Chapter IV

Creating a new problem-solving assessment tool

My goal was to develop an evaluation tool that can obtain a detailed assessment of a student's strengths and weaknesses in technical problem solving. I used a novel approach that avoids the two primary difficulties in evaluating complex problem solving skills. The first is untangling problem solving from the subject content knowledge necessary to solve problems. A student may be a good problem solver but simply does not know the particular area of inquiry addressed in the problem, or a student could be very familiar with the given problem so that "solving" it on a test is merely writing down a memorized response. The second difficulty is that most problem solving assessments can only evaluate the composite result of using the many different skills that are important in solving complex problems. This is inadequate to usefully evaluate and guide instruction. These difficulties have confounded many past attempts to measure or improve problem solving skills in math and science. Most attempts have failed or resulted in a test limited to only a few specific skills or an isolated content area (Hestenes, Wells & Swackhamer, 1992; and Hestenes and Wells, 1992; Cummings, 2005).

I have created an evaluation tool that analyzes students' general problem solving skills. The solver must read through a script that has two people working through a realistic scenario that involves solving a complex problem. The solver responds to a series of questions as they read through the script. This script provides scaffolding that avoids the common difficulty in problem solving evaluations that one weakness will prevent evaluation of other skills that may be subsequently needed.

The series of questions forces the subject to engage in solving the problem within the scenario under the guise of evaluating the other two people who in effect act as group members.

For this dissertation, I have created a version of this instrument that is effective in an interview situation at identifying and evaluating students' problem solving abilities in 44 distinct areas and I created a written version with scoring rubric that evaluates 90% of these 44 skills. These skills cover several categories of mathematical aspects of problem solving, as well as broader categories such as metacognition and motivation. This instrument is suitable for use with middle school students up through college students; but, has been developed and tested with a wide range of adults, including a high school drop out, college students and professionals with Ph.Ds. in science with a very wide range of backgrounds and abilities. These measured category strengths and weaknesses show good correlation with abilities displayed short term, while solving physics problems, and over extended periods of time in a wide variety of other contexts such as class activities throughout a semester and job performance by employees.

Validation studies were performed throughout development in the form of interviews, comparison with teacher and employer evaluations, and two series of rigorous comparisons to problem solving in physics. The first series consisted of comparing written results of the evaluation tool to interviews of students solving an involved introductory mechanics problem. The second series compared results of interviews that used the problem solving instrument to a semester of interviews on quantum mechanics problem solving that were conducted by an independent interviewer. All of these comparisons showed that students have the same strengths and weaknesses in problem solving regardless of the subject matter.

The written version of the instrument is currently designed to be given in a paper and pencil format; however, future plans include creating an online, computer graded version that is capable of accurately measuring the problem solving skills identified in my rubric. Once this on-line version has been satisfactorily refined and validated, a comparable version for post instruction will be created.

This chapter opens with a brief section on the motivation for this study and then introduces the evaluation tool and how it was constructed. The design section is followed by descriptions of validation interviews as well as two rigorous studies that verify the problem solving skills measured by this instrument are the same skills that students use to solve mechanics and quantum mechanics problems. Following the research section will be discussion on the implications of the findings of this study and how these specific problem solving skills can impact further research and teaching of problem solving.

Motivation

What is critically needed is an evaluation tool of specific problem solving skills that does not require content knowledge to evaluate. I have utilized a new approach to evaluating problem solving in an effort to create such an instrument. Over the past couple of years I have carried out sufficient implementation of a problem solving evaluation tool to show that this novel approach has great promise.

This instrument evaluates 44 separate and important problem solving skills that emerged from the design interviews, in a way that does not require any physics content knowledge from the student. In addition these skills are shown to be the same skills used by students, in conjunction with their physics content knowledge, when solving physics problems. This instrument is successful at measuring all 44 skills when used in an interview format. In written (non-interview) format it's capable of measuring 90% of the skills.

This survey is unique not only in avoiding the need for content knowledge but more importantly it unravels the individual skills from the whole. There are many facets to solving problems and students are unsuccessful for many different reasons. Identifying students' strengths and weaknesses before instruction is crucial for effective education. Usually the reasons for a student's lack of success in a course are not understood and instruction is ineffective because it is frequently focused on the wrong issue. This evaluation tool can identify the specific problem solving strengths and weaknesses of the student on the first day; allowing teachers to focus their instruction on the appropriate areas.

