# Sketch Worksheets: A Sketch-based Educational Software System

**Panrong Yin[†], Kenneth D. Forbus[†], Jeffrey Usher[†], Brad Sageman[††], Benjamin D. Jee[†††]**

[†]Qualitative Reasoning Group, Northwestern University,
2133 Sheridan Road, Evanston IL 60208
pennyyin@u.northwestern.edu, {forbus, usher}@northwestern.edu

[††]Department of Earth & Planetary Sciences, Northwestern University,
1850 Campus Drive, Evanston IL 60208
brad@earth.northwestern.edu

[†††] Department of Psychology, Northwestern University
2029 Sheridan Road, Evanston, IL 60208-2710
b-jee@northwestern.edu

## Abstract

Intelligent tutoring systems and learning environments can provide important benefits for education, but few have been developed for heavily spatial domains. One bottleneck has been the lack of rich models of visual and conceptual processing in sketch understanding, so that what students draw can be interpreted in a human-like way. This paper describes *Sketch Worksheets*, a form of sketch-based educational software that mimics aspects of pencil and paper worksheets commonly found in classrooms, but provides on-the-spot feedback and support for richer off-line assessments. The basic architecture of sketch worksheets is described, including an authoring environment that allows non-developers to create them and a coach that uses analogy to compare student and instructor sketches as a means to provide feedback. A pilot experiment where sketch worksheets were used successfully in a college geoscience class in Fall 2009 is summarized to show the potential of the idea.

## Introduction

Almost no intelligent tutoring systems exist for spatial topics. For example, at the 2009 International AI & Education conference only one paper out of 60 involved spatial reasoning. This means that the benefits of ITS's and intelligent learning environments are unavailable for highly spatial disciplines, which includes many aspects of science, engineering, and mathematics. A key bottleneck is sketch understanding: People bring rich conceptual structures and a powerful visual apparatus to creating sketches. Language plays an important role in human-to-human sketching, both because most people are not artists and because most spatial concepts do not have easily recognized visual symbols, unlike alphabets and schematics. In Figure 1, for example, the shape of the sketch constituents (which are called *glyphs*) are governed by the shape of the entities being described. *Open-domain sketch understanding* systems (Forbus et al, 2008) use specialized interfaces to avoid the need for recognition, large knowledge bases to provide a model for rich conceptual structures, and models of visual perception to encode sketches in human-like ways. We believe that such systems can help bring the benefits of intelligent tutoring systems and learning environments to highly spatial domains.

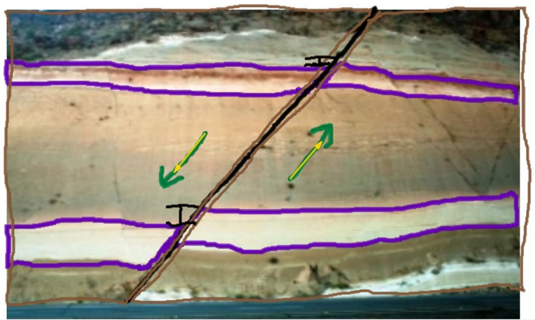


**Figure 1: A sketch worksheet for geoscience, showing the instructor's solution**

Paper-based worksheets are a staple in many classrooms. For example, in a geology class, students might be asked to highlight a fault on a photograph, or draw the layers of the Earth. Sketching is a valuable way of learning spatial relationships. Consequently, we are developing *sketch worksheets* as an electronic analogue to pencil and paper worksheets, with the goal of making them as widely usable as pencil and paper worksheets are today. With pencil and paper sketching, feedback is delayed, and assessment is time-consuming and difficult. With sketch worksheets, we can provide rapid feedback to students, and hopefully

Yin, P., Forbus, K., Usher, J., Sageman, B. & Jee, B. (2010). Sketch Worksheets: A Sketch-based Educational Software System. Proceedings of the 22nd Annual Conference on Innovative Applications of Artificial Intelligence.

make assessment simpler and more efficient, thereby improving learning.

