

On the other hand, studies of foundational skill-building often focus on path effects. A typical study may provide two forms of instruction so that it can compare the robustness of the learning they generate. For instance, Juffs and Eskenazi are comparing explicit to implicit methods for teaching ESL reading vocabulary. Students either click on a new word in a text in order to pop up its definition, or they must infer the word's meaning from its context in the text. A learning event space for this study is:

Start

1. Student ignores the to-be-learned word → Exit, with little learning
2. Student clicks on the word, studies the definition, and tries to learn its meaning → Exit, probably with explicit learning
3. Student tries to infer the word's meaning from context → Exit, probably with implicit learning

The study tries to insure that the manipulation (providing definitions or not) has little affect on the frequency of path 1. In particular, the students are tested on the target vocabulary and the tests count as part of the students' grade. This should keep the frequency of path 1 low in both the explicit condition and the implicit condition. This allows the researchers to interpret results in terms of path effects—is learning from an definition better or worse than learning from context? Many studies in the fluency cluster have this same analysis. They employ a learning event space that has two or more "good" exits; they try to control path choices; they measure robustness and attribute it to the effects of the paths.

3.5 A few final comments about learning event spaces

The PSLC has just begun to analyze its studies in terms of learning event spaces. The analyses below were inspired by discussions inside the Executive Committee, then developed by the Co-directors working individually with each project leader. This initial group comprises a significant fraction of the PSLC membership and the idea of learning event space analysis is beginning to spread more widely. We have observed PSLC faculty, post docs, and PhD students working hard to think with the PSLC conceptual framework and struggling to make the link between the macro and micro levels. It appears that learning event space analysis is helping members to make that link.

The concept of a learning event space clearly separates the designers' impact on learning from the student's impact. Basically, the student has final control over path choice. The designer can only provide opportunities; it is up to the student to take them or not. This is why one often finds that highly motivated students can learn well from almost any instruction, and that even the most carefully designed instruction fails to benefit poorly motivated students.

The concept of a learning event space also explains why it is so important for the PSLC to study learning *in vivo*, rather than using paid subjects who are students in research universities. The frequency distribution of path choices partly determines learning outcomes, and the frequency distribution of path choices in an authentic situation are probably quite different than those of a laboratory situation.

Learning event spaces also explain the PSLC's dedication to collecting and analyzing log data from tutoring systems. Log data often partially reveal student's paths through learning event spaces.

As mentioned earlier, the PSLC is striving to develop a common theoretical framework for explaining robust learning. This may lead eventually to a theory, but it is not itself a theory. Learning event spaces are a component of this framework. They too are not a theory, but will hopefully accelerate our progress toward a theory.

4 Coordinative learning cluster

The Coordinative Learning cluster is one of two sense-making clusters. (The other one is Interactive Communication.) The studies in the Coordinative Learning cluster tend to focus on varying the types of information available to learning, and in particular, on the impact of having learners coordinate two or more types. For instance, a study might compare various presentations of material, such as different combinations of written text, spoken text and diagrams; another might contrast the content of the instructional material, such as exercises meant to teach procedures versus those meant to reinforce concepts.

Most of the studies are testing methods for increasing sense-making processes and decreasing shallow tactics such as guessing or copying from hints. Thus, their expected benefits can be analyzed in terms of changing students' path choices. Various methods for changing path choices are being explored.

Perhaps the most common generic technique is modifying the instruction to make the sense-making path easier. For instance, in the Aleven and Butcher study discussed below, students must coordinate two types of information: a geometry diagram and some geometry statements. The statements refer to parts of the diagram via symbolic names, such as $\angle$ABC. The experimental conditions test different ways of making it easier for students to coordinate the diagrams with the text, such as coloring the symbolic terms and the corresponding diagrammatic elements with the same colors. The hypothesis is that this will make the sense-making paths easier to follow, and thus increase the frequency of beneficial path choices.

A second common technique is instructing students on the benefits of the sense-making path before training begins, and then to prompt them at each opportunity in order to remind them to take the sense-making path. This of course does not guarantee that they will take the better path, but it should raise the frequency. This technique is being used, for instance, in the Ogan and Aleven study, where prompting is used at critical locations in a film in order to increase self-explanation of cultural events in the film.

A number of Coordinative Learning projects are comparing alternative positive learning event paths or the effects of *coordinating* results from alternative paths. For instance, Booth, Siegler, Koedinger and Rittle-Johnson are comparing two paths, recall of knowledge components vs. (re-)construction, in solving algebraic equations. Students who have both conceptual and procedural knowledge of the domain are expected to use both paths, while students who lack conceptual knowledge are limited to using only the recall path to solve problems. Booth et al. will prompt students to self-explain correct and incorrect worked examples to provide them with (or reinforce) the conceptual, construction pathway; previous research has shown that self-explanation of both correct and incorrect answers improves students' strategy use (Siegler, 2002). This study will vary the types of exercises presented (procedural alone versus procedural plus self-explanation) in order to examine the impact of having both paths on measures of robustness. In similar work in a different domain, Davenport and Klahr are comparing use of the conceptual and procedural paths for solving chemistry problems; they will vary the media of the information presented (equations vs. equations plus diagrams) and prompt students to self-explain the diagrams to coordinate their conceptual understanding with their procedural knowledge. In addition, two language projects, Pavlik's and Liu and Perfetti's, are exploring how to facilitate better vocabulary learning through evoking alternate retrieval paths (e.g., foreign spoken to English written, picture to foreign written, etc.) both during learning and at post-test.

4.1 Visual-Verbal Learning in Geometry (Aleven & Butcher)

This project is attempting to use visual affordances to decrease the effort associated with coordinating diagrammatic information with textual information and verbal domain knowledge. Ultimately, we seek to understand how coordination of visual and verbal information supports processes of robust learning, thus informing the development of learning technologies in domains that use multiple representations. The visual supports that we are building into the Geometry Cognitive Tutor should make it easier for students to integrate visual and verbal information during geometry problem solving, thus encouraging them to choose a deep-learning path through the learning event space.

The existing Geometry Cognitive Tutor requires that students calculate answers for geometry elements in a diagram, such as the measure of angle ABC. In order to answer correctly students should engage in many cognitive activities, including figuring out which angle in the diagram corresponds to the symbolic term, "∠ABC." Finding the referent of "∠ABC" requires locating relevant letters in the diagram and recognizing the geometry relationship(s) that they combine to form (e.g., tracing the lines between points A, B, and C to identify the angle). Determining the reference of a symbolic term increases the cognitive processing required to formulate answers. More importantly, students also must recognize relevant diagram features (angles, arcs, line segments, etc.), encode the symbolic terms of those features, and maintain known and unknown quantities as related to the diagram features and the goal information.

Using in vivo experiments, we are comparing student learning when different methods are used to modify these cognitive processes. One method uses *contiguity*, where students work directly with the visual problem diagrams instead of entering answers in a separate table. In the existing Geometry Cognitive Tutor, symbolic labels on the table where students work ask for the measure of ∠ABC. In the experimental version of the tutor (the "contiguous" interface), a box is placed in the diagram at ∠ABC as well as in the other to-be-solved locations. The student interacts directly with the diagram and the tutor provides the student with feedback and hints, if necessary. Accepted answers are displayed directly in the diagram instead of in the separate table. The contiguous interface developed in this research is also being used in the examples research by Renkl, Aleven & Salden, where it is important that the tutoring system relieve the student of most of the work so that the student can engage in deep learning paths.

Let us analyze the learning events that occur under both the control and experimental conditions. First, consider the most obviously influenced knowledge component—the skill of matching elements in a diagram with their symbolic references. In the control condition –where students work in a table that is separate from the diagram—student must repeatedly transition between the table information and the diagram representation. In contrast, students in the experimental condition—where students work directly with the diagram and see answers displayed within that diagram—must map between the original problem statement and the diagram in order to determine the given information but subsequently they can reason directly with information displayed in the diagram without continual mapping to information, answers, and labels in the table. Thus, students in the control condition receive more practice in mapping between symbolic terms and diagram elements than student in the diagram condition. However, this skill has been practiced frequently before the angle unit of the geometry tutor where the manipulations occur. Thus, the small increase in strength provided by the extra practice of the control condition may be undetectable.

Now let us consider the newly taught knowledge components—the geometry principles about supplementary angles, complementary angles, isosceles angles, sum of angles in a triangle, etc. Earlier work (Aleven & Koedinger, 2002) suggests that students either apply a shallow guessing strategy (e.g., if two angles look equal in the diagram, they are probably equal) or they try to apply meaningful target knowledge components. If they fail to enter a correct answer on the first attempt, they get feedback and may get a hint from the tutor. Students either try to learn from the feedback and hints or not, then resubmit a new answer. In addition to numerical answers, in the existing geometry tutor students also explain their problem-solving steps by naming the geometry principle that justifies their answer. Another planned feature extends this explanation by asking students to explain how the named geometry principle applies to the diagram. However, it is important to note that the learning event space is consistent for each of these tutored features. For each of these potential learning events, the learning event spaces are:

Start

1. Students apply a shallow guessing strategy
   1.1. They enter a correct answer → Exit having not learned much.
   1.2. They enter an incorrect answer → Start
2. Students apply the target knowledge.
   2.1. They enter a correct answer → Exit having learned something.
   2.2. They enter an incorrect answer → Start
3. Students ask for and get a hint → Start

If students generate the correct answer using a shallow guessing strategy, they learn little about the target knowledge. If they generate the correct answer by applying the target knowledge, perhaps following some instructional hints from the tutor, then they learn something about the target knowledge components. Moreover, if students get one or more hints before exiting via 2.1, then the contextual features encoded with this learning event are less valid, because they include hints that will be absent during post-testing. That is, students who takes the path apply(2) → correct(2.1) are more likely to recall the target knowledge components during future learning events than those who takes the path apply(2) → incorrect(2.2) → apply(2) → correct(2.1) or the path guess(1) → incorrect(1.2) → apply(2) → correct(2.1). If we knew the frequency of student's paths through this graph, we can predict roughly how much they learned.

We hypothesize that elimination of the extra work of finding the referent of symbolic angle names, continually mapping between diagram and table, and maintaining relevant feature quantities will have 2 major effects on the frequency of the students' paths through this graph:

- *Errors*: Control students may make errors when finding the referent of a symbolic angle or mapping between found diagram values in a table and their location in the diagram. This increases the frequency of the incorrect answers (1.2 and 2.2), which in turn reduces learning.
- *Estimated effort*: Let us hypothesize that some students choose a path based in part on the estimated effort required to follow the path. In particular, if they estimated that the deep learning path is considerably more effort than the shallow learning path, then they are more likely to take the shallow learning path. Thus, reducing the estimated effort of the deep learning path, as supported by the interactions in the contiguous diagram representation, should increase learning.

Errors and estimated effort affect path choice, but there also is a path effect that deserves consideration, as it is likely to impact meaningful learning:

- *Increased feature validity:* We predict that students who are able to reason directly with diagram representations will attend more closely to the geometric features and relations to which geometry principles apply. That is, even if two students, one in the contiguity condition and one in the traditional condition, both exit via applying geometric knowledge successfully (path 2.1), they will have encoded different features, and in particular, the learning of the contiguity condition students will have higher feature validity. Students who learning using the contiguous representations may better recognize deep features of situations in which the newly learned knowledge components (the geometry principles) apply and, thus, may more successfully apply the target knowledge. That is, students learning directly with diagrams may not only more frequently *choose* path 2, but are more likely to be able to proceed along path 2.1. This would not only increase performance during tutor practice, but would also accelerate future learning by increasing the frequency of path 2.1 in subsequent learning events.

Although the predicted results could be explained by the hypothesis that reducing cognitive load increases learning (Chandler & Sweller, 1991; Sweller & Chandler, 1994), the theoretical framework provides an alternative hypothesis that succinctly explains learning benefits without using circular definitions (reduced cognitive load is often used both to describe the format of beneficial materials and to explain their benefits). Moreover, our hypothesis makes predictions about what paths students make through a learning event and generalize to several types of learning events (answers, geometry principles, etc.). The hypothesized connection between learning outcomes and hypothesized learning paths will be assessed when the log data from the geometry tutor are analyzed.