A way to measure specific problem solving skills that are broadly useful has implications that go beyond the school environment. These general problem solving skills are the specific tools that are vital to success in the workforce. My preliminary research has confirmed work by Mayer and Heller (Mayer, 2003; Heller & Reif, 1984) that a person's strengths and weaknesses in problem solving define how they tackle all problems in their life whether they are physics problems or problems within the workplace. With this being the case, by measuring and thereby improving these skills through education, the beneficial impact of education will be greatly improved.

Many educators use the term “problem solving” to cover a very broad selection of tasks including answering back of the chapter textbook problems. As mentioned previously, I will use the specific definition: “Problem solving is cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver.” (Mayer, 1992) This requires that the classification of a problem be based on the solvers response to the problem rather than the task itself. If a person is an ‘expert’ in their field, then it is very likely that a task that is a problem for students will be an exercise for the expert. I am specifically addressing how people handle a task that is a *problem* for them. I can then break down the skills used to tackle such a problem into specific items.

Assessment Tool

I have designed an evaluation tool to assess a range of problem solving abilities. It has been written to analyze each specific ability independently of the others. If a student is unable to use certain skills, they will still be able to complete the instrument. This allows me to analyze their skills in all areas of problem solving regardless of particular weaknesses, as long as they are able to read English text at approximately 4^{th} grade level or above. The instrument has been written to focus on the problem solving skills that are useful in a broad range of applications. There are some skills such as content knowledge that may be useful in only one specific discipline. Knowing that the integral of sine is cosine is not a helpful skill when balancing your checkbook. However, there are other skills such as number sense that would be valuable in many types of problems. The literature search revealed that educators and researchers have a wide range of varying opinions about problem solving that has been thoroughly discussed in the literature review contained in Chapter 1. Some will say that skills are only domain specific (Chi, 2006 ; Mayer & Whitrock, 2006) while others say there are general problem solving skills that need to be categorized and better understood by the community (Maloney, 1993; Anderson & Schunn, 2000; Ross, 2007). My hypothesis, which is supported by my preliminary work discussed below, is that there are general skills and that the conclusion that all skills are domain specific is a result of using inadequate analysis tools. The tools used to reach this conclusion have been unable to untangle the content knowledge, which obviously does not transfer, from the individual aspects of problem solving.

The goal of content-independent evaluation is accomplished via the instrument’s unique construction. It consists of a scenario about a wounded eagle. This scenario is followed by a passage that features two interns discussing the solution to the problem of saving the eagle. Nearly all of the specific facts that are required to solve the problem are imbedded within the story, except a few everyday common knowledge facts such as the solver’s own weight. The wounded eagle problem is modeled after the Jasper Woodbury problems developed by the Cognition and Technology Group at Vanderbilt University (CGTV, 1997). The Jasper problems are stories about a character named Jasper Woodbury who encounters some sort of complicated problem in his everyday life (ie. transporting an injured Eagle via an ultra-light plane) and were designed for use in 5th and 6th grade classrooms with groups of students. These are very involved problems with no less than 14 steps to solution including a wide range of plausible solutions. This insures that the solution is not immediately obvious yet it is very solvable. None of my 30 interview subjects has been able to immediately see the solution. My interview pool includes physics faculty and yet these problems are successfully solved by 5th grade students if given enough time. (CGTV; 1997)

The Jasper problems were designed for video; but, I have created a written summary of the story (Figure I) because it is not feasible for instructors to administer a problem solving evaluation on video. Screen shots taken from the video are embedded within the summary to capture many of the visual cues provided by the video. The discussion between the two interns provides a group scaffolding environment which guarantees progress towards a solution. A picture of each intern, with their name, appears next to their individual quotes so that the discussion can be easily followed. Questions have been placed at crucial points throughout the interns' discussion. These questions ask the solver’s opinion on how the conversation and problem solving process between the interns is progressing and ask the solver to provide bits of factual knowledge, planning and procedural ideas. These questions require the solver to engage in solving the problem while the interns act as group members keeping the solution process moving forward. The answers to these questions provide information about the solver’s problem solving skills in 44 separate areas.

The scenario that I created, to motivate the solver to analyze the interns, tells the solver that they are a hiring manager in a company where they are responsible for deciding which one of two excellent summer interns to hire permanently. The purpose of the scenario is to provide scaffolding to the problem so that the solver cannot get stuck, to motivate the solver to listen closely to what the interns are saying and to remove much of the pressure that a solver normally feels when taking a test. Most subjects that have been interviewed get involved in the scenario and behave as if they are providing helpful information by evaluating the interns and rarely exhibit signs that they are conscious of being evaluated themselves.