Figure 1 shows a sketch worksheet for a physical geology class, where the ink illustrates the instructor's solution. The task for the students is to identify the main fault in the photograph (black diagonal line in the middle), noting the marker beds (outlined in purple), hanging wall and foot wall (large regions in brown), and indicating which way the marker beds are moving and what their displacements are (the green arrows and black I-shaped annotations, respectively). When a student tackles this worksheet, they see the photograph but not the instructor's sketch. They draw their own sketch, creating glyphs and labeling them with concepts from a set provided by the instructor. When the student asks for feedback, the student's sketch is compared to the instructor's sketch, and the differences are used to provide suggestions, based on instructor-authored information stored with the worksheet. After several iterations of improving their sketches, students turn them in, and they are further analyzed off-line for assessment purposes.

We first review CogSketch, the open-domain sketch understanding system we use. Next we describe the basic architecture of worksheets, and illustrate via an example. Then we describe a pilot experiment where worksheets were used in a physical geology course, which provides evidence that this model can be useful in real-world classrooms. We close with related and future work.

## Background: CogSketch

Most sketch understanding systems treat understanding as a matter of recognizing ink, or ink plus speech, as a member of a limited number of predefined symbols (e.g., military symbols (Pittman et al, 1996), electronics/UML diagrams (Alvarado et al, 2002)). This limits them to expressing a small, fixed set of concepts. Recognition-based interfaces can be of great practical value – handwriting recognition is the most successful example – but our goal is fundamentally different. A key insight of our approach is that in human-to-human sketching, recognition is a catalyst, not a requirement (Forbus et al 2008). When people sketch with each other, we typically also talk, using language to label the intended meaning of pieces of ink, or of the spaces defined by the ink. CogSketch uses menus to enable users to select the concept(s) that their ink depicts from an OpenCyc[1]-derived knowledge base (KB) with over 58,000 concepts, including our own additions to support visual, spatial, qualitative, and analogical reasoning. While neither perfect nor sufficiently broad coverage for all that we do, OpenCyc contents have proven to be a useful engineering approximation for conceptual knowledge.

CogSketch is designed to model the perceptual and spatial understanding that people bring to sketching. It is based on two key hypotheses:

(1) *Perceptual processing produces qualitative spatial representations*. Qualitative representations quantize continuous properties, making meaningful units that can be manipulated symbolically. Constructing representations of segments, regions, volumes and relationships between and within them is, we argue, one of the key functions of perception.

(2) *Structure-mapping processes are used in visual reasoning*. Gentner's structure-mapping theory (Gentner, 1983) defines analogy and similarity in terms of comparison of structured representations. Qualitative representations provide visual and spatial structure which is used in analogical operations of matching, retrieval, and generalization. These analogical operations are used to identify similarities and differences, and form components in models of larger-scale cognitive processing. For example, analogical matching and generalization have been used to model geometric analogy problems (Lovett et al 2009) and Raven's Progressive Matrices test (Forbus et al, 2008). The Structure-Mapping Engine (SME) (Falkenhainer et al, 1989) is built into CogSketch, and is used in coaching for comparing the instructor and student sketches.

A user draws a glyph in CogSketch by pressing a button when they start and stop drawing. Manual segmentation is not as elegant as automatic techniques but it is far more reliable. For example, timeouts or pen-up constraints commonly used in sketch recognition systems are quite problematic for students who are trying to draw complex shapes while thinking hard about the domain. The entity that the glyph depicts is explicitly represented internally, and can be categorized through conceptual labeling.

There are three kinds of glyphs. *Entity glyphs* represent physical or conceptual objects. Examples in Figure 1 include the fault and the marker beds. *Relation glyphs* describe relationships between entities. For example, in describing the carbon cycle, arrows are used to depict transfers between different reservoirs of carbon. *Annotation glyphs* provide additional information about other glyphs. For example, the arrows in Figure 1 indicate the direction of movement, and the I-shaped annotations indicate displacement. The meaning of relation glyphs and annotation glyphs, like entity glyphs, are specified by menus which are populated with concepts from the KB.