4.2 Tutoring a meta-cognitive skill: help-seeking (Aleven & McLaren).

This project is essentially trying to *explicitly teach* students to take better paths though the learning event space. More specifically, it is attempting to tutor a meta-cognitive skill, which is to seek help when and only when the help will benefit the student's learning. As mentioned earlier, the geometry tutor can be analyzed in terms of the following learning event space:

Start

1. Students apply a shallow strategy, such as guessing or copying from a hint
   - 1.1. They enter a correct answer → Exit having not learned much.
   - 1.2. They enter an incorrect answer → Start
2. Students apply the target knowledge.
   - 2.1. They enter a correct answer → Exit having learned something.
   - 2.2. They enter an incorrect answer → Start
3. They ask for & get a hint from the tutor → Start

Aleven et al. (2004) found that student often used inappropriate help-seeking strategies. Two strategies predominated. *Help abusers* often follow the path 3 → 3 → …→ 3 → 1 → 1.1. That is, they ask for hints until they reach the bottom out hint, then use the copying strategy. *Help refusers* often follow the path 1 → 1.2 → 1 → 1.2 … → 1 → 1.1. That is, they guess until they get the correct answer. Both abusers and refusers exit via 1.1, so they do not learn much.

The goal of the help seeking tutor is to get student to exit via 2.1—this should cause learning. Moreover, the help seeking tutor tries to get students to take as few cycles through hint(3) as possible. The fewer the hints, the fewer the invalid features encoded along with the target knowledge application, so the more likely the recall during post-testing.

The help-seeking tutor differs from the regular tutor in that it occasionally inserts some extra meta-cognitive advice at 1.2 and at 3. This could have two benefits:

A. First, the advice may convince students to deliberately change their help-seeking strategies and choose path (2) more frequently.
B. Second, the student could learn that guessing (1) and asking for a hint (3) sometimes slow them down with extra advice, so in order to increase their speed, they should choose applying their knowledge (2) more frequently.

A preliminary analysis of an *in vivo* study's data suggests that the help seeking tutor did increase domain learning, but it apparently did not increase appropriate help seeking behavior when it was turned off. If this result holds up in the final data analysis, then it suggests that explanation B holds but explanation A does not. That is, the manipulation succeeded in changing the frequency of the students learning paths while the help seeking advice was active, but it did not succeed in changing their beliefs about the utility of help-seeking strategies.

4.3 Handwriting in algebra learning (Anthony, Yang & Koedinger)

In the Algebra LearnLab course's tutoring system, and many other tutoring systems such as the ones used in the Physics and Chemistry Learnlab courses, algebraic expressions are entered with a keyboard. Anthony, Yang and Koedinger are determining if the acquisition of equation solving skill in algebra is facilitated by using handwriting instead. The claim is mostly that handwriting will affect student's choice of path, and in particular, that it will encourage them to travel a deep learning path more frequently.

In these study, students alternated between copying an example and solving a similar problem with the example still present. When solving the problem, they entered the final numerical answer into the computer. If the answer was incorrect, they were asked to try again. After three tries, they were shown a correct solution and required to copy it. There are two learning activities here, copying an example and solving a problem, we must analyze them separately.

When copying the example (or the solution to a problem after three incorrect answers), students have the choice of self-explaining each line as they copy it or not. Moreover, if they do try to self-explain a line, they may or may not succeed. Thus, there are three options:

Start
1. Do not try to self-explain the line → Exit, with no learning
2. Try to self-explain how this line was derived from the preceding one
   2.1. Find a correct explanation → Exit, having learned
   2.2. Fail to find a correct explanation → Exit, having some learning

When students are using the keyboard to copy a line, they take longer and report more subjective difficulty than when they use handwriting to copy the line. The extra time and perceived effort of keyboarding may cause them to more frequently choose to not try to self-explain the line. This is one explanation for why the learning gains were higher with handwriting than keyboarding.

Next we discuss the second activity, wherein students solved an equation with a similar example present. This activity affords a choice at each line of:

Start
1. Try to copy the corresponding line in the example
   1.1. Some numbers in the example line are not in the example's first line → Start
   1.2. Otherwise, map the first line of the example to the first line of the problem and apply that substitution to the example line while copying it → Exit, with little learning
2. Try to apply the target knowledge
   2.1. Write a correct equation → Exit, having learned
   2.2. Write an incorrect equation → Exit, having mislearned
   2.3. Fail → Start
3. Refer to the corresponding line in the example, try to self-explain it and try to apply the inferred transformation to the problem's line
   3.1. Write a correct equation → Exit, having learned
   3.2. Write an incorrect equation → Exit, having mislearned
   3.3. Fail → Start

Because handwriting is easier than keyboarding, students might be more likely to choose apply(2) or self-explain(3) than copy(1). Thus, making good paths easier may be causing students to choose them more frequently.

It might seem that the error-based explanation used with the Aleven and Butcher geometry contiguity study should also apply here, assuming that typos are more common than handwriting errors. However, unlike the Aleven and Butcher study, where errors were increased only along the deep learning path, the errors here would probably be increased along both the shallow learning and deep learning paths. Moreover, the feedback on errors doesn't occur until an answer is entered. When an incorrect answer is entered, the students must try again. On this second attempt, they already have a list of equations, so they are more likely to refer to the example in order to check their work (this is a new, 4th path, similar to copying (1) in the graph above). Such checking would allow them to find the error, but it would not increase their learning. However, it might convince them to use the copying path(1) on the next problem. Thus, errors do not appear to have a direct effect on learning as they did in the Aleven and Butcher study; their effect appears to be indirect.

This analysis in the preceding paragraph is complex enough that it would probably become clearer if it were redone formally. That is, the learning event space would be treated as a stochastic finite state network by equipping each choice point with a probability distribution across the choices. The effects of errors could be shown not to affect the learning outcomes unless the errors also shifted the probability distributions.

The bottom line is that the Aleven and Butcher study can be explained by either errors or estimated effort, whereas this study can only be explained by estimated effort. Thus, if the two studies have different results, we may know why. If we tried to use "cognitive load" to explain them both, differential results would require us to label one study's cognitive load "germane" and the other study's cognitive load "extrinsic." (Sweller, ref).

At this writing, a preliminary lab study has been completed. Students in the handwriting condition had a mean gain score of 16.7% versus 13.4% for students in the keyboarding condition. Although this trend is in the expected direction, it was not statistically reliable, perhaps due to the short duration of the study (1 hour) and incomplete assessment of students' prior competence. The research plan calls for repeating this study *in vivo* so that student incoming competence can be factored out and the training time can be longer.

4.4 Note-taking technologies (Bauer & Koedinger)

These studies compare the learning gains achieved by different technologies for note taking. The hypothesis is that different technologies make different paths through the learning event space easier or harder; this affects students' path choices, which in turn affects their learning.

Students are given a text to study and one of the following methods for taking notes on the text:

- No-paste. Students use a text editor that prohibits copy-paste.
- Paste. Student can either copy-and-paste or type notes.
- Restricted paste. Like paste, but the tool permits only key ideas to be selected, not arbitrary pieces of text.
- Selection. When students click on text with a key idea, they see a menu of possible notes; they select the one they want. Only one is correct. If an incorrect one is selected, students are asked to try again.

Study 1 used just the first two conditions. It found that learning gains were larger in the no-paste condition than the paste condition. When a student has just read a paragraph and wants to take a note on it, the student has the following choices

Start

1. Think hard about the paragraph, perhaps rereading it, decide on its key ideas
   1.1. Find a passage in the paragraph with the key ideas & copy it (or highlight it; this is option is available in study 3) → Exit, having learned
   1.2. Type the key ideas in → Exit, having learned
   1.3. Type an incorrect representation of the idea → Exit, having mislearned
2. Think little about the paragraph and decide to note some plausible summary such as the lead sentence or the final sentence
   2.1. Copy it → Exit, having learned little
   2.2. Type it in → Exit, having learned little
3. Think little about the paragraph, and decide to note the entire paragraph.
   3.1. Copy it → Exit, having learned little
   3.2. Type it in → Exit, having learned little

In the typing condition, only options 1.2, 1.3, 2.2, and 3.2 are available. Options 2.2 and 3.2 often require more keystrokes than 1.2, because the copied passages are often longer than the key ideas. In fact, 3.2 is quite unlikely (as observed in previous studies), as it can involve an inordinate amount of work. Thus, in the typing condition, students may actually save time by finding the key ideas (1) instead of typing a summary (2) or entire paragraph (3). Thus, the estimated effort for the key ideas path in the typing condition may be lower than the estimated effort for the paths that involve less learning. When pasting is provide, the estimated efforts flip, so that copying summary sentences (2) or entire paragraphs (3) becomes faster and easier than locating and typing in key ideas (1). Thus, the estimated effort explanation used with the two preceding project also applies here and explains why the typing (no-paste) condition does better than the paste condition.

The second study will include the restricted paste and the selection conditions. These offer new options to the student:

Start

1. Think deeply about the paragraph, perhaps rereading it, and decide on its key ideas
   1.1. Cannot find a passage in the paragraph with the key ideas → Start
   1.2. Restricted paste tool: Find a passage in the paragraph with the key ideas & try to copy it
      1.2.1. Tool accept the selection → Exit, having learned
      1.2.2. Tool reject the selection → Start
   1.3. Selection tool: Find a passage in the paragraph with the key ideas & select it to get a menu of possible notes
      1.3.1. Find and select the appropriate note→Exit, having learned
      1.3.2. Find and select an inappropriate note; receive “try again” feedback → Start

2. Think little about the paragraph and decide to note some plausible summary such as the lead sentence or the final sentence
   2.1. The restricted tool accepts the selection
      2.1.1. Self-explain the selection → Exit, having learned
      2.1.2. Do not self-explain the selection → Exit, having learned little
   2.2. The restricted tool rejects the selection → Start
   2.3. The selection tool accepts the selection
      2.3.1. Find and select the appropriate note→Exit, having learned
      2.3.2. Find and select an inappropriate note; receive "try again" feedback → Start
3. Think little about the paragraph, and decide to note the entire paragraph.
   3.1. Copy and paste multiple times, until you have the whole paragraph→Exit, having learned little
   3.2. Rapidly select each sentence within the paragraph to get the menu. For each idea:
      3.2.1. Find and select the appropriate note→Exit, having learned
      3.2.2. Find and select an inappropriate note; receive "try again" feedback → Start

In the restricted paste condition, there are two ways to learn (1.2.1 and 2.1.1). The second one seems rather implausible—if the students were not engaged enough to think deeply about the paragraph, then they are probably not inclined to wonder why the tools accepted the passage they selected. Thus, if the tool increases learning, it is probably by encouraging exiting via 1.2.1. The restricted paste condition does this by requiring students to pay attention to the material they are noting and increasing the effort of surface strategies. For example, noting an entire passage would require multiple copy-paste actions.

In the selection condition, there are three ways of learning (1.3.1, 2.3.1, 3.2.1). In actuality, every time the student takes a note this tool requires an additional task of discrimination between options, even where the student had surface level goals (2, 3). Students would not learn where they select an incorrect entry.

4.5 Knowledge component construction vs. recall (Booth, Siegler, Koedinger & Rittle-Johnson)

This project is attempting to tease apart two kinds of learning: knowledge component construction vs. knowledge component recall. It is primarily a study of path effects, but the later studies also attempts to manipulate path choices in order to increase the frequency of knowledge construction relative to recall.

Suppose that the algebra tutor, in the unit on combining like terms, asks the student what to do next when solving 2x+1 = 3-4x for x. The learning event space is:

Start

1. Apply a shallow strategy, such as guessing or copying from a hint
   1.1. Enter a correct response (e.g., add 4x to both sides) → Exit, learning little

1.2. Enter an incorrect response → Start
2. Recall or partially recall the correct algebraic operation
   2.1. Enter a correct response → Exit, refining and strengthening the knowledge
   2.2. Enter an incorrect response → Start
3. Construct the correct algebraic operation from more primitive knowledge components
   3.1. Enter a correct response → Exit, having constructed a new KC
   3.2. Enter an incorrect response → Start
4. Ask for a hint → Start

The primitive knowledge components used in construction(3) could be

A. Being able to identify the variable terms and the constant terms
B. Understanding that it is possible to combine the variable terms together and the constant terms together, but not the variable terms with the constant terms
C. Understanding that 3 – 4x is the same as 3 + (-4x) and that the negative is part of the modifier of x on the right side of the equation
D. Understanding that in order to remove a negative term from a side you must add it to both sides (or to remove a positive term you must subtract it from both sides)
E. Being able to carry out the addition (or subtraction) procedure effectively

The knowledge component constructed during construction(3) may generate a correct response in this case, but may nonetheless be buggy. For example, while solving the equation 5 + 4x = 3 + 6x, a student may construct a knowledge component that says that the first step should be to subtract from both sides the variable term that comes after the operator on the left hand side; in this problem, 4x. However, when it is recalled and used later, it may generate an incorrect response (2.2), such as for the problem 2 – 7x = 4 + 3x, where the application of the new knowledge component would lead the student to incorrectly subtract 7x from both sides. In general, a student may have multiple applicable knowledge components available for recall(2), some of which may be buggy. The tutor's feedback probably changes the relative frequency of recalled knowledge components; this may be described by memory theories (e.g., ACTR).