Design Process

Cross Disciplinary problem solving skills

The design process began with the compilation of problem solving skills from various sources. These included a literature search, interviews with experts in a range of fields, interviews with students solving problems using computer simulations and finally a summary of important skills provided by teachers and employers.

Before beginning this project I taught introductory physics, space flight, science for non-science majors and elementary education courses for several years. These classes were normally very small, with as few as 18 students in a course. This provided the unique opportunity to closely observe individual students throughout multiple semesters and in some instances from course to course. While working with students it became clear that students had certain strengths and weaknesses in the way they approached new problems, regardless of the topic. Based on these observations I formed the hypothesis that students have a set of specific skills that they use to solve problems. A student's ability to use these skills does not vary based on the topic being studied; however, certain skills are more or less important for different topics and types of problems.

Expert interviews were conducted with people who are considered proficient in their field. These included physicists, bio-physicists, biochemists, biologists, chemists, mathematicians, accountants, and business owners. These interviews consisted of asking the interviewee to summarize how they tackle difficult problems in their field and then to follow up with any differences they had observed between necessary skills for different types of problems. This question was particularly informative when talking to biochemists or biophysicists – people who have cross-disciplinary knowledge such as medicine, biology, chemistry, and physics. The results were revealing in that the skills that were indicated as necessary were quite consistent across discipline (e.g. ability to synthesize information) with the only differences stemming from the nature of the specific scientific knowledge or research process necessary in that particular field – i.e. content knowledge.

The next step for compiling problem solving skills consisted of characterizing students that were interviewed for the Physics Education Technology (PhET) project discussed in Chapter 3. During this project I had six students who where interviewed approximately five times throughout the fall semester and four students were interviewed approximately five times during the spring semester about various PhET computer simulations covering a wide range of physics topics. These simulations ranged in content and level of sophistication; but, were always covering a topic of science that the student had not yet received formal instruction on. Consistently occurring strategies, strengths and weakness that each student used throughout the semester were identified using a list of skills that was structured from the problem solving skills compiled from the literature search and discussions with experts. This analysis revealed that each individual student had a consistent routine or set of skills that they used to tackle each of the various simulations. In most cases, the student's routine slowly evolved to become more sophisticated but did not vary in reaction to the content of the simulation. For example, one student was very strong at using abstract information that she had learned in class yet had a hard time seeing connections between her personal experiences and the ideas she learned in class or saw in the simulation.

Finally I had the unique situation of employing three of my former students. This provided an opportunity for a careful comparison of these former students' approaches to their work at their job and their performance on course work. This revealed strengths and weaknesses that were consistent across both environments. The course work included physics problem solving while the workplace required serious multi-tasking of accounting duties, filing and handling customer questions and complaints.

Compilation of the above information provided support for my hypothesis that aside from specific content knowledge, there are many problem solving skills that a person uses in every aspect of their life. In short, they have a set of tools that they use to tackle a problem when they encounter it – whatever environment they are in.

ASSESSMENT TOOL PROBLEM CHOICE

It was important that the assessment task be a problem for a wide range of subjects and did not require the solver to bring in outside content knowledge. To see a person's strengths and weaknesses, they must be actively engaged in the struggle of problem solving. This means the problem that is evaluated must be a problem and not an exercise for all subjects. It also has to be of sufficient difficulty that the solver can make continual progress towards a solution and not get stuck due to a lack of information or extreme difficulty.

Jasper Woodbury Adventures

The Jasper Woodbury series of problems was designed for 5th and 6th grade math students by the Cognition and Technology group at Vanderbilt University's Peabody College of Education and Human Development (CGTV, 1997). These problems were designed for students to solve in groups over several class periods. The Jasper series is the culmination of several years of effort to create enriched learning environments that foster the development of mathematics problem-solving skills.

Studies found that when students solve their first Jasper problems, they have a very hard time determining what to ask. Data shows that as students are given more questions to ask, they have a higher probability that they'll solve the problem. Classes using three or four Jasper adventures over a school year were compared with control classes on several measures. Aggregate pretest scores were equivalent for both groups. Post test data indicate Jasper students performed as well as or better on standardized tests, even though the Jasper classes had spent three or four weeks less on the regular math curriculum. Jasper students also demonstrated superior performance on one-, two, and multi-step word problems. Finally, Jasper students scored much higher on planning and sub goal comprehension problems than their control counterparts. (CGTV, 1997)

Studies also showed that the Jasper problems fit with many different teaching methods. A nine state implementation of the project showed that Jasper problems had a positive impact on standardized math scores, complex problem solving skills and attitudes toward mathematics. Attitudes included self-confidence, math utility, interest in math and feelings about math challenges. In attitude surveys Jasper students showed less anxiety toward mathematics and were more likely to see mathematics as relevant to everyday life. Jasper students were also more likely to appreciate complex challenges (CGTV, 1997).