The result of sketching is a structured relational representation that combines visual, spatial, and conceptual entities and concepts, formally represented using concepts from the KB. This allows a combination of visual, logical, and analogical reasoning to be performed with them. The visual and spatial aspects of this representation are automatically computed based on the user's ink, as interpreted through the lens of their conceptual labeling.

A CogSketch sketch consists of one or more *subsketches*. The *metalayer* provides a view where all of the subsketches are themselves glyphs, so that they can be related and annotated like other glyphs. For sketch worksheets, one subsketch contains the teacher's solution and the student's work is done on another.

[1] http://www.cyc.com/opencyc

## The Sketch Worksheet Workflow

Figure 2 illustrates the sketch worksheet workflow. We use the fault identification worksheet from Figure 1 as a running example.

### Expert authors worksheet

Worksheets are authored via a special-purpose interface built into CogSketch. Creating a worksheet involves four steps: (1) Creating the problem statement, (2) drawing the solution sketch, (3) selecting the subset of concepts that the student will see, and (4) providing coaching advice. We discuss each in turn.

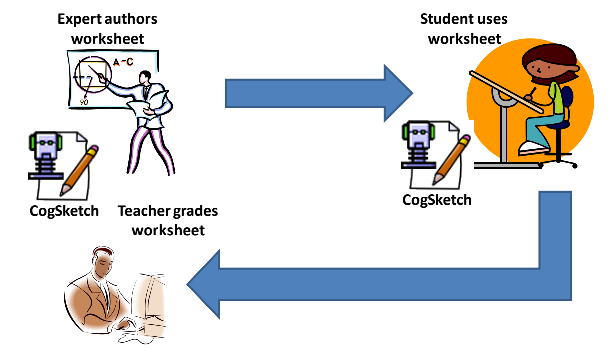


**Figure 2: The sketch worksheet workflow.**

The problem statement is a piece of text that will be displayed at the bottom of the worksheet, indicating what the student is supposed to do. For example, in the fault identification worksheet, students are given eight steps to follow, involving drawing and labeling glyphs, e.g., “Outline each half of the two prominent marker beds that allow you to identify the fault, and label them.”

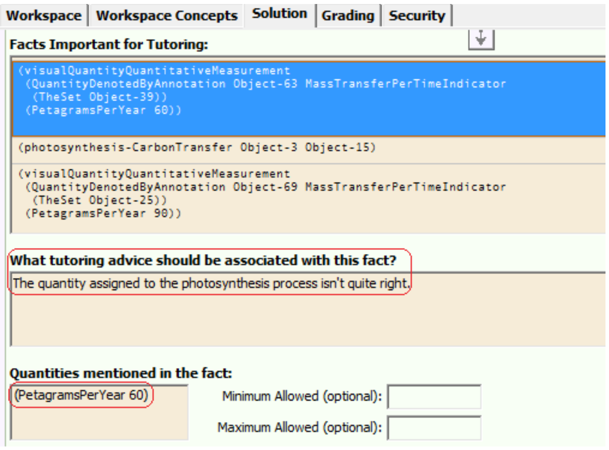
The solution sketch is the sketch against which student sketches are compared, in order to generate advice while interacting with the student, and for evaluating assessment rubrics when scoring worksheets. If an image is part of the worksheet, it must be provided. (Worksheets do not require images – for example, a worksheet asking students draw the layers of the Earth does not, because that would reveal too much of the answer.) The author draws whatever glyphs are needed and conceptually labels them appropriately. At this stage they have access to the full set of concepts in the KB.

Once the solution sketch is finished, authors then select the concepts that students will have available to use. All concepts mentioned in the solution sketch are available automatically, but for pedagogical purposes, typically some distracters are added as well. The names for these concepts that the student sees can be edited, as well as the explanatory text available when the student is browsing the concepts. This is important because the default concept names and comments are ontologist-friendly, not student-friendly. Currently worksheet authors have access to the full range of concepts in the KB, and finding what they need for a particular worksheet requires some familiarity and skill with the ontology. Importantly, students never see this level, unless the author exposes it.