The first study inserted assessments of small knowledge components, such as the ones listed above, before and after the normal training of the algebra tutor. Preliminary results from this study show that buggy knowledge components were associated with use of related buggy strategies on written pre- and posttests. The investigators will also attempt to fit the above model to the student's paths through the learning events as partially observed through log data.

The second and third studies will determine the impact on learning caused by inserting training on the small knowledge components into the normal training of the algebra tutor. This will be done using self-explanation exercises that require students to explain which knowledge components describe why a given strategy is a good or bad step to take in solving the problem. For instance, to train students on the importance of identifying like terms (for component C above) one exercise might show a start state of a problem (2x + 1= 3 – 4x) and a first step in the problem (3x = 3 – 4x) and ask the student to explain why the step was incorrect (e.g., it combined terms that weren't like terms: 2x and 1). This training should influence in predictable ways the relative distribution of paths through the learning event space above. For instance, when students are trained on knowledge components that they appear to be ignorant of (as determined by a pretest), then they will more likely take the construction(3) path because they notice that they tend to be successful when they do. Thus, this manipulation should affect students' path choices, and hence their learning.

4.5 Adding diagrams of acid-base solutions (Davenport, Klahr & Koedinger)

This project focuses on coordinative learning involving chemistry diagrams and equations. Although the hypothesis is that including diagrams will help learning, the main work is in finding out why. This involves studying both *path choices* (do diagrams encourage students to take successful deep learning paths more frequently?) and *path effects* (is deep learning involving a diagram+text more robust than deep learning involving just a text?).

In chemistry, a system (such as an aqueous solution) can be represented symbolically as a chemical equation or more concretely as a diagram that depicts individual molecules. Chemical equations (e.g., $H_2PO_4^- + H_2O \leftrightarrow H_3O^+ + HPO_4^{2-}$, $pK_a = 7.2$) have the advantage of representing the overall properties of the system; specifying which chemicals react with each other in which proportions. Experts can readily use information from chemical equations to solve problems and make predictions, such as how the pH will change when a strong acid is added to a solution.

Molecular diagrams that depict underlying microscopic processes can also represent chemical solutions and may allow students to visualize chemical reactions more concretely. These diagrams portray cartoons of molecules as color-coded dots and make salient information that may be lost in an equation. For instance, a chemical's concentration can be seen as the dots per square inch.

In the first of several studies of the impact on diagrams on chemistry understanding, Davenport et al. are comparing instruction on creating buffering solutions with and without diagrams. Both forms of instruction involve a derivation of a generic equation for solving buffering problems, and an example of its application to a specific problem. The control instruction is entirely text, including mathematical and chemical equations. The experimental instruction includes a series of diagrams that illustrate the chemical processes and the derivation of the solution equations. After instruction, students practice solving buffering problems with the aid of a tutoring system. They can refer to the instructions as they do so. When problem solving is completed, they take a closed-book post-test that includes questions to test conceptual understanding and transfer of knowledge.

Learning can occur either while studying the instruction or during the practice sessions, so let us consider the learning event spaces separately. During instruction, let us assume that a learning event corresponds to the student reading a line of text which may or may not have a diagram associated with it. The space for such a learning event is:

Start
1. Read the material without trying to understand it deeply → Exit, with little learning
2. Try to self-explain the line of text
   2.1. Succeed → Exit, with learning
   2.2. Fail → Start
3. If there is a diagram present, try to self-explain the diagram and text together
   3.1. Succeed → Exit, with learning
   3.2. Fail → Start

The control subjects will never be able to take path 3, because their instruction has no diagrams. Thus, the benefits of the experimental manipulation, assuming they exist, could be due to the popularity of path 3 and/or the higher probability of successful self-explanation using it (i.e., $P(3.1 \mid 3) > P(2.1 \mid 2)$). Thus, the manipulation may have part of its effect via path choice: Diagrams are better than text at enticing students to self-explain, and they may more frequently allow a successful self-explanation. A second explanation involves path effects: A successful self-explanation of a diagram+text may create a robust learning than a successful self-explanation of the text alone.

During problem solving, let us assume that students are asked a series of questions via menus, type-in boxes or other gestures (e.g., pouring a chemical into a simulated beaker). They get immediate feedback and hints when they make an incorrect entry. At each such prompt, they have an opportunity to exercise a small amount of knowledge, so let us assume each one is a learning event. Viewed abstractly, the learning event spaces are:

Start
1. Apply a shallow strategy, such as guessing or copying from a hint
   1.1. Correct response → Exit, with little learning
   1.2. Incorrect response and feedback from the tutor → Start
2. Retrieve from memory some knowledge and apply it
   2.1. Correct response → Exit, with learning
   2.2. Incorrect response & feedback from the tutor → Start
3. Refer to the instructional materials and apply them shallowly, e.g., by copying the equations syntactically
   3.1. Correct response → Exit, with little learning
   3.2. Incorrect response → Start
4. Refer to the instructional materials, self-explain them if necessary, and apply them deeply
   4.1. correct response → Exit, with learning
   4.2. Incorrect response → Start

Even if students have the same knowledge coming into the problem solving session, the addition of diagrams to the instruction may change the paths they choose. The diagrams may make path 4 more common than path 3, for instance. This is another possible explanation for why diagrams may help students learning. It is based on path choices during problem solving.

The first version of this study was run at the University of British Columbia in late March. 75 students agreed to participate and completed the study. All data were passed to the PSLC DataShop and is currently in the process of being analyzed.

The second version of this study is currently underway in a chemistry course at Carnegie Mellon University. Approximately 150 students are anticipated to participate and all data will be collected in the PSLC DataShop.

4.6 Co-training of Chinese characters (Liu, Perfetti, Mitchell & Wang)

In order to determine if an important phenomenon in machine learning is found in human learning, Liu, Perfetti and Wang investigated a novel method for teaching students the English meanings of Chinese characters. When students are learning to translate into English the written and spoken forms of a Chinese character, the standard training procedure is to give them feedback on their translation. According to co-training results in machine learning (Blum & Mitchell, 1998), having students produce translations *without feedback* should also help their learning, given that the spoken and written Chinese characters are both presented. In machine learning terminology, the English translations are the "labels" so the standard instruction is viewed as learning from labeled examples, where each presentation consist of either the spoken form + label, the written form + label or both the spoken and written forms + label, where this last presentation is considered an example pair. On this view, the experimental instruction is viewed as learning from *un*labeled example *pairs*.

The key idea is that when observing the unlabeled example pairs, students can use the more familiar element of the pair (e.g., the written character) to generate a possible English translation that is used as "feedback" for learning the translation of the less familiar element of the pair (e.g., the spoken form). According to the Blum and Mitchell cotraining results, the benefits of training with unlabelled pairs should be strongest when one element of the pair is much more familiar then the other element of the pair.

Here is the learning event space when both forms are presented to the student without feedback:

Start
1. Don't even try to retrieve the meaning → Exit, with no learning
2. Try to retrieve meaning
    - 2.1. Spoken form provides confident guess
        - 2.1.1. Attend to written form, and strengthen association of written form to meaning and to spoken form → Exit, with learning
        - 2.1.2. Don't attend to written form, so strengthen only the association of spoken form to meaning → Exit, with some learning
    - 2.2. Written form provides confident guess

    2.2.1. Attend to spoken form, and strengthen association of spoken form to mean and to written form → Exit, with learning
    2.2.2. Don't attend to spoken form, so strength only the association of written form to meaning → Exit, with some learning
  2.3. Neither spoken nor written form provides a confident guess, so associate only the association between them → Exit, with little learning

If at least one of the forms is already learned, which is likely given the prior training with labeled examples, then path 1 and 2.3 should be rare. Thus, as long as students engage in the task, then they should mostly exit via 2.1.1, 2.1.2, 2.2.1 or 2.2.2, and thus learn. This would illustrate our claim (and Newell's) that learning at this level of description is often determined by the information itself rather than the information processing architecture (human vs. machine).

Moreover, if the written form is more familiar than the spoken form or vise versa, this may possibly increase the chances for path 2.1.2 and 2.2.2 to be taken, and thus increase the likelihood of triangular association. That is, manipulating the correlations among elements of the unlabelled example pairs should impact student's path choices, and hence their learning.

In all conditions of the experiment, students were pre-tested, trained on labeled examples, perhaps given additional training, and post-tested. In testing, students translated either the written form or the spoken form when it was presented in isolation. The conditions of the experiment were:

- *Unlabeled, uncorrelated pair*: Students are shown an unlabeled example pair, where one element of the pair (written or spoken form) was more familiar (as measured on the pretest) than the other element of the pair.
- *Unlabeled, correlated pair*: Students were trained with unlabeled example pairs, where the elements of the pairs were equally familiar.
- *Unlabeled single element*: Students were trained with unlabeled non-pairs, that is, a written form or a spoken form, but not both.
- *Control*: No training.

The prediction based on co-training theory was supported by preliminary results from the study:

- *The value of "unlabelled" learning trials.* In learning the meanings of Chinese characters and spoken syllables, learning was facilitated by the addition of trials that did not provide the meaning (unlabelled trials) that followed trials in which meanings had been provided (labeled trials). However, this effect was restricted to unlabeled trials in which cross-modal pairs (spoken syllable and character) were presented; it was absent when only one (spoken) or the other (written) modality was presented.
- *The value of written forms to support the learning of meaning.* The study found a very large advantage for the presentation of written characters compared with their corresponding spoken syllables for learning a form-meaning pair.

- *Benefits of uncorrelated examples.* This is still being assessed.

4.7 Co-training in the self-correction of speech errors (McCormick, O'Neill & Siskin)

This project can also be viewed as testing the Blum and Mitchell co-training results, but in a much more applied setting than the Liu, Perfetti and Wang study. Indeed, the control and experimental conditions differ enough that it is difficult to analyze the study in terms of learning event spaces, and a complete analysis has not yet been accomplished.

In order to correct errors in spoken second language, a common activity is to have students record 2 or 3 minutes of speech on a given topic, have the teacher record oral feedback on the student's speech, and have the student study the teacher's feedback. There are several known problems with this approach. First, students may not process the teacher's feedback and attempt to learn from it. Second, students depend upon the teacher for feedback and do not practice self-correction, which is a meta-skill that they can use after leaving the course.

McCormick et al. have developed a new way to involve students in the correction of spoken language. Students (1) record a 2 or 3 minute speech, (2) transcribe their spoken language verbatim, (3) note their errors and indicate corrections, and finally (4) submit everything to the teacher. The teacher not only checks the student's error analyses, but records an oral analysis of errors that the student overlooked. Students are required to complete a digital report after listening to the teacher's feedback. Essentially, this task differs from the conventional one in three ways:

- Students transcribe their speech.
- Students attempt to find and correct their errors.
- Students are required to complete a digital report after listening to the teacher's feedback.

Let us analyze the potential impact of the first two activities, transcription and self-correction.

The students have multiple opportunities to reflect on their language output. First, the students can reflect while they are producing the spoken language output. This opportunity does not involve direct manipulation of path choices. Second, the students repeatedly listen to short segments of their spoken language samples, pause, and transcribe the speech they heard during which they have the opportunity to reflect on its correctness. Third, students can listen to their samples and view their transcription in the section of the program in which they take notes on their errors and corrections. During this process they again have the opportunity to reflect on correctness. When they reflect, they have the following options:

Start

1. Process the phrase shallowly and consider it correct → Exit, with little learning
2. Process the phrase deeply, considering both its orthographic and/or audio forms
   2.1. Consider it correct
      2.1.1. The phrases is correct → Exit, with some learning

    2.1.2. The phrase is incorrect → Exit, with some mislearning
  2.2. Consider it incorrect
    2.2.1. Properly understand the error and execute a proper correction → Exit, with learning
    2.2.2. Either there is no error, the error is misunderstood or the error is miscorrected → Exit, with perhaps some mislearning

In order to make this activity effective, students should most frequently exit via 2.1.1 or 2.2.1. On the one hand, the teacher and the software have little control over the errors and phrases that the student will encounter, so they cannot adjust the difficulty of the self-correction activity to keep it within the students' capabilities. However, students transcribe their speech, and they may be more capable of detecting errors in the written version than the spoken version. Thus, they are taking advantage of co-training, just as Lui, Perfetti and Wang did in one of their studies. Although this is not the sole factor causing benefits for this approach over the conventional one, and it may not even be a particularly large factor, it is an important one for the PSLC theoretical framework, as it shows how machine learning can help us understand human learning.

The new self-correction activity is so different from the old activity, wherein teachers corrected the students' errors, that their learning event spaces hardly overlap. Thus, it will take more space than this forum affords to carefully analyze the differences between the old activity and new one.