Adventure at Boone's Meadow

This problem was chosen because it fits the requirement of being a problem for a wide range of subjects as mentioned earlier. It is carefully crafted so that it has a range of solutions and makes the problem accessible to a wide range of people and yet still provides enough of a challenge that it has been a true problem for all subjects. The problem is created in what the CGTV group calls a Natural Learning Environment with an embedded data design. Natural learning environment means the context is a real life scenario that includes activities, subjects, ideas and information that a typical person has encountered in their everyday life. Embedded data design means that all content knowledge is provided. Real life scenarios and embedded data minimize unnecessary cognitive load and provide motivation for the students to solve the problem.

The problem is also more realistic than typical classroom problems in that lots of information is seen and heard by the characters in the story over the course of a few days and then they stumble upon a problem. This leaves the solver wishing they had paid attention from the start. This is a problem solving skill that is often not needed in school but is very valuable in the workplace.

"Rescue at Boone's Meadow"

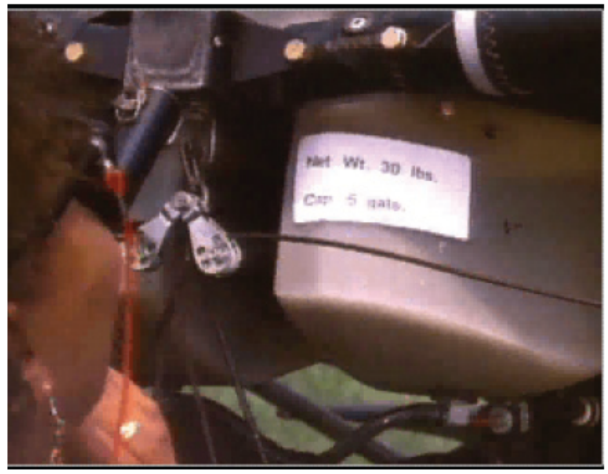
Story Summary

Larry Peterson, a friend of Jasper Woodbury, flies an ultralight plane over Cumberland City. Soon, Larry begins to teach Emily Johnson to fly the ultralight. He gives her some information about the plane: Its total weight is 250 pounds and it can carry a payload of up to 220 pounds. Larry explains that payload is the weight the plane can safely carry in addition to its own weight; payload includes the weight of the pilot, the fuel, and cargo. Emily says she doesn't see any place for cargo. Larry then shows Emily a box used for carrying extra cargo. "Here it is but I hardly ever use it since the box alone weighs ten pounds." Says Larry.

Emily comes closer to the ultralight so she can see as Larry is teaching her. He explains that the propeller does the pushing, just like it does with a boat; the wing does the lifting. He then demonstrates how the unique shape of the wing helps lift the plane. He shows Emily that the air over the top of the wing has further to go than the air under the bottom. He says "This causes the air over the top to move faster since it has further to go. The difference in speed creates a difference in pressure. Here let me show you."

Larry demonstrates with a piece of paper. Blowing over the top of the paper lowers the pressure on the top and the paper lifts into the air. Emily tries and produces the same result.

A few days later Larry teaches Emily about the engine of the ultra light. He tells her that his ultra light's engine was originally used for a snowmobile, so it uses regular fuel and not aviation fuel. Emily points out that one and one-half gallons of fuel are left in each of the two sides of the fuel tank. She asks Larry how far he flew on the two gallons missing from the tank. He tells her that he had filled up the fuel tank in the morning and had flown over to Headlyville and back, which was about 30 miles in total. She asked him how long that took. Larry replied, "My rule of thumb is one mile every two minutes ‹ on a calm day, that is." "Headlyville? They've got an airfield over there?", asks Emily. "Nope, just give me a field 100 yards long to get this little darlin' off the ground and you're in business," says Larry.



A few weeks later Emily takes her first flight. Emily, Larry, and Jasper go out to supper to celebrate. At the restaurant, Jasper talks about his plans for a fishing trip. He says that he plans to drive the 60 miles from Cumberland City to Hilda's Service Station and then hike to his favorite fishing spot, which is about 18 miles or so on foot. Larry mentions that he flew his ultra light to see Hilda the previous week, "I set the ultra light down in that field next door. Good golly, I thought old Hilda and her eyeballs were in for a permanent separation."