Coaching advice is provided by using dialogs in the authoring environment to select particular facts in the solution sketch, marking them as important. In general, when a fact is missing, what will be said to the student can be specified (e.g., “Is this really the location of the hanging wall?” when qualitative spatial relationships involving the hanging wall are not satisfied). If the fact concerns a numerical quantity, allowable error bounds and advice when a value is high or low must be provided (Figure 3, from a different worksheet). Glyphs can also be marked for quantitative comparison, where a tolerance and advice if incorrect are specified (e.g., “Your drawing of the major fault isn’t quite right.”).

In addition to instructor-authored advice, two kinds of advice are automatically generated. If an important entity is missing, i.e., an entity mentioned in an important fact or whose ink is to be quantitatively compared, the Coach will complain about it. Whether or not extra entities in the student’s sketch are allowed can also be controlled in the authoring environment.

For security, worksheets can be password protected, so that the student cannot see the solution sketch nor can they modify it via the authoring environment.



Facts Important for Tutoring:

```
(visualQuantityQuantitativeMeasurement
 (QuantityDenotedByAnnotation Object-63 MassTransferPerTimeIndicator
  (TheSet Object-39))
 (PetagramsPerYear 60))
(photosynthesis-CarbonTransfer Object-3 Object-15)
(visualQuantityQuantitativeMeasurement
 (QuantityDenotedByAnnotation Object-69 MassTransferPerTimeIndicator
  (TheSet Object-25))
 (PetagramsPerYear 90))
```

What tutoring advice should be associated with this fact?

The quantity assigned to the photosynthesis process isn't quite right.

Quantities mentioned in the fact:

(PetagramsPerYear 60)

Minimum Allowed (optional):

Maximum Allowed (optional):

**Figure 3: Providing coaching advice in the authoring environment**

### Students use worksheet

When a student opens a worksheet, they see the problem statement plus any starting materials associated with the worksheet. This can be either a blank screen, a photograph or diagram that they are supposed to annotate, or even an initial sketch that they are supposed to modify. As they tackle the worksheet, they can get advice by pressing the feedback button. They can then improve their sketch, by adding, removing, and/or redrawing glyphs, repeating the cycle until they are satisfied.

To generate advice, the Coach uses SME to compare the student's sketch with the solution sketch. SME takes as input two descriptions, the *base* and *target*, and produces one or more *mappings*, consisting of correspondences between entities and statements in the base and target. Each mapping also includes *candidate inferences*, which are facts that can be projected from one description to the other for pattern completion, and a numerical similarity score indicating the structural quality of the match. Here the solution sketch is the base and the student's sketch is the target, with the mapping telling the Coach exactly how elements of the solution sketch and student sketch correspond. Constraints specified in the authoring environment on quantity values and quantitative ink comparisons are projected via the candidate inferences of the mapping. Missing important facts are detected by candidate inferences involving important facts. The Coach uses these results along with the strings entered via the authoring environment to generate textual feedback. When what glyph being referred to might be ambiguous, graphical highlighting is used.

### Teacher grades worksheet

For convenience, we have added a simple grading interface. The authoring environment provides the ability for instructors to assign points for including important entities (or not including extra entities), important facts being true, quantitative sketch outlines being correct, and numerical values being within tolerance. Given a filled-out worksheet, the grading system produces an HTML page with the student's sketch and a detailed breakdown of their performance, based on the instructor-supplied rubrics.