Because this in-vivo project is still underway, present findings at this time are not available. However, we would like to mention a few anecdotal observations. First, all levels have been able to identify and self-correct errors to some degree. Second, students are able to identify and self-correct errors in the areas of grammar, vocabulary, and pronunciation, with grammar the area with the most frequent corrections. Third, the issue of what constitutes "correction" is complex. Data show that students do the following:

- correct already-correct language forms (as in 2.2.2),
- incorrectly correct incorrect forms (2.2.2.),
- partly and correctly correct incorrect forms (2.2.2.), or
- completely and correctly correct incorrect forms (2.2.1).

The data from the correction component of the project may provide insight into robust language learning. For example, correcting an already-correct language form may be an indication that the learner has not robustly learned that form, even though it was produced correctly. Therefore, it is important to view correct output as potentially underestimating what students know.

Although our project is currently positioned in the co-training cluster, we believe that the process of the RSA connects to all four clusters: co-training, dialogue, refinement, and fluency. Because the student identification and self-correction process occurs without concurrent input from the teacher, students are engaged in *co-training* – they are going beyond classroom instruction and practice to work on their own. Overall, the work of catching errors is shared by multiple agents – the student during and after speech and the instructors when providing feedback on the speech – in what could be viewed as asynchronous *dialogue*. The action of noticing errors and correcting errors may *refine* learner interlanguage. Finally, over time and in conjunction with their other language instruction, practice, and experiences, this refinement of interlanguage may facilitate *fluency* in that forms that the learners have consistently noticed and corrected may be accessed faster.

4.8 Personalization and example studying in chemistry (McLaren, Koedinger & Yaron)

This project is studying a pair of manipulations, both intended to influence path choice and hence learning. The study uses a 2x2 design, that is, it includes all combinations of two changes to the instruction used in the chemistry LearnLab course.[2] One change, motivated by Clark and Mayer's (2003, p. 133-138) personalization principle, was to reword the tutor's messages to the student to make them more personal and informal. For instance, instead of saying "This type of problem should be solved as follows…" the tutor would say, "Here's how I think you ought to solve this kind of problem…" The second change, motivated by Trafton and Reiser's (1993, p. 1022) paradigm of alternating examples with problems, was to replace every other problem-solving task with an example-studying task. That is, for half the tasks, instead of generating a solution to a problem, students would study the solution to the problem instead. The problems were stoichiometry problems, so each solution involved the creation of an equation, where each equation is composed of multiple terms, which in turn are composed of multiple elements. For example, the following equation is the solution to one problem.

(0.58 mol AsO2- / 100 kL solution) * (1 kL solution / 1000 L solution) * (106.9 g AsO2- / 1 mol AsO2-)
= 0.00062 g AsO2- / 1 L solution

As they build the equation and solve the problem, the individual elements of the equation provided by the student are the numbers (e.g., 0.58, 100, etc.), units (e.g., mol, kL), and substances (e.g., AsO2-, solution) in each term. Thus, a learning event for the stoichiometry tutors is the creation of individual equation elements.

The learning event spaces of the example-studying and problem-solving task have no paths in common, even though they are similar in many ways.

The learning event space for studying one element of an example is:

Start
1. Process the element shallowly, e.g, by paraphrasing it or just reading it → Exit, having learned little
2. Self-explain the element → Exit, having learned

[2] Actually, to be precise, we did not change existing instruction in the chemistry course but, rather, developed our own, new tutored problems. The plan is to add these tutored problems to the chemistry LearnLab course, but they are not part of the course now.

The learning event space for generating a solution line during problem solving with the tutor's help is:

Start
4. Students apply a shallow strategy, such as guessing or copying from a hint
    4.1. They enter a correct answer → Exit having not learned much.
    4.2. They enter an incorrect answer → Start
5. Students apply the target knowledge.
    5.1. They enter a correct answer → Exit having learned something.
    5.2. They enter an incorrect answer → Start
6. They ask for & get a hint from the tutor → Start

The personalization manipulation does not affect the learning event spaces, but could potentially affect the student's choice of paths. However, preliminary results suggest that this manipulation did not make a difference in learning gains. While there are several potential reasons why no difference was found, one possibility is that results found in laboratory studies (e.g., Moreno and Mayer, 2000) do not transfer to the "wild" i.e., *in vivo* classwork.

The manipulation wherein half the problems were replaced with examples could affect learning if it changes the relative frequency of good and bad paths through the two learning spaces. Preliminary results suggest that students who both solved problems and studied examples learned more than those who only solved problems – but not at a significant level. This suggests that when studying examples, students more often decided to self-explain (they chose path 2), whereas when students were solving the equivalent problems, they more frequently exited via 3.1 than 4.1. It is not clear yet why the path choices came out the way they did, but because the difference was not significant further studies are required.

4.9 Understanding culture from film (Ogan, Aleven & Jones)

This study explores methods for teaching and assessing students' understanding of French culture in the context of the French LearnLab course. Its main manipulation is to insert some questions at just the right place while students are watching a film. This is intended to modify their path choices and increase the frequency of deep learning.

The model used for the development of intercultural competence was proposed by Lo Bianco, Liddicoat and Crozet (1999). It involves an interactive cycle that begins with authentic source material and, with guidance in noticing unique cultural elements, leads to self-reflection and subsequent verbal output that is evaluated by peers and instructors. Training is done with short (1 to 2 minute) film clips as source material that have been chosen by instructors to highlight aspects of French culture that they would like students to learn. Two types of knowledge components are assessed: primarily, the acquisition of cultural elements such as attitudes and values with analytical post-test questions, and secondarily, cultural perspective-taking skills done by evaluating the online, typed discussions after the training. The transcripts are coded, blind to condition, in order to evaluate the cultural perspective-taking skills of the students.

In the control condition of this study, students read a summary of the film's plot up to the point of the film clip, view the film clip and discuss it online afterwards. The experimental condition is the same, except that the software stops the film just before a key moment in which a cultural event occurs, and asks students about what they think will happen next and what they have seen so far. The first question is presented in a menu format to suggest several appropriate responses and lightly constrain students to actual cultural possibilities. In the second question, students are given space to explore their hypothesis and provide evidence from the clip or their knowledge of the French culture to support their reasoning. The experimental condition's software also asks questions after the clip is completed. For instance, questions might ask about how the characters were dressed, their attitudes, and how they interacted with members of other social groups in the film. These questions have students select their response from a menu so that the software can give feedback.

For analysis, each of the tutor questions can be treated as a learning event. If students reflect deeply enough on the question and film, then they should be able to acquire a component of culture knowledge, but they may choose not to do so:

Start

1. Use a shallow strategy, such as guessing or typing "I don't know"
    1.1. Software provides negative feedback → Start
    1.2. Software accepts the response → Exit, with little learning
2. Reflect on the film and the question, trying to learn
    2.1. Software provides negative feedback → Start
    2.2. Software accepts the response → Exit, with learning

These questions assist students with the second and third steps of the Liddicoat model, drawing the students' attention to aspects of the film that they could, in principle, have noticed without the questions' help, then asking them to self-reflect. Thus, for each question, there is at least one episode in the film clip where the appropriate information is observable. These episodes comprise the learning events for the control condition. Note that although the episodes are in the film, they are not highlighted by the software in the control condition. Thus, their learning event space is quite simple:

Start

3. Notice the cultural implications of the episode → Exit, with learning
4. Do not notice the cultural implications → Exit, with little learning

If, as initial results show, the addition of questions to the film clips does increase learning, then it could be that students in the experimental condition have a higher good path frequency than the control students. The good path frequency is the number of students exiting via a good path (i.e., (i.e., 2.2 or 3) divided by the number of students exiting.

Unfortunately for the simplicity of our analysis, these are not the only times when learning can occur. The discussion after the film clip is not only an assessment but also an opportunity to learn, and answering a question may cause students to notice and learn different things as they watch the film after answering the question. This type of study, which allows the set of learning events to vary across individuals, is difficult to completely analyze in terms of learning event spaces. This illustrates why the theoretical framework needs both a macro level of description and a micro level—a study like this can be complex and hard to understand at the micro level and yet be simple and easily understood at the macro level.

4.10 Does learning from examples improve tutored problem solving? (Renkl, Aleven & Salden)

This project is mostly a study of the changes in path choices that occur when a tutoring system includes partially worked examples. The basic idea is that when a tutor relieves a student of some of the work in generating a line by providing part of it, then students are more likely to engage in deep learning to fill in the rest. However, the instruction must be engineered so that students still become autonomous problem solvers—they eventually can do all the work themselves.

In earlier work comparing mixtures of example studying to problem solving, it was found that "fading" the example studying into the problem solving is highly effective (e.g., Atkinson, Renkl, Merrill, 2003; Renkl & Atkinson, 2003). To illustrate the fading procedure, suppose all the problems can be solved in exactly four lines. Students would first see a worked example, where all four lines leading up the answer were provided. Next, a new problem is presented in which the first three lines are displayed, and the last line has been "faded out" so the student must generate it and the answer. The third problem has only the first 2 lines provided; the student must generate the last 2 lines and the answer. Eventually, the student is solving the whole problem and providing all 4 lines. This "fading of examples" training has been shown in several experiments to be more effective than other, more conventional mixtures, such as alternating between example studying and problem solving.

All the earlier work was conducted with *untutored* problem solving—if the student made a mistake, the software gave no feedback. In six studies, 3 taking place in Freiburg and 3 in Pittsburgh, we will determine if the advantages of fading occur when the problem solving is tutored. The studies will take place in Circles unit of the Geometry LearnLab course.

One analysis treats each step of the solution of a geometry problem as a learning event. A step has 2 parts: Given a geometric component, such as an angle, type in its value and the justification (reason) for the component having that value. For instance, given $\angle$ABD, enter the value (35º) and the justification (e.g., "Exterior angles applied to line DBC"). Solving a problem may require one or more such steps. A glossary of possible justifications and other forms of help are available. One learning event space for a single step is:

Start

1. The tutor provides the value and prompts the student for a justification.
    1.1. The student generates a correct justification via a deep strategy such as self-explanation → Exit, with learning
    1.2. The student generates a correct justification via a shallow strategies such as guessing → Exit, no learning
    1.3. The student's justification is incorrect and the tutor gives feedback → Start
2. The student generates the value via a shallow strategy such as guessing or copying it from a hint
    2.1. The value is incorrect and the tutor gives feedback → Start
    2.2. The value is correct; the tutor says so and prompts the student for a justification.
        2.2.1. The student generates a correct justification via a deep strategy such as self-explanation → Exit, with some learning
        2.2.2. The student generates a correct justification via a shallow strategies such as guessing → Exit, no learning
        2.2.3. The student's justification is incorrect and the tutor gives feedback → Start
3. The student generates the value by trying to apply geometry knowledge
    3.1. The value is incorrect and the tutor gives feedback → Start
    3.2. The value is correct; the tutor says so and prompts the student to self-explain it.
        3.2.1. The student generates a correct justification via a deep strategy such as self-explanation → Exit, with learning
        3.2.2. The student generates a correct justification via a shallow strategies such as guessing → Exit, no learning
        3.2.3. The student's justification is incorrect and the tutor gives feedback → Start
4. The student asks for and receives a hint → Start

Experiment 1 has two conditions. In the problem-solving condition, students must enter the value and justification for every step; thus they can never take path 1 and always exit via path 2 or path 3. In the example condition, the tutor sometimes provides the value, while the student always provides the justification.   In this condition, the first multi-step problem in each unit has all the values provided by the tutor, and subsequent problems have fewer and fewer values provided by the tutor. When the tutor provides the value for a step, then the students exit via path 1. When the tutor does not provide the value, then the students exit via path 2 or path 3.

The "best" exit is via 3.2.1, because that means that the student has reasoned out both the step's value and its justification. Exiting via 1.1 or 2.2.1 means that the student has thought out the justification of the step but not the value, so those paths may elicit some learning. The other exits probably involve even less learning.

Preliminary analyses of the first experiment showed that there is an aptitude-treatment interaction. Besides the not very surprising finding that the learners with high prior knowledge outperformed learners with low prior knowledge irrespective of the learning condition (example versus problem-solving), learners with low prior knowledge profited more from the problem-solving condition whereas learner with high prior knowledge came off better in the example condition. This pattern of results is exactly the opposite of what would be predicted by previous research findings on example-based learning (cf. Renkl & Atkinson, 2003).