For desert Emily orders a scoop of strawberry ice cream in a dish. Larry tries to ask nicely for lemon Jell-O in a sugar cone but the veteran waitress is set in her ways and will have none of it. She says the Jell-O is in the Jell-O section and the sugar cones are in the ice cream section. "Sir in all my years in the restaurant business I've never seen it done that way. I truly never have and I don't expect I ever will!" Jasper sees a better solution than attempting to reason with the waitress. Instead he asks for a strawberry ice cream cone and a dish of lemon Jell-O. When the desserts come, he simply transfers the ice cream to the dish and the Jell-O to the cone. Larry says "Jasper, you're a genius." Emily replies "I second that."

Their bill comes to $17.50. Emily suggests they include a 20% tip for all the grief Larry gave the waitress, and they agree to split the check equally. They each put money down on the table: Jasper puts $11.00 down, Emily puts $12.00, and Larry puts $9.00 down. Larry calculates the total bill and makes change for each of them.

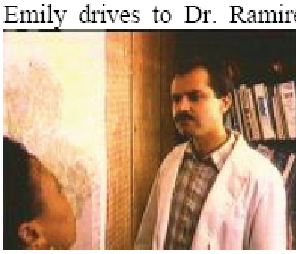
Before leaving the restaurant, Emily and Larry weigh themselves. The scale shows that Larry weighs 180 pounds.

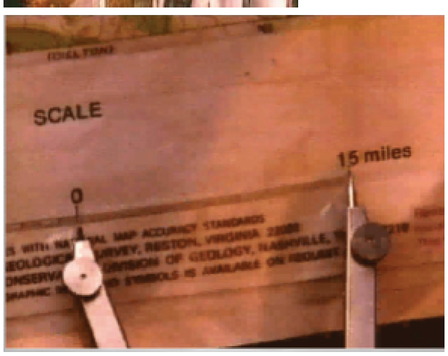
While fishing, Jasper catches a nice looking trout. He fries it over a campfire and as he's taking his first bite, he hears a gunshot. Jasper packs up and goes to investigate. He discovers that an eagle has been shot. After giving first-aid to the eagle, he makes an emergency call to Hilda on his two-way radio.

A customer in a convertible drives up to Hilda's, a gas station along a highway in the middle of nowhere. Hilda is pumping gas for her customer as Jasper radios for help. When Hilda is finished, the gas pump shows that the customer got a total of 13.9 gallons and that gas costs $1.259 per gallon. Her customer records his mileage and tells Hilda that he got 312 miles on his last tank of gas. Hilda says "I'm real happy for ya fella. That'll be $17.50", and he pays for it with a $20.00 bill.

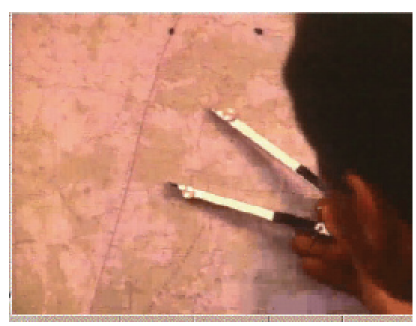
When Hilda answers Jasper's emergency call, Jasper tells her about the wounded eagle and explains that he needs to get it to Dr. Ramirez, a veterinarian in Cumberland City, ASAP. Jasper tells Hilda that he is at Boone's Meadow, which is about a six hour walk from her service station. He asks Hilda "call Emily Johnson and explain the situation, she'll think of something."

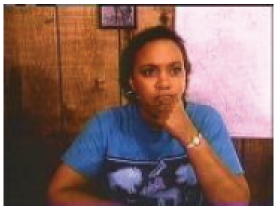
Emily drives to Dr. Ramirez's office. They go into his office where he has a map of the area on his wall. He marks the locations of his office in Cumberland City, Boone's Meadow, and Hilda's. Dr. Ramirez points out that Hilda's is right off the highway and that there are no roads leading into Boone's Meadow. Emily asks how much a bald eagle weighs. Dr. Ramirez estimates that it would weigh about 12 pounds.





On the map, Dr. Ramirez uses a pair of calipers to determine that the distance by air between Boone's Meadow and Cumberland City is about 65 miles. He tells Emily that most planes need about 2,000 feet of runway and Boone's Meadow is just half that long. Before he leaves, Dr. Ramirez tells Emily that the sooner he can treat the eagle, the better chance he has of saving it.