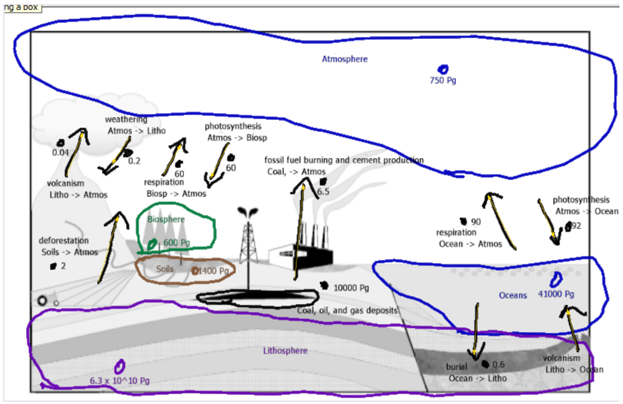


**Figure 4: Carbon Cycle worksheet, instructor solution**

## Classroom Pilot Experiment

In Fall 2009 a pilot study was conducted in a physical geology course taught by Brad Sageman at Northwestern University. The first worksheet assignment was an optional laboratory exercise, which went so well that a second assignment was developed, which was required for all students. The first assignment involved four worksheets. Three of these were fault identification worksheets, using different photographs, since learning this skill that requires practice on the part of students. A fourth worksheet, involving the chronological reconstruction of geological processes from a diagram of rock layers and geological structures, was also used. Students were given a tutorial in using CogSketch. 10 students out of 28 signed up for this extra-credit laboratory assignment. Our goals for this experiment were modest: Can students actually use sketch worksheets? The answer is yes: Students were able to complete the worksheet exercises. Student performance also seemed to be improving over the course of the exercises. The data log collected automatically by CogSketch includes the Coach's operation and timing information, which gives us a window onto student performance. There was a significant negative correlation between the time to complete a worksheet and how many worksheets they had done so far ($r(58) = -.46$, $p<.01$), i.e., students took less time on later worksheets. Moreover, there was a strong negative correlation between the number of remaining feedback items and the number of completed worksheets (r(58) = -.73, p<.01), meaning that students have fewer remaining feedback items on later worksheets.

The second assignment, which was required for all students, involved a carbon-cycle worksheet. Students were given a diagram and told to sketch on it the reservoirs for carbon, arrows indicating transfers between reservoirs, the amount of carbon stored in each reservoir, and the amount of carbon transferred per year via the transfers between reservoirs. Figure 4 illustrates the instructor's solution.) The students were able to complete the carbon cycle worksheet, and performed reasonably well on it.

### Challenges raised by the pilot study

Besides providing evidence that the worksheets were generally well-received and potentially effective for instruction, putting the worksheets to use in a real-world classroom was also useful for highlighting several problems with the current system. There were some usability problems, the worst being that students tended to just put numerical values on the labels of annotations in the carbon cycle worksheet, instead of filling out the numerical values in the annotation dialog. We believe that clearer instructions and feedback would help with this particular class of problem. Another problem is that while the mappings involving the reservoirs were always correct, the annotations were sometimes mapped incorrectly, due to poor representation design. This meant that the grading system could not be used to automatically score this assignment. We have changed the representation conventions for annotations to fix this problem.

A more subtle issue is illustrated in Figure 5. Compare the marker beds (black horizontal regions) here to Figure 1. In the instructor's sketch, they are drawn all the way to the edge of the photograph. Recall that when quantitative shape requirements are used, an error threshold is used to describe how much the student's ink can deviate from the instructor's ink. That method tacitly assumes that every boundary is equally meaningful. In Figure 5, even though the student has not drawn the marker beds all the way to the ends of the photograph, they have drawn them far enough to reassure a human instructor that they have the "right idea", because in fact the vertical boundaries of the marker beds are quantitatively accurate. Finding robust ways to correctly detect such reasonable variations, and natural ways for instructors to express such constraints, is an interesting challenge.