Based on the data analyses already available and the rationale of learning paths, a preliminary explanation for this unexpected finding is the following: *Learners with low prior knowledge* had difficulties particularly in providing correct justifications, whereas determining the correct value of a step – in part with the help of hints (via path 4) – was easier for them. Thus, in the example condition, paths 1.2 and 3.2.2 prevailed (shallow processing of the justification), whereas in the problem condition a successful determination of the value of a step was more frequent (path 3.2.2). Because path 3.2.2 involves some learning whereas path 1.2 involves no learning, and students in the problem solving condition could not take path 1.2, they learned more than students in the example condition.

Now we consider *learners with high prior knowledge*. In the example condition, there were many productive self-explanation episodes (path 1.1) that prepared them well for later faded steps so that they could directly (i.e., without the support of hints, path 4) apply geometry knowledge to generate both a value and a justification (path 3.2.1). In the problem solving condition, this "preparation" was missing, leading to suboptimal exits (problem solving with partial understanding, such as paths 2.2.2 or 3.2.2). Further analysis of the on-line data collected during the learning phase will show whether this preliminary explanation holds.

5 Interactive communication cluster

The Interactive Communication cluster contains projects and studies that focus on the impact on learning of a student communicating interactively with another agent, such as a tutor or a peer. Although many of the studies involve natural language dialogues, the interactive communication could be in other media was well as. However, the more features the communication shares with natural language dialogues, the better it fits in this cluster. Key features include sharing initiative and control over the interaction, exploring an idea at arbitrary depth and weak constraints on content.

Like the other sense-making cluster, the Interactive Communication cluster tends to focus on path choices, as opposed to path effects. It studies the impact on learning of methods for increasing the frequency of students' sense-making activities. Perhaps the major method studied in the cluster is varying the degree of interactivity of the communication. The hypothesis is that low (or no-) interaction instruction, such as reading a text, will be marred by a low frequency of sense-making. For instance, many students would read the text shallowly rather than self-explaining it deeply. On the other hand, high-interaction instruction, such as a natural language dialogue, should increase the frequency of sense-making paths in the learning event space. In particular, eliciting a line of reasoning from a student via a series of questions, as opposed to telling them the line of reasoning via a series of statements, should decrease the frequency of shallow processing. Two projects (Katz; Hausmann & VanLehn) are exploring the impact of interactivity on learning.

Another method for changing path frequencies is to manipulate the students' role in the interaction. For instance, McLaren, Rummel & Spada are comparing the learning of students who take turns acting as tutor and tutee, vs. students who are coached by a computer tutor. Hausmann and Chi are comparing peers who were taught to co-construct knowledge while adopting either an elaborative role or a critical role with respect to their partner's proposals.

A third method for manipulation of path choices is to first briefly instruct students on the desired form of sense-making. This occurs before the domain-specific training. During training, students are reminded or prompted to use the deep learning strategy. This method is a component of the studies by Craig & Chi and by Hausmann & Vanlehn.

One of the key features of interactive communication is that the communication itself only weakly constrains the content, and this feature can make analysis difficult. In particular, when the conversation can be about any topic, as when two human students solve problems together (Hausmann & Chi; McLaren, Rummel & Spada), then the learning event analysis may not be very useful, because the learning events across dialogues are seldom similar enough that they can be merged into a learning event space. For this reason, many of the studies have used forms of communication that are not as open-ended as unconstrained speech.

5.1 *Deep, rhetorical questions during example studying (Craig & Chi)*

When students study a line in a problem's solution, they generally have the following options:

Start

1. Study the line shallowly (e.g., read & paraphrase it) → Exit, with little learning
2. Self-explain the line → Exit, with learning

Several earlier studies have shown that learning increases when the software displays each line separately and gives a content-free prompt, such as "Please explain how this line was derived." This increases the frequency of self-explanation (2) compared to shallow studying (1). This explanation has been confirmed by taking verbal protocols during example studying (Chi, 2000). It demonstrates that a relatively simple intervention can cause a significant change in path frequencies and hence learning.

Working in the physics LearnLab, the Craig and Chi are trying to further increase the amount of sense-making by replacing the content-free prompt with a deep-content rhetorical question, as suggested by Gholson & Craig (in press). That is, deep-content rhetorical questions are expected to increase student's preference for deep learning paths, and hence increase their learning.

Learners watch a series of short videos, each showing a step in solving a problem and each ending in either a self-explanation prompt or a deep, rhetorical question. The questions are rhetorical in that students are not given a means for answering. Moreover, it is hypothesized that the increase in the amount of sense making occurs even when only *some* of the example lines are followed by deep, rhetorical questions.

After students studied examples with and without deep rhetorical questions, they solved problems on Andes, the tutoring system used in the Physics LearnLab course. Although scoring of the Andes problems has not been completed, preliminary analyses of log data from the study shows a trend (p =.08) where the amount of time required to complete an Andes problem on the topic was decreased for learners who watched a video with deep rhetorical questions (M = 338.25s) below that of learners that watching a video with the identical content (M = 716s). Thus, rhetorical questions may be associated with faster subsequent problem solving, but we do not yet know whether it is associated with increased accuracy.

5.4 Does it matter who generates the explanation? (Hausmann & VanLehn)

Hausmann and VanLehn are comparing the learning that results from either being exposed to an explanation or generating it oneself. The main focus is on the effects of the two paths. In particular, if students follow a deep learning path that involves generating an explanation, will their learning be more robust than if they follow a deep learning path that involves comprehending an explanation that they have heard?

Students study a physics example line by line, which is presented as a video of the experimenter solving an isomorphic problem in Andes. For half the students, the example lines are incomplete (the explanations connecting some lines are missing) whereas for the other half, the example lines are complete. Crossed with this is an attempted direct manipulation of the student's path choice. Half the students were instructed and prompted to self-explain the line, while the other half were prompted to paraphrase it. Thus, here is the learning event space:

Start

1. Process the line shallowly, e.g., paraphrasing it
    - 1.1. There is nothing more to learn → Exit, with learning
    - 1.2. The line is incomplete; its explanation is missing → Exit, with little learning
2. Try to process the line deeply, e.g., self-explain it
    - 2.1. There is nothing missing from the line → Exit, with learning
    - 2.2. The line is incomplete; its explanation is missing
        - 2.2.1. The attempted self-explanation succeeds → Exit, with learning
        - 2.2.2. The attempted self-explanation fails → Exit, with perhaps less learning

The key issue is the robustness of the learning produced by path 1.1 compared to paths 2.1 and 2.2. That is, this experiment was initially designed to study path effects, as opposed to path choices. Hausmann and VanLehn hypothesize that learning by paraphrasing a complete example line (1.1) is just as robust as learning from self-explaining a complete line (2.1) or an incomplete line (2.2). Hypotheses similar to this have recently been proposed in the literature (Klahr & Nigam, 2004; Nokes & Ohlsson, 2005; Rittle-Johnson, 2006). An alternative hypothesis, related to the generation effect (Jacoby, 1978; Slamecka & Graf, 1978), is that the learning from paraphrasing a complete example (1.1) is less robust than the learning by self-explaining a complete (2.1) or incomplete (2.2) example. In the present experiment, verbal protocols were recorded as the students studied in order to determine whether the attempted manipulation of paths succeeded, and in particular, how often students successfully self-explained incomplete examples (2.2.1) versus unsuccessfully self-explained incomplete examples (2.2.2).

At this reporting, the data have been collected but the data analysis is not yet complete.

5.2 Self-explanation vs. interactive dialogue (Katz)

When students study a lengthy scientific explanation, they may or may not self-explain each step thoroughly. Perhaps one could increase the frequency of taking a deep learning path at each step by having the tutor ask a series of questions that "walks" the student through a multi-step explanation. That is, instead of telling the student each step, the tutor tries to elicit it from the student with a question. This can be viewed as an attempt to use increased interactivity to increase the frequency of deep learning paths through a learning event space.

Katz tried this in the physics LearnLab by giving students a problem to solve, asking them questions about it such as "If the initial velocity increases, does the block slide further?" then having them either read an explanation (hoping that they would self-explain it) or walk through the explanation's line of reasoning with the tutor (interactive dialogue). In a variant of the interactive dialogue condition, students were given the opportunity to ask the tutor questions, via follow-up questioning menus (mixed initiative dialogue). Although both forms of interactive dialogue and self-explanation invite students to follow the same line of reasoning, they are just different enough that it makes it clearer if they are analyzed separately.

When studying the solution in the self-explanation condition, at each line, students have these choices:

Start
1. Process the line shallowly → Exit, having learned little
2. Tries to self-explain the line
    2.1. Student successfully self-explains → Exit, with some learning
    2.2. Student fails → Exit, with little learning

On the other hand, when the tutor is coaching them down a line of reasoning, and the tutor has just asked a question, the choices are these:

Start

3. Student answers the question shallowly, e.g., by guessing
    3.1. Response is correct → Exit, having learned little
    3.2. Response is incorrect
        3.2.1. tutor starts a subdialogue → Start
        3.2.2. tutor states correct response
            3.2.2.1. Student successsfully self-explains the response → Exit, with some learning
            3.2.2.2. Student doesn't try or fails → Exit, with little learning
4. Student tries to answer the question from knowledge
    4.1. Response is correct → Exit, having learned
    4.2. Response is incorrect
        4.2.1. Tutor starts a subdialogue → Start
        4.2.2. Tutor states a correct response
            4.2.2.1. Student successfully self-explains → Exit with some learning
            4.2.2.2. Student doesn't try or fails → Exit, with little learning
5. Student asks a follow-up question to fill in a knowledge gap that prevents him from answering the tutor's question, and tutor responds → Start
6. Student tries to answer the question by analogy to the original problem's solution
    6.1. Response is correct → Exit, having learned
    6.2. Response is incorrect
        6.2.1. Tutor starts a subdialogue → Start
        6.2.2. Tutor states a correct response
            6.2.2.1. Student successfully self-explains → Exit with some learning
            6.2.2.2. Student doesn't try or fails → Exit, with little learning

The hypothesis is that the interactive dialogue is more effective than text because thoughtfully and successfully answering questions (4.1) is more frequent than successful self-explanation (2.1). That is, the hypothesis is that the manipulation affects path choices.

Unfortunately, student participation in all three forms of intervention was low. Among 70 treatment subjects (across the three post-problem questioning conditions) who completed at least one Andes problem, only 52 responded to one or more questions; only 14 treatment subjects responded to all 22 post-problem questions. Consequently, in analyzing post-test performance, Katz compared "yoked pairs" of students who responded to at least five questions with those who did not respond to any questions, across conditions. The former significantly outperformed the latter, with respect to pre-test to post-test gain scores on a test that emphasized conceptual understanding of physics (in particular, work and energy concepts). Regression analyses which treated the amount of participation in post-problem questioning as a continuous variable also revealed that the total number of post-problem questions that a student responded to had a significant positive effect on post-test score ($p = .03$), as opposed to the number of problems he or she completed ($p > .10$). However, analyses of final exam scores—for all problems and for the work and energy problems focused on in the study—did not reveal a significant effect for the number of questions that a student responded to. The number of work-energy problems that the student solved had a stronger positive impact on final exam performance than did the questioning intervention, but this may be because questions on the final exam were quantitative rather than qualitative. Hence, it appears that the effectiveness of the treatment was stronger for conceptual understanding ("sense-making"—as measured by the pre-test and post-test questions) than for quantitative problem-solving ability. Stated differently, transfer from conceptual understanding to problem-solving skill was not promoted by the intervention.