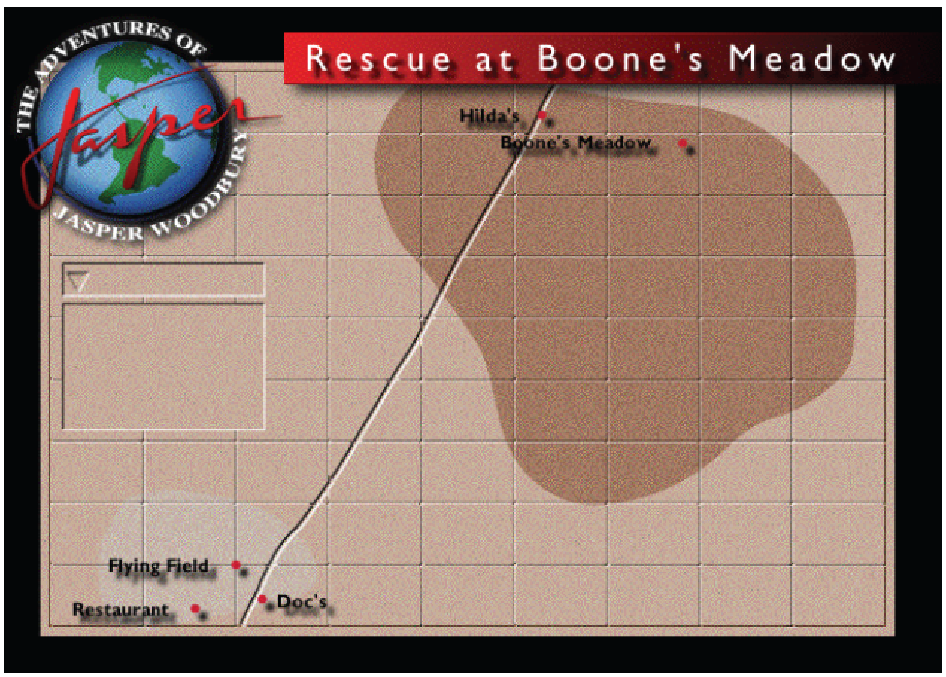
Emily plans for the eagle's rescue. She uses the map to determine the distance by air between Boone's Meadow and Hilda's. Next, she calls Larry, who is just down the road. She learns that Larry is available to fly, that the ultra light is fueled up and ready, and that "the winds are calm as a church on a Monday morning."

Emily thinks about the information she has gathered. She estimates that if the ultra light is used in the rescue, she had better add five minutes for each stop.

Challenge for Interns

Emily wants to know two things:

- The quickest way to move the eagle to Cumberland City?
- And how long will that take?



Please begin the Survey now to follow along with the Interns' problem solving process.

DESIGN, INTERVIEWS AND VALIDATION

The first step to the validation was verifying that the instrument is asking what is intended and that solver responses can be clearly interpreted in only one way. As versions of the evaluation instrument were created, validation interviews were performed. A total of 30 people have been interviewed with the various versions of the instrument. The Jasper story summary was revised slightly whenever there was any confusion by an interview subject. After the first ten or so interviews, the Jasper story summary required no further modifications. The script between the two interns and the questions that are placed throughout the script needed substantial revisions throughout the first 20 or so interviews. After many iterations, the questions within the script are now sufficient to create an interview environment that requires only a brief introduction of think-aloud interviews and no further input from the interviewer. The conversation between the interns and the questions throughout, keep the conversation between the three (interns and the solver -interview subject) flowing without any external input from the observer. Below I will go through the three main stages of the instrument creation process.

Initial Version

The initial version of the evaluation tool was created by writing a brief summary of the video Rescue at Boone's Meadow. The summary included all (almost all) of the useful as well as the distracting facts presented in the video. Next I took screen shots from the video at key points that included necessary and sometimes not so necessary data. These screen shots were carefully placed with the text that directly applied to the scenes. At the end of the summary, the challenge is repeated and a map is provided. This is the same ending that is left for the students on the video. The Jasper story is followed by a script with two students (at this stage the intern scenario had not yet been created) solving the problem presented within the summary.