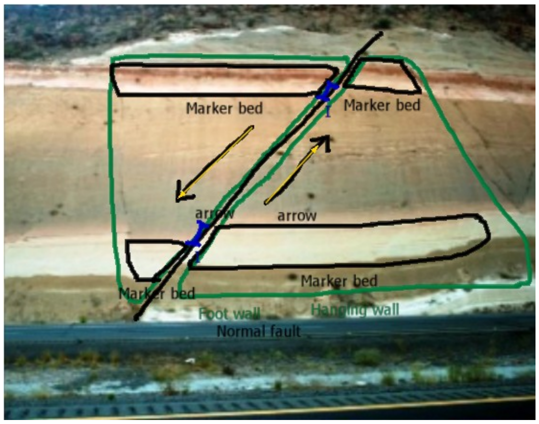


**Figure 5: While not literally satisfying the instructor's rubrics, this sketch is considered to be correct.**

## Related Work

Some sketch recognition systems combine ink recognition with speech recognition to increase the accuracy of both (e.g., Adler & Davis, 2004; Pittman et al, 1996). While this approach can be useful for domain experts, it has several problems for educational use. First, training the speech system requires considerable overhead, typically including additional training every time new vocabulary is introduced. This is not a good fit for education applications, when new vocabulary is being added constantly. Second, in many circumstances such as computer labs and classrooms, talking is frowned upon. Third, it does not provide the links to conceptual knowledge and visual processing that our approach provides, and that are necessary for many educational applications.

Sketch recognition technologies are strongest when deployed in domain-specific systems. For example, Newton's Pen (Lee et al, 2007) is designed to provide help for students constructing free body diagrams and equilibrium equations, and Kirchhoff's Pen (De Silva et al, 2007) is designed to help students learn to apply Kirchhoff's current and voltage laws when analyzing simple circuits. Both systems require highly stereotyped drawings involving a relatively small number of parts. (Newton's Pen runs in a pen with a computer inside, and thus further limits students to sketching diagrams in a specific order due to on-board computational constraints.) By contrast, nothing in the sketch worksheet architecture presented here is domain-specific. Worksheets have been built for biology exercises, for example.

Our approach of leaving the user's ink alone has much in common with the idea of informal perceptual user interfaces (Landay et al, 2002). We differ in trying to provide visual and conceptual understanding of what is drawn (cf. Forbus & Usher, 2002).

## Discussion and Future Work

This paper describes sketch worksheets, a simple sketch-based educational software system that relies on open-domain sketch understanding. Sketch worksheets use simple interface mechanics that avoids the training required by today's recognition technologies, allowing learners to focus more on the subject matter. They provide on-the-spot feedback and support iterative refinement of student answers. They gather internal timing data for assessment purposes, a source of information that otherwise would require video data or one-on-one human attention. The authoring environment enables people who are not CogSketch developers to create worksheets, if they are willing to learn about the underlying knowledge base and the representation conventions it uses. Finally, the worksheet model is designed to be simple and general-purpose: if the appropriate KB concepts (or close stand-ins) can be found, a worksheet can be created. Generality is due to the use of robust visual processing (especially qualitative spatial representations) and analogical reasoning, so that the system can interpret sketches visually and compare them in human-like ways.

The initial classroom pilot study provided encouraging results. Given the lack of a control group and a balanced intervention, we do not yet have evidence of improved learning outcomes. While not bug-free, the initial experience was sufficiently positive that students wanted more such exercises. Students were able to complete the exercises and did reasonably well on them.

There are a number of directions we are taking in future work. We discuss each in turn.

### More sophisticated coaching

As Figure 5 shows, finding ways to express the subtleties needed for interpreting sketches can be challenging. We think that in this particular case, annotations on the shape itself to indicate portions that must be quantitatively correct could solve the problem, but this is an empirical question. More generally, moving in the direction of tutoring, equipping the system with models of the domain of the worksheet, is a promising direction.

### More classroom studies with more worksheets

We are still in the early stages of formative evaluation, so growing our pool of collaborating instructors is important.

Finding schools that have large enough classes where we can find balanced sections to provide a control group is something we are working on. We also plan on expanding the pool of worksheet users to high school and middle school, and broadening the range of topics to include physical sciences and biology.