5.3 Conceptual vs. quantitative applications of knowledge (Katz)

In a new project, Katz is comparing two different physics tasks that are intended to apply the same knowledge components either quantitatively or conceptually. Students in the conceptual condition are given a task that has two parts: students first solve an ordinary problem, and then they answer qualitative questions about it. The questions usually pose a qualitative change in the problem and ask about the difference it would make on the solution, e.g., "If the initial velocity were larger, would the block slide further?" After posing the overall question and getting the student's initial answer, the tutor walks the student through a line of reasoning that answers the question by asking a series of questions. At each question, there are several choices:

Start

1. Process the question shallowly, e.g., by guessing
    - 1.1. Response is correct → Exit, having learned little
    - 1.2. Response is incorrect
        - 1.2.1. tutor starts a subdialogue → Start
        - 1.2.2. tutor states correct response
            - 1.2.2.1. Student successfully self-explains the response → Exit, with some learning
            - 1.2.2.2. Student doesn't try or fails → Exit, with little learning
2. Tries to answer the question from knowledge
    - 2.1. Response is correct → Exit, having learned
    - 2.2. Response is incorrect
        - 2.2.1. Tutor starts a subdialogue → Start
        - 2.2.2. Tutor states a correct response
            - 2.2.2.1. Student successfully self-explains → Exit with some learning
            - 2.2.2.2. Student doesn't try or fails → Exit, with little learning
3. Tries to answer the question by analogy to the original problem's solution
    - 3.1. Response is correct → Exit, having learned
    - 3.2. Response is incorrect
        - 3.2.1. Tutor starts a subdialogue → Start
        - 3.2.2. Tutor states a correct response
            - 3.2.2.1. Student successfully self-explains → Exit with some learning

3.2.2.2. Student doesn't try or fails → Exit, with little learning

In the quantitative condition, none of this occurs. Instead, students solve a few more quantitative problems than students in the qualitative questioning condition. (This is intended to control for time on task.) The learning event space for quantitative problem solving is similar to the ones above, but the analogy section (3) is absent and the knowledge used in trying to answer qualitative questions is an abstraction of the quantitative knowledge used in solving extra problems in the control condition. In particular, for the control task, a learning event space is:

Start

4. Ask for a hint
    4.1. Try to learn from the hint → Start, with some learning
    4.2. Ignore it → Start, with little learning
5. Guess an equation or copy it from a bottom out hint
    5.1. correct → Exit, with little learning
    5.2. incorrect & negative feedback from the tutor → Start
6. Try to generate an equation by applying knowledge
    6.1. Correct → Exit, with learning
    6.2. Incorrect & negative feedback from the tutor → Start

The conceptual and quantitative tasks are designed to employ the same knowledge components. For instance, in the quantitative condition, students would need to apply the definition of kinetic energy (which is a knowledge component) and write KE1 = 0.5*m*v1^2, whereas in the conceptual condition, students would need to use the same knowledge component to answer "If the initial velocity increases, what happens to the initial kinetic energy?" They could do that by recalling the formal definition, mentally noticing that making the v larger makes v^2 larger, which makes 0.5*m*v^2 larger, so the initial kinetic energy is larger. This is path 2. Alternatively, they can look at their solution to the original problem, find an equation that has KE0 and v1 in it, and answer the question by propagating an increase in v1 through the written equation. This is path 3.

Going along path 2.1 or 6.1 involves applying the target knowledge component and learning. However, the qualitative path 2 involves keeping the principle in memory while propagating the perturbation through it. This may be difficult to do reliably, so students may make more errors and/or take paths 1 or 3 instead. On the other hand, the retrieval features on path 2.1 are similar to those on the conceptual tasks on the post-test. Thus, the results turn on the relative frequency of paths 2.1 and 6.1 trading off with transfer from those learning events to the post-test.

5.4 Peer tutoring scripted collaboration (McLaren, Rummel & Spada)

McLaren, Rummel and Spada are studying the impact of scripted collaboration on learning. This section discusses one of their studies, which involves a script that has homogenous pairs of students taking turns playing the roles of tutor and tutee. Although the learning event spaces are complex and do not adequately capture all the differences between the conditions, the basic hypothesis is that the learning of both tutor and tutee increases because both tend to take deep learning paths more frequently than they would if they were in either an unstructured collaboration or working alone.

In the Peer Tutoring Script, which structures the collaboration in one condition of this study, there are three phases:

- *Preparation*: One student (the tutor) solves an algebra problem with the Cognitive Tutor Algebra (CTA) tutoring system providing feedback and hints. The student also prepares specifically for tutoring by identifying difficult skills and generating questions and explanations.
- *Collaboration*: The tutor acts as the feedback and hint provider while a second student, the tutee, solves the same problem using the CTA interface. Students are located at different computers and communicate via chat.
- *Assessment*: When they consider the problem done, students hold a structured and guided discussion on what occurred and what they have learned.

Let us analyze the first two phases separately. Analysis of the Assessment phase is omitted for brevity.

In the preparation phase, the student who will later play the role of tutor has the following options at each step in solving the problem:

Start
1. The student applies a shallow strategy, such as guessing or copying from a hint
    1.1. The student enters a correct answer → Exit having not learned much.
    1.2. The student enters an incorrect answer → Start
2. The student applies the target knowledge.
    2.1. The student enters a correct answer → Exit having learned something.
    2.2. The student enters an incorrect answer and get feedback → Start
3. The student asks for & get a hint from the tutor → Start

Because students know that they will soon have to tutor another student on exactly this same problem, they are less likely to apply shallow strategies and more likely to reflect deeply about their strategies. Thus, the tutor role should encourage better path choices.

In the collaboration phase, the two students (whom we call the tutor and the tutee) have the following options:

Start
4. The tutee applies a shallow strategy, such as guessing or copying from a hint
    4.1. The tutee enters a correct answer
        4.1.1. The tutor considers it correct → Exit, with little learning for the tutee and tutor
        4.1.2. The tutor considers it incorrect and gives feedback → Start, with some mislearning for the tutor
    4.2. The tutee enters an incorrect answer

        - 4.2.1. The tutor considers it correct → Exit, with little learning for both tutor and tutee
        - 4.2.2. The tutor considers it incorrect and gives feedback → Start, with some learning for the tutor
5. The tutee applies the target knowledge.
    - 5.1. The tutee enters a correct answer
        - 5.1.1. The tutor doesn't process the answer and merely agrees that it is correct → Exit, with tutee learning and tutor not learning
        - 5.1.2. The tutor thinks hard and considers it correct → Exit, with both having learned something.
        - 5.1.3. The tutor considers it incorrect
            - 5.1.3.1. The tutee objects and wins the argument → Exit, with some learning for both
            - 5.1.3.2. The tutee acquiesces or the tutor wins the argument → Start, with some mislearning for both
    - 5.2. The tutee enters an incorrect answer
        - 5.2.1. The tutor doesn't process the answer and merely agrees that it is correct → Exit, with tutee mislearning and tutor not learning
        - 5.2.2. The tutor considers it correct → Exit, with some mislearning for both
        - 5.2.3. The tutor considers it incorrect
            - 5.2.3.1. The tutee objects and wins the argument → Exit, with some mislearning for both
            - 5.2.3.2. The tutee acquiesces or the tutor wins the argument → Start, with some learning for both
6. The tutee asks for and gets a hint from the tutor → Start

Although there are many paths available, the key observation is that the paths where the tutor makes mistakes (e.g., failing to notice the tutee's error) have mislearning on them for both tutor and tutee, whereas the paths where the tutor makes no mistakes and wins any arguments that occur contain learning for both participants. This analysis predicts that learning gains will be higher when a more competent student plays the role of tutor. It is important to note that because the tutor has the answers to the problem available, and has solved the problem previously, mislearning is less likely to occur.

When this activity is compared to unscripted collaboration, there is a clear advantage. During unscripted collaboration, it often occurs that one student dominates the interaction. That is, the dominant student sits at the keyboard and solves the problem while the other student watches. For the dominant student, this means taking mostly path 2 and avoiding 1 and 3, so the dominant student probably learns well from this activity. The other student learns only if she or he self-explains most of the dominant student's actions. That seems unlikely, so the non-dominant student probably learns little.

On the other hand, the scripted collaboration, particularly when the tutor is competent, probably engages both students. The tutee is likely to avoid guessing (4), to ask for hints if necessary (6) and to otherwise try hard to infer a correct response (5). The tutor knows that the assessment phase will reveal errors that the tutor has overlooked, so the tutor should seldom take exits 5.1.1 or 5.2.1 and instead self-explain the tutees actions in order to check them. Thus, both tutor and tutee should take paths of high engagement.

When scripted collaboration is compared to solving a problem alone on the tutoring system, it is less clear that the scripted collaboration provides an advantage. Unlike the human tutor, the tutoring system always gives appropriate feedback, so there is no danger of mislearning. Thus, the key factor is how often the solo student takes the shallow learning path (1). When solo students are compared to the tutors, it is likely that the tutors will take path (1) less frequently during the preparation phase than the solo students, because they know that they need to teach this problem's solution later. Moreover, the tutors often cover the same knowledge component applications later during the collaboration phase. Thus, it is quite likely that the tutors will learn more than the solo students.

However, it is less clear what will happen when the tutees are compared to the solo students. The tutees have the same opportunities to take the shallow learning path (4) as the solo students. Perhaps the presence of the tutor may motivate them to avoid path 4, but a dominant tutor might encourage the tutee to engage in social loafing and take path 4 frequently. However, one advantage that the tutee has in the script is the opportunity to engage in more dialogue with the human tutor than with the computer tutor.

Thus, this analysis predicts that the tutor in a scripted interaction will probably learn the most, followed by the dominant student in an unscripted interaction. The solo students and the tutees in a scripted interaction should have moderate learning gains. The non-dominant student in an unscripted interaction should have the smallest gains.

This analysis is also limited in the fact that it focuses almost entirely on the peer tutor in his/her role as corrector. The students are also expected to engage in dialog with one another, and the tutor is also expected to monitor the skills of the tutee. Applying knowledge during these activities is likely to facilitate the acquisition of both collaborative and cognitive knowledge components, and lead to more learning compared to the individual condition. The shallow application of collaborative knowledge in these scenarios is not expected to be beneficial.

6 Fluency and refinement cluster

Fluency is primarily a standard of performance. Smooth performance that approximates that of the expert is called "fluent." Fluency in language includes those attributes of skill recognized by native speakers, which especially entail speech that captures the morphosyntactic structures of language. Automaticity is often cited as a feature of fluency. However, automaticity is best understood as one of the features of knowledge use that enables fluency.

Refinement is a cognitive process that can contribute to fluency. It is the acquisition of relevant features and the deletion of irrelevant features. Repeated practice with a given knowledge component involves trials that are often not completely identical. For example, repeated exposures to the French word “pôt” may include phonetically variable exemplars. To form a phonemically robust representation, the learner must engage in a process of feature pruning or refinement. This refinement begins with a single episode (a learning event) and superimposes subsequent episodes as instances of the same category. This leads to refinement of the spoken syllable.

Projects in the Fluency cluster focus mostly on path effects. Their learning event spaces often provide more than one exit that should entail learning, and their studies try to understand which form of learning is more robust. For instance, a study by Tokowicz has students learn the French words for common objects which are portrayed by line drawings. Half the line drawings present the object in a non-standard orientation, so that a student would have to mentally rotate the drawing in order to recognize the object. This can be analyzed as learning space with two desirable exits, both involving a student trying to associate a meaning with the French word, but only one path’s learning includes mental rotation. Tokowicz hypothesizes that the mental rotation should increase the robustness of the learning. In other words, the research question involves the effects of going down the mental rotation path as opposed to going down an equivalent learning path that doesn’t involve mental rotation (i.e., the objects are drawn in their standard orientation).

A variety of different path effects are being compared by Fluency cluster studies, but the most intensively studied comparison involves manipulating the features that are activated at the time of the learning event. In de Jong, Perfetti and DeKeyser’s study, certain syntactic features are activated (primed) by presenting them as part of a comprehension task immediately before the target production task. This should increase the chance of the students getting the production task right on the first try, and hence increase their learning. Other studies, including those of Lui and Perfetti, MacWhinney, Tokowicz and Pavlik, manipulate feature activation by varying the order in which items are presented for learning. Placing two items that share features close together in the training sequence tends to prime or activate the shared features during the second item’s learning event. This should increase the chance of a successful response on the first try, which should increase the robustness of the learning.

Although these studies tend to focus on path effects, their manipulations may also have an impact on path choice. For instance, if activating some relevant features tends to make a deep learning path more likely to succeed on the first try, then repeated successes would tend to make students choose the deep learning path more frequently. Even in these relatively simple tasks, students have path choices that cannot be ignored.

6.1 Successful recall vs. unsuccessful recall plus feedback (de Jong, Perfetti, DeKeyser)

In this project, de Jong, Perfetti and DeKeyser are studying a manipulation that could affect the probability that a deep learning path will produce a correct answer, and that this will lead to more robust learning.

In the task domain of French grammar learning, they studied a production task, where students must write a grammatically construction of a specified type, alternating with a comprehension task, where they are given an instance of a construction and asked to identify its type. In the control task the production and comprehension items do not alternate.

In the production task, students enter an answer only once. If it is incorrect, they receive negative feedback and the correct answer, and then go on to the next task. Thus the choices are:

Start
1. Use a shallow strategy, such as guessing
    1.1. Correct answer → Exit, having learned little
    1.2. Incorrect answer & feedback
        1.2.1. Deeply process the feedback → Exit, having moderate learning
        1.2.2. Shallowly process the feedback → Exit, with little learning
2. Try to apply the relevant knowledge components
    2.1. correct answer → Exit, having learned
    2.2. Incorrect answer & feedback
        2.2.1. Deeply process the feedback → Exit, having moderate learning
        2.2.2. Shallowly process the feedback → Exit, with little learning

If this task is preceded by a comprehension task, then it changes in the following way. The comprehension items *prime* the production items, increasing the likelihood of a correct answer given that students are trying to apply their knowledge. According to the theoretical framework, path 2.2.1 introduces features into the context (namely, the feedback itself) that will be absent later when recall is needed. Thus, the learning from 2.2.1 should be less effective than from 2.1.