The first set of interviews included three separate interviews that each took about an hour. That hour of working through the evaluation tool revealed a wealth of information about each person's problem solving abilities. I had known each of the three interview subjects quite well before performing the interviews; yet learned many new things about each during that hour. The setting created by the story summary and the students' (interns') struggles solving the problem created a dialog with the interview subject that revealed specific strengths and weaknesses about these subjects. These strengths and weaknesses correlated with other behaviors that I'd observed over the years but had not fully understood. Now, after one hour with the evaluation instrument, I could describe specifically why these particular people were successful at some tasks and would struggle with others.

These interviews were very effective but it was clear that the instrument had room for improvement. All three interviews required substantial participation by me as the interviewer to keep things moving and to probe the interview subject for specific information. The subjects were focused on solving the Jasper problem and minimized the parts of the script that included the students (interns). The subjects used any facts offered by the students (interns) but expressed annoyance at their involvement in the problem. The subjects also had a hard time keeping track of which student (intern) was talking and who had provided information. In short, the students (interns) were not taking on any sort of personality or importance within the evaluation. Although useful, this version did a mediocre job relative to later versions of revealing the subject's full range of problem solving skills because the students' (intern's) conversation was not successfully scaffolding the solution process and the subject was not carefully analyzing the students (interns); hence, not providing enough information for the interviewer to unravel a lot of the specific skills being utilized by the interview subjects. Additionally the interviewer's role was much too involved.

Second Version

The second version of the instrument introduced the intern scenario. A brief introduction was provided before the Jasper story explaining to the person completing the instrument that they were to play the role of a manager in a company who was responsible for choosing which one of two excellent summer interns would be offered a permanent position with the company. The permanent position would require strong problem solving skills. To assist in the manager's decision, the interns would be asked to solve a very involved problem together.

This scenario creates a context that lifts the focus of evaluation off of the interview subject, who is completing the evaluation instrument, to the interns. It puts the interview subject (solver) into the position of a helper, someone providing a service. It removes the performance goal (Dweck, 1999) and in my interviews the act of evaluating the interns quickly transitioned the interview subjects who began in *performance mode* to *active engagement* (Adams et al; 2008) with the scenario of evaluating the interns. Some took longer than others to engage but all subjects that were interviewed became immersed in the problem within 10 minutes and in most cases much less.

This scenario also allowed me to evaluate skills that are important for individual problem solving and skills used in a group. The interviews were all conducted on an individual basis so each subject was forced to engage in solving a problem and they could not act as a passive observer. The interns act as group members and all interview subjects discussed these two characters as they would real people and reacted to their comments as if they were real. This interaction provided insight into the subject's ability and interest in evaluating and integrating other group member's input and ideas.

Every attempt was made to give each intern a unique, likable, but imperfect personality with characteristics commonly found amongst college students. Two young ladies were chosen to avoid any issues stemming from preconceptions about men's versus women's competence or intelligence. There have been external comments and concerns that one lady is clearly Caucasian and the other appears to be Caucasian, Hispanic or possibly African American depending on the quality of the copy of the evaluation tool. However, during actual interviews, I have not seen a single indication that this has affected the interview subject's (solver's) analysis of the problem or the interns. Interestingly, I have definitely seen indications that the intern's gender brings in many preconceived ideas (some positive and others negative). Fortunately, these ideas have always been applied equally to each intern.

Presumably this is because they are both young ladies of the same age and level of attractiveness.

Pictures of each intern were included beside each comment made by the intern with different fonts used for each intern's words. This structure was provided to make it easy to see who was talking so that minimal cognitive load was spent keeping track of the interns. Questions were written in a third unique font. The questions were followed by plenty of space to write an answer. This addition of pictures and space to answer questions, as the solver progressed, created pages with mostly white space. This structure creates an inviting, non-intimidating environment that is easy to navigate and reduces a solver's cognitive load. When asked to stop and consider which intern they were leaning towards hiring, solvers were able to easily flip back through the previous pages and find the passages they were looking for.

The structure also provides a way to scaffold what is seen by the solver and when. If I wanted a solver to move on to a new idea, I switched pages so that the previous conversations were not in view while presenting the new idea. If there was information or a plan that I was hoping to see emerge from the solver, presentation of the idea or plan was not offered by either intern until after the solver had been prompted and given a chance to produce this fact or idea. Then, on a following page, the item or suitable clue would be introduced, such as a comment by an intern, "I remember Emily said something about 1 ½ gallons gone in each tank and Larry said he'd flown somewhere that morning. Do you remember how far?" At this point a solver's response to the new idea would be elicited by another question to identify how close they were to contributing this idea or if they even felt that it was useful and valid.