### Worksheets as assessment tools

There is already some evidence from a laboratory experiment that worksheets could be useful in automated assessments. Jee et al. (Jee et al, 2009) found that when experts versus novices drew geological processes, or marked up images with geological features, there were distinct and easily recognizable differences between the two groups. In process diagrams, experts tended to include more arrows, which in such diagrams indicate the processes that are occurring and relate different aspects of the cycle to one another, and tend to begin their diagram with such information. For geological features, the experts tend to draw more geologically relevant features, often in an idealized manner. This cannot be attributed to differences in drawing skill, since control photographs (e.g., fruit, lasagna) were indistinguishable across the two groups. Importantly, the same pattern of results holds for sketching from memory, for copying, and tracing. This suggests that comparisons of student sketches in a very simple copying task could be diagnostic of their mental models of the domain.

### Lower entry barriers to worksheet authoring

In the long run, we want domain experts and teachers to be able to author worksheets without AI experts at their elbows. One key barrier is understanding the ontology. We plan to provide a natural language layer for concepts and facts, so that worksheet authors can select concepts and work with facts without knowing predicate calculus nor having an intimate knowledge of the OpenCyc ontology.

## Acknowledgements

This research was supported by the Spatial Intelligence and Learning Center, NSF SLC Grant SBE-0541957.

## References

Adler A., and Davis R. 2004. *Speech and Sketching for Multimodal Design.* In Proceedings of the 9th International Conference on Intelligent User Interfaces.

Alvarado C., Oltmans M., and Davis R. 2002. *A Framework for Multi-domain Sketch Recognition.* In Proceedings of AAAI Spring Symposium on Sketch Understanding.

Falkenhainer, B., Forbus, K., and Gentner, D. 1989. *The Structure-Mapping Engine: Algorithm and Examples* Artificial Intelligence, 41, 1989, pp 1-63.

Forbus, K. and Usher, J. 2002. *Sketching for Knowledge Capture: A Progress Report.* In Proceedings of IUI'02, San Francisco, California.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., and Wetzel, J. 2008. *CogSketch: Open-domain Sketch Understanding for Cognitive Science Research and for Education.* In Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling, Annecy, France.

Gentner D. 1983. *Structure-mapping: A Theoretical Framework for Analogy.* Cognitive Science, 7 (1983), 155-170.

Hammond, T. L. 2007. *A Perceptually-based Language to Simplify Sketch Recognition User Interfaces Development.* MIT PhD Thesis.

Jee, B., Gentner, D., Forbus, K., Sageman, B. and Uttal, D. 2009. *Drawing on Experience: Use of Sketching to Evaluate Knowledge of Spatial Scientific Concepts*. In Proceedings of the 31st Annual Conference of the Cognitive Science Society, Amsterdam, The Netherlands.

Landay J. A., Hong J., Klemmer S., Lin J., and Newman, M. 2002. *Informal PUIs: No Recognition Required*. In Proceedings of AAAI Spring Symposium on Sketch Understanding.

Lee W., De Silva R., Peterson E. J., Calfee R.C., and Stahovich, T. F. 2007. *Newton's Pen – A Pen-based Tutoring System for Statics*. In Proceedings of the Eurographics Workshop on Sketch-Based Interfaces and Modeling.

Lovett, A., Tomai, E., Forbus, K. & Usher, J. 2009. *Solving Geometric Analogy Problems through Two-Stage Analogical Mapping.* Cognitive Science **33**(7) 1192-1231.

Pittman J., Smith I., Cohen P., Oviatt S., and Yang, T. 1996. *Quickset: a Multimodal Interface for Military Simulations.* In Proceedings of the Sixth Conference on Computer-Generated Forces and Behavioral Representation, pp. 217-24.

Wetzel J. and Forbus K. 2009. *Automated Critique of Sketched Mechanisms*. In Proceedings of the 21st Innovative Applications of Artificial Intelligence Conference, Pasadena, California.

De Silva R., Bischel, T. D., Lee W., Peterson E. J., Calfee R.C., and Stahovich, T. F. 2007. *Kirchhoff's Pen: A Pen-based Circuit Analysis Tutor*, In Proceedings of the 4th Eurographics workshop on Sketch-based interfaces and modeling.