Note that this training procedure does not afford the shallow strategy that involves copying a surface feature of the comprehension task to the production task's answer. For instance, after doing a comprehension task involving "J'ir**ais** en vacances," copying the verb ending "ais" to the production task and entering "Il manger**ais** un gateau." (correct answer is manger**ait**) leads to an incorrect answer. This unsuccessful strategy discourages the shallow (1) paths. Only copying of the deeper feature of modality (conditional versus indicative) leads to a correct answer. This suggests that any benefits observed can be attributed to path effects rather than path choices.

6.2 Implicit vs. explicit instruction on word meanings (Juffs & Eskenazi)

Juffs and Eskenazi are comparing the effects of two different paths to successful acquisition of new vocabulary items. When following one path, students learn from explicit instruction; when following a different path, they learn from a kind of implicit instruction.

When a student cannot recall the meaning of a word, they can construct a meaning either from its context in the text they are reading, or from a pop-up definition. In particular, ESL students read a text containing "target" words, which are words that they missed on the pre-test. The target words are in bold font. For half the students, clicking on a bold word pops up its definition. The other students do not have this feature available. When a student encounters a bold word, this is their learning event space:

Start

1. Recall the meaning of the word & it fits the context → Exit, having strengthened and refined the knowledge component
2. Fail to recall a meaning that fits the context; do not try to learn the word → Exit, having learned little
3. Fail to recall a meaning that fits the context; try to learn the word's meaning
    - 3.1. Click on the word
        - 3.1.1. Infer a correct meaning → Exit, having learned explicitly
        - 3.1.2. Infer an incorrect meaning → Exit, having learned a buggy meaning
        - 3.1.3. Either do not study the definition or fail to infer a meaning → Start
    - 3.2. Infer the meaning of the word from context
        - 3.2.1. Infer correct meaning → Exit, having learned implicitly
        - 3.2.2. Infer incorrect meaning → Exit, having learned a buggy meaning
        - 3.2.3. Fail to infer a meaning → Start

If the definition feature is unavailable, then path 3.1 is unavailable. (Although students can look up the word offline in a dictionary, this occurs rarely during this study.) This means that any learning that occurs is implicit (3.2.1 and 3.2.2).

When a student tries to infer a word's meaning from a definition (3.1), the cognitive processes can be complex. The student is reading a definition written in the student's second language, English in this case, and may not understand the definition correctly. The student may alternate between reading the definition and examining the word's context in the training text. The more complex this inferential process becomes, the more invalid the retrieval features created by it. This is somewhat less of a problem when inferring a words meaning without the definition being present (3.2), as even when that process is complex, the perceptual features are more similar to ones available on post-test due to the absence of a definition.

Although the study was designed under the assumption that shallow strategies (path 2) would be rare in both conditions, observation of students during pilot testing showed that many were clicking on definitions but not trying to infer the meanings of words. That is, they were taking the path 3.1.3 → 2. A multiple choice comprehension test was added at the end of each passage, and teachers increased their vigilance. It appears that this behavior has decreased. However, the study does not control nor measure the frequency of shallow learning (2), which may be affected by whether or not definitions are available. Future work could address this issue with denser data, such as concurrent verbal or written protocols, self-reports or eyetracking.

6.3 Video vs. audio-only training of pronunciation (Liu, Perfetti & Wang)

Study 2 of Liu, Perfetti and Wang compared two methods of teaching students how to pronounce Chinese characters. Students are repeatedly presented with a Chinese character, its English translation and a spoken pronunciation of it. The pronunciation is either a video headshot of a person speaking the Chinese, or only audio of the spoken Chinese character with a still picture of the speaker. When a student experiences such a presentation, there are several choices

Start
1. Process the inputs shallowly → Exit, with little learning
2. Process the video deeply → Exit, with learning
3. Process the audio deeply → Exit, with learning
4. Process both audio and video deeply → Exist, with robust learning

In the condition where only the audio is available, then paths 2 and 4 cannot be followed.

The results showed that video training was more effective than audio-only training in some testing tasks. There are two explanations that need to be teased apart in future work.

- *Path choice*: The video could have been more engaging than the audio-only training. That is, in the video condition, path 1 was less frequent than it was in the audio-only condition. The longer average viewing time for video plus audio condition in the logged data partially supported this possibility.
- *Path effects:* If the shallow learning path was equally frequent, then the results suggest that the extra information provided by the image of the speakers face provides value added over the information in the audio alone.

This study's results illustrate how an instruction manipulation can be a complete success at the macro level of the theory—video really was more effective than audio—but questions still remain at the micro level.

6.4 Basics skills training (MacWhinney)

Every teacher knows that students must learn the basic skills before trying to learn more complex skill that call upon the basic skills. However, the details of this pedagogical principle are not well understood, in part because the relationships between basic and complex skills seem highly variable. In order to elucidate this relationship, MacWhinney and his colleagues are investigating a variety of basic skills in French and Chinese learning, including the following:

- *French dictation*: Given a spoken French word or phrase, write it in proper French orthography, spelling it correctly.
- *French inflection*: Given a written phrase or sentence with incomplete or incorrect inflectional morphology (gender, verb conjugation, etc.), the student provides the correct inflectional markings.

- *Speeded Chinese tone recognition*: Mandarin has four main tones, a neutral tone, and some tone assimilations. Students listen to input words and then type number keys to indicate the correct tones.
- *Pinyin dictation:* After hearing a spoken Mandarin word, students enter the pinyin which includes tone markings.
- *Chinese character dictation*: After hearing a word, students use the Chinese input method to enter the correct character.
- *Character dictation to pinyin*: After seeing a Chinese character combination, students type in the corresponding pinyin. This is essentially a vocabulary learning task.

All these studies use the same experimental procedure. Students are trained on half the training items and tested on all of them. This measures near transfer based on knowledge components that are shared among training items. Students' performance in subsequent exercises in the course that use these basic skills is also tracked. This measures transfer from basic skills to complex skills. One analysis of the learning events space is:

Start

1. The student uses a shallow approach, e.g., guessing
   1.1. Correct response → Exit, with little learning
   1.2. Incorrect response & feedback
      1.2.1. Study the feedback & try to learn from it → Exit, with some learning
      1.2.2. Ignore the feedback → Exit, with little learning
2. Without analysis or decomposition of the stimulus, the student retrieves a response
   2.1. Correct response → Exit, having strengthened the knowledge component
   2.2. Incorrect response & feedback
      2.2.1. Study the feedback & try to learn from it → Exit, with some learning
      2.2.2. Ignore the feedback → Exit, with little learning
3. Analyze or decompose the stimulus to find familiar parts or features, and infer a response from them
   3.1. Correct response → Exit, having constructed a knowledge component
   3.2. Incorrect response & feedback
      3.2.1. Study the feedback & try to learn from it → Exit, with some learning
      3.2.2. Ignore the feedback → Exit, with little learning

In the first study (French dictation), it was found that students gained just as much on untrained items as they did on trained items. This suggests that students used path 3 extensively during both training and testing, and that the trained items shared many parts or features with the untrained items. A fine-grained analysis of errors and latencies in the log data should allow determination of what those shared parts and features are.

The theoretical framework suggests that students will learn more from partially familiar items than from totally unfamiliar items. Suppose students are given an item where all but one of the parts and features are familiar. The students are more likely to take path 3 than path 1, and to take path 3.2.1 over path 3.2.2 if they respond incorrectly. Thus, they should learn about the single part or feature of the training item that was unfamiliar to them. On the other hand, if an item has few familiar parts, then students are more likely to take path 1 than 3, and thus they will learn very slowly.   This prediction can be tested with log file data.

It also suggests improving the tutoring system so that it presents items only when they are partially familiar. This can be done statically, via a cleverly designed sequence of training items; or dynamically, via tracking the individual student's mastery of fine-grained knowledge components and choosing items that are partially familiar to this particular student.

6.5 First language effects on second language grammar acquisition (Mitamura)

This study examines the impact of the grammar of a student's first language on the student's learning of the grammar of a second language. It is expected learning will be slowed down when the grammar of the first language conflicts with the grammar of the second language. Negative transfer may also occur with respect to robust learning measures, such as retention.

Specifically, students will be taught the distinction between the definite and indefinite articles in English. Students' first languages will either lack articles (e.g., Korean, Japanese) or have an article grammar that differs from the English one (e.g., Arabic). Students will practice using a simple fill-in-the-blank tutoring system, which gives immediate feedback and hints when an incorrect article is supplied. Thus, the learning event space is

Start
1. Guess
    1.1. Entry is correct → exit, with little learning
    1.2. Entry is incorrect → Start
2. Use the article of one's first language
    2.1. Entry is correct → exit, with possibly mistaken learning
    2.2. Entry is incorrect → Start
3. Try to apply knowledge of English article grammar
    3.1. Entry is correct → Exit, with learning
    3.2. Entry is incorrect → Start

The second set of paths (2, 2.1, 2.2) are only available to students whose first language has articles. However, even if the students' first language has articles, they may not choose path 2. They may rapidly learn that choice 2 often leads to errors (2.2) so they may stop using their first language as the default solution. In that case, the expected negative transfer may not occur during learning.

After both groups of students have reached mastery on the English articles, it will be interesting to see if their first language affects the robustness of their learning. For instance, retention tests might show that first-language errors, which were stamped out during training, have reappeared. That is, when several weeks have passed since training, the memories of English articles are no longer as recent, so students whose first language's articles conflict with the English articles may display higher error rates. If this kind of negative transfer were found, it would tend to disconfirm the path independence hypothesis (Klahr & Nigam, 2004; Nokes & Ohlsson, 2005), which is that when students reach a certain level of competence, it doesn't matter how they got there; their subsequent performance, including both retention and acceleration, will be the same.

6.6 Optimizing the practice schedule (Pavlik)

The order in which factual content items appear during drill may affect the rate of learning as well as the retention, transfer and perhaps even acceleration of future learning. These effects of practice scheduling can help us understand the basic mechanisms of robust learning by showing how learning events for knowledge components result in long-term performance, how knowledge components can transfer to new contexts and how the current state of the knowledge component space affects future learning.

In earlier work, Pavlik (2006, in press) developed a scheduling algorithm that optimized the spacing of practice items by balancing between the competing advantages of better immediate correctness caused by narrow spacing of practice and better long-term retention caused by widely spaced practice. In the models and experiments that underpinned the algorithm, the practice items were all of the same type (e.g., a Japanese word was presented, and an English word was entered by the student) and there were few strong inter-item relationships. However, in second language classes, there are often strong inter-item relationships. For instance, one item may present an English word and elicit its Chinese translation, while another item may present the Chinese word and elicit its English translation.

In the first of a series of studies, Pavlik is having English-speaking students learn Chinese words. The experiments use 6 item formats of prompt-response: Hanzi-English, sound-English, Pinyin-English, Hanzi-Pinyin, sound-Pinyin, English-Pinyin (where "sound" means the Chinese pronunciation of the word; the other stimuli are written). The training obeys the following policies. On the first presentation of an item, both the stimulus and the response are presented and the student is invited to study them. On other subsequent occurrences of the item, the stimulus is presented and the student must begin entering a response within 6 seconds. If they fail or enter an incorrect response, they are shown the correct response. Thus, the learning event space for an item presentation could be analyzed as follows:

Start
1. The item is an example (i.e., the response is displayed)
   1.1. Student studies the pairing deeply → Exit, with learning
   1.2. Student does not study the pairing deeply → Exit, with little learning
2. A response is needed, and student tries to recalls the pair using all available related information

2.1. Response is correct → Exit, having strengthened the target knowledge
2.2. Response is not correct → return to path 1

The first study compared two conditions, both using the previously developed algorithm to present items in optimally spaced schedules. In the Mixed condition, all items were available to be picked by the algorithm. In the Blocked condition, two item pools were used: (1) pairs where the response was in English, and (2) pairs where the response was in Pinyin. The algorithm switched selection between the two pools in blocks of 60 trials during learning.

The impact of this manipulation (better, but slower performance in the Mixed condition) on learning was partly determined by the path choices. It is likely that students in the Mixed condition followed path 2.1 more frequently than students in the Blocked condition because Mixed condition subjects benefited from more recent presentations of related information (this appears to be transfer dependent on the recency of the transferable learning events). Here, two pairs are considered related if they refer to the same knowledge component. For instance, a test where the pinyin ma1 is presented and the English word mother is expected as a response draws on the same knowledge component (the pinyin orthography for ma1) as a trial presenting the Hanzi character for ‘mother’ and requiring a pinyin typed ma1 response. Blocking artificially separates many such opportunities for transfer making transfer less likely

A long-term assessment session in this study (2 days) suggested that a portion of the learning is robust, though condition differences were not significant. This long-term assessment showed that the 40 minutes of practice resulted in 39% recall for the first test trials for each of the 24 items (averaged across the 6 possible trial types).

6.7 Semantic grouping during vocabulary training (Tokowicz)

As is often the case in second language vocabulary instruction, the French LearnLab course currently presents vocabulary items in blocks of semantically-related words, such as a group of clothing vocabulary items, a group of animals, of moods, vegetables, etc. This study will compare semantically grouped presentation with semantically mixed presentation, wherein words from several semantic groups are intermingled in random order. Each student will learn some groups of words in each presentation format. Only the order in which words are presented will be manipulated; the various exercise formats currently used in the course will remain unchanged. Thus, the learning event spaces will be the same in both the semantically grouped and semantically mixed presentation conditions.

In order to analyze the learning event spaces, let us divide semantic features into *category* features (e.g., the “vegetable” feature), *sub-category* features which discriminate among members in a category (e.g., the “carrot” vs. the “broccoli” features) and *other* features (e.g., “green” vs. “orange”). When students are practicing a block of items that all come from the same semantic category, the category feature is highly activated because all the items have the same category feature. Thus, the feature may be come associated with the responses as a contextual feature. However, the student is striving to distinguish among different members of the category, so the students should focus attention on subcategory features or other features. For instance, if the student has just successfully retrieved the French word for “carrot” and now faces “broccoli,” the student should be focusing attention on images, mnemonics or other links for “green” or “broccoli.” Deliberately focusing on the category feature “vegetable” will not help. On the other hand, when blocks of items have mixed category memberships, then focusing attention on the category feature might help. Thus, in the semantically homogeneous blocks, the category feature is associated to the response “involuntarily” as a contextual feature, while in semantically heterogeneous blocks, the category feature might be associated more deliberately to the response, should the student chose to focus on it. This suggests the following learning event space:

Start
1. The student applies a shallow strategy, such as guessing or copying from a hint
    1.1. Response is correct → Exit, with little learning
    1.2. Response is incorrect; tutor gives a hint → Start
2. Focusing on a category feature, student recalls something
    2.1. Response is correct → Exit, with strengthening of the category feature
    2.2. Response is incorrect; tutor gives a hint → Start
3. Focusing on a subcategory or other feature, the student recalls something
    3.1. Response is correct → Exit, with strengthening of the non-category feature
    3.2. Response is incorrect; tutor gives a hint → Start

The manipulation should affect the frequency of path 2 vs. path 3. It should also affect whether the category features are highly activated contextually. These differences should affect students’ ability to recall both the word-level responses (given “apple,” say the French word for “apple”) and category memberships (given “apple,” say the French word for “fruit”).

However, the manipulation could affect the frequencies of path 1. For instance, the students might find learning words in semantic groups to be more boring than learning them in semantically mixed order, so they take path 1 more frequently. On the other hand, learning the words in semantic groups might give students a sense of accomplishment when a group is finished (e.g., "Wow, now I know all of the vegetables!"), motivating them to try harder and take path 1 less frequently. Thus, it is simply not clear whether the manipulation will have an effect on the students’ path choices. However, this information may be discernable from log data.

In addition to path choice differences, the PSLC theoretical framework suggests that the features that are salient during learning should match those that are salient during testing in order to increase recall accuracy and speed. When words are learned in semantically grouped lists, the features that are salient are unlikely to be the same as those that would be required to process a single word either in isolation (as when one attempts to remember a word), or in larger linguistic context (as when one attempts to comprehend a word in context or speak/write a word in a sentence), because the word meaning will be emphasized. Therefore, the semantically mixed training condition is more likely to facilitate processing in these contexts. By contrast, benefits of semantically grouped learning may be expected when the task is to identify category membership.

6.8 Mental rotations during vocabulary acquisition (Tokowicz)

For easily drawn objects with unambiguous names (e.g., piano, ashtray, fire extinguisher), it is common to use pictures during second language vocabulary training. The pictures are invariably drawn in their natural orientation. This study manipulates the orientations of such drawings. Essentially, it compares a "normal" vocabulary learning with and without a concurrent mental rotation task.

Students will learn French vocabulary under three training conditions: (1) in association to English words, (2) in association to pictures in usual orientations, and (3) in association to rotated pictures. For example, in condition 3, a student may be shown a drawing of an ashtray that has been rotated 180 degrees and asked to learn that it is a "cendrier." A learning event spaces for this study is:

Start

1. The student applies a shallow strategy, such as guessing or copying from a hint
   1.1. Response is correct → Exit, with little learning
   1.2. Response is incorrect; tutor gives a hint → Start
2. Student applies a deep strategy
   2.1. Response is correct → Exit, with some learning
   2.2. Response is incorrect; tutor gives a hint → Start

At this level of analysis, although the paths *available* to students are not affected by the manipulation, the *frequency* of paths may be affected, and this may be discernable from the log files. It could also be that the rotated pictures take somewhat more time and effort than the unrotated pictures. (Latency can be determined from the log data.) If so, the extra effort may increase the frequency of the shallow learning path (1). On the one hand, the extra attention required by the rotated pictures may increase the strength of the connection between the meaning of the word and the label in French, and therefore performance will be faster and more accurate in the rotated picture condition relative to the standard orientation picture condition.

By contrast, when students are asked to translate words that were learned in association to drawings in usual orientations, they will show more interference from the dominant language, which will lead to weaker knowledge components and less valid features; therefore, less accurate performance would be predicted.

Thus, both Tokowicz studies predict an advantage for French vocabulary when the training features match the testing features and when knowledge components are strong. The important research issue is whether this effect is robust. Although there has been some research on these issues in laboratory studies of paired associated learning, this study will take place in the French LearnLab. Students are accessing their evolving vocabulary for many exercises besides the vocabulary exercises where the manipulations occur. Diligent students may study the vocabulary privately, whereas less diligent students may not complete the vocabulary exercises where the manipulations occur. Against this real world "noise," will the benefits of feature matching be detectable? If so, will the words trained in the experimental conditions always maintain an advantage, or will the control words eventually catch up? Is there any differential acceleration of future learning as a result of richer semantic networks that would be expected in the experimental conditions?

6.9 Visual enhancement of Chinese tone learning (Wang, Lui & Perfetti)

This study manipulates the information available to students who are learning to hear the difference between the tones of spoken Chinese. The basic hypothesis is that providing a visual display of the speech waveform may help students infer the tone's identity as they hear it, and this might increase their learning.

There are 5 tones in Chinese, and discriminating among them is difficult for most students to learn. In the Chinese LearnLab course, students are trained to discriminate among the tones by listening to a spoken Chinese syllable while reading its Pinyin representation, selecting the tone they heard from a menu, and receiving immediate feedback on their selection. This drill continues throughout the year as the students' vocabulary increases. Pinyin is a standard orthography for representing the sound of Chinese. It uses numbers to represent tones. Thus, /ma/4 represents a syllable with tone 4, whereas /ma/2 represents a completely different syllable that employs tone 2. When the Pinyin is present during tone training, the numbers are left off.

The manipulation studied here involves changing the information available during the training. In one experimental condition, students saw a wave form displaying acoustic information in the speech, heard the speech and read the Pinyin. In another experimental condition, they saw the acoustic information and heard the speech, but the Pinyin was absent. In the control condition, students saw no wave form but saw Pinyin and the tone numbers. The assessment was done with speech only; neither Pinyin nor the wave form were presented during testing.

When all three types of information are present (speech, wave form and Pinyin), the students in the experimental condition can attend to any combination of them. To display all combinations as paths in the learning event space would be tedious and unhelpful. The following abstract learning event space is perhaps a more helpful analysis:

Start
1. Use a shallow strategy, such as guessing
   1.1. Response is correct → Exit, with little learning
   1.2. Response is incorrect → Start
2. Use the speech alone to categorize the tone
   2.1. Response is correct → Exit, having strengthened some correct knowledge
   2.2. Response is incorrect → Start

3. Use the wave form to guess the tone, “subtract” the Pinyin from the speech, and associate the remaining speech with the presumed tone
    3.1. Response is correct → Exit, having learned
    3.2. Response is incorrect → Start
4. Ignore the wave form (if it is present), “subtract” the Pinyin from the speech, and try to categorize the remaining speech
    4.1. Response is correct → Exit, having learned
    4.2. Response is incorrect → Start
5. Ignore the Pinyin (if it is present), use the wave form to guess the tone
    5.1. Response is correct → Exit, having learned only a little
    5.2. Response is incorrect → Start
6. All other strategies
    6.1. response is correct → Exit
    6.2. Response is incorrect → Start

The basic idea, which is only crudely represented above, is that students may learn best when they can somehow subtract the non-tonal speech (represented by the number-less Pinyin) from the given speech, and at the same time know with reasonable certainty what tone this residual sound represents. The wave form may be supplying the latter piece of information, namely the tone that the sound represents. On this view, exiting via path 3.1 provides the maximal learning.

Only the first experimental condition allows path 3, since only it provides the wave form and the Pinyin along with the speech. Thus, one would expect students to learn more from it, which is exactly what was found in a LearnLab study spanning the whole fall semester of the Chinese LearnLab course.

In particular, analyses of learning curves across the 8 lessons over a semester showed a significant main effect of training condition. The experimental group for whom learning of the tones is accompanied by visual presentation of pitch information as well as Pinyin had significantly lower error rates than the other experimental group and the control group for Lessons 2, 5, 6, and 7.

The experimental group also showed a larger pre-test to post-test gain on the Tone Judgment Task, compared to the control group. The improvement was shown on more difficult items when the two Chinese syllables had the same rhyme and required a NO response, or when the two syllables had different segmental information and required a YES response.

7 References

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147-170.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004). Toward tutoring help seeking (Applying cognitive modeling to help-seeking skills). In J. C. Lester, R. M. Vicari & F. Paraguacu (Eds.), *Intelligent Tutoring Systems: Seventh International Conference: ITS2004* (pp. 227-239). Berlin: Springer.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science, 26*(1), 85-112

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology, 95*, 774-783.

Bloom, B., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook I: Cognitive Domain*. Green, NY: Longman.

Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of Eleventh Annual Conference on Computational Learning Theory (COLT),* (pp. 92–100). New York: ACM Press.

Bransford, J. D., Brown, A.L., & Cocking, R.R. (1999). *How people learn: Brain, mind, experience and school*. Washington, D.C.: National Academy Press.

Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher, 26*(8), 4-16.

Bruer, J. T. (2002). Avoiding the pediatrician's error: How neuroscientists can help educators (and themselves). *Nature Neuroscience, 5*, 1031-1033.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chi, M. T. H. (2000). Self explaining expository texts: The dual process of generating inferences and repairing mental models. In Glaser, R. (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates.

Clark, R. C., & Mayer, R. E. (2003). *e-Learning and the Science of Instruction : Proven Guidelines for Consumers and Designers of Multimedia Learning*. San Francisco: Jossey-Bass.

Dienes, Z. & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences, 22*, (5).

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. San Diego: Academic.

Gholson, B., & Craig, S. D. (in press). Promoting constructive activities that support learning during computer-based instruction. *Educational Psychology Review*.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267-333.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *Journal of the Learning Sciences. 5*(1), 23-49.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*(6), 649-667.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 12*, 391-438.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science, 15*(10), 661-667.

Krashen, S. (1994). The Input Hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-78). San Diego: Academic.

Lo Bianco, J., Liddicoat A. J. , and Crozet C. (1999) *Striving for the Third Place: Intercultural Competence through Language Education*. Language Australia, Melbourne.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). New York: Oxford University Press.

MacWhinney, B., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Moreno, R. and Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Ed. Psych., 93*, 724-733.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned? *Cognitive Science*, *29*(5), 769–796.

Pavlik, P. (2005). Modeling order effects in the learning of information.

Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, *51*, 73-75.

Nokes, T. J., & Ohlsson, S. (2005). Comparing multiple paths to mastery: What is learned? *Cognitive Science, 29*(5), 769–796.

Ohlsson, S. (1993) Abstract schemas. *Educational Psychologist*, *28*(1), 51-66.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Learning and Verbal Behavior, 6*, 855-863.

Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm for instructional theory: Volume II* (pp. 5-30). Mahwah, NJ: Erlbaum.

Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist, 38*, 15-22.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1-15.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: *Human Learning and Memory, 4*(6), 592-604.

Siegler, R. S. (1996). *Children's thinking*. Upper Saddle River, NJ: Prentice Hall.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). New York: Cambridge University.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*, 185-233.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Trafton, J. G. and Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.) *Proceedings of the 15th annual conference of the Cognitive Science Society*, 1017-1022.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.