Further Interviews and Revisions

Interviews using the second version were much more successful. Interviewees paid substantially more attention to the interns, who were seen as two distinct "people" in the eyes of the interviewee. However, there was some difficulty keeping their names straight and at times what contribution belonged to which intern. This problem was alleviated on a later version by adding each intern's name to their picture – Sara and Jasmine.

Many refinements were made to the evaluation tool as further interviews were performed In the early stages a few changes were necessary to the summary of the story. Small clarifications in wording and a little further explanation of some irrelevant details. For example, in early interviews, some of the interview subjects had tried to work through an irrelevant scene where the characters pay for dinner and divide up the change. Not enough information had been given in my summary to actually determine each person's correct change. This had caused frustration with a couple of solvers who attempted to check this. Data was added from the video so that this mini-problem within the summary could be solved if a person wanted to. I did not want someone to determine information was irrelevant only because I did not include all of it. In later versions, when all information was included, some students have used it to determine each person's correct change after dinner and were not frustrated. Now solvers must deduce from the story that this information is unnecessary, rather than simply basing their decisions on an artificial construct of the evaluation instrument.

Based on comments arising during approximately the next 15 interviews, many additional questions were added to the script to probe progress with the problem solving and at times to facilitate the solution of the problem. These additional questions were recorded during interviews and added to the script in appropriate places. This process of interviewing and adding probing questions was an iterative process that was applied to all versions of the evaluation instrument.

I asked a series of follow-up questions about the subject's experience solving the problem, after the s/he had finished the portion of the script involving the interns. The follow-up questions were designed to accomplish several goals: First, to give the solver a chance to demonstrate other ideas they may have wanted to try but, the direction of the script did not allow for them. Secondly, to probe the subject's ideas about how faster and slower times could have been accomplished (alternate solutions or mistakes). Finally, I asked a series of self-reflective questions that helped me evaluate the subject's metacognitive skills and motivations. These follow-up questions were added in writing to the evaluation tool. After each interview, the written versions were refined until this section of the evaluation tool was also able to take the place of my verbal questioning.

Current Version

The current version of the instrument (Appendix A) typically requires 1 ½ to 2 hours to complete in an interview situation with minimal direction. The subject is told

Table I – Problem Solving Skills

to begin reading and to follow the written directions. Typically a person spends ten minutes reading the brief description of the intern scenario followed by the story about Jasper at Boone's Meadow. At the end of the story the subject starts reading the script between the interns that contains questions about both the story and the intern's conversation. At this time, I typically need to provide a little direction reminding the subject to describe what they were thinking verbally while completing the questions. The job of interviewer then became one purely of observer. All desired information is acquired by simply listening and prompting the subjects to think-aloud if s/he falls silent. The questions within the script and the interns' discussion provide the same scaffolding for the problem solving process that I, as interviewer, had originally provided with questions. The evaluation tool itself elicits all necessary verbalization from the interview subject to identify their strengths and weaknesses.

Problem Solving Categories

Summaries of each subject were written up after each interview. During the first year, the focus was on producing descriptions of each solver which identified their unique problem solving abilities - what their strengths and weaknesses were and what defined their problem solving. I was pretty excited to identify 10 distinct problem solving skills during my first three interviews. During the next series of interviews, each student's strengths and weaknesses were identified and in each case this combination was unique. As these interviews progressed, the list of skills began

Knowledge – have	Beliefs, Expectations & Motivation	Processes – do
Math – basic add/sub/mult/div	Confidence	Acquires Info 1st time through
Math – equation formation	Attribution (takes responsibility for their actions)	Plan ideas (What – ask questions)
Reading comprehension	Judgment of information based on the source	Plan way to get answer (How)
Spatial – mapping	Wants to solve the problem for self	Plan – big picture (Visualization)
Previously known facts	Wants to solve the problem for interviewer	Keep problem framework in mind
Real World knowledge	Wants to succeed on the "test"	Connect steps and pieces
Knowledge of own Strengths	Interested in the context of the problem	Check calculations of others
Knowledge of own Weaknesses	Enjoyed solving the problem	Aware of how others helped
Number Sense	Enjoyed analyzing interns	Meta-process – step outside of problem solving to see if own actions are useful.
Estimation	Enjoyed complete experience	Skepticism
Ability to analyze interns	Real life vs. student	Estimation
	Careful/Thorough	Creativity
		Adaptability (shifts direction easily)
		Can throw out useless info
		Judgment of reasonable issues
		Judgment of importance of number values (is it material)
		Tie in personal experiences
		Tie in info provided by another
		Scientific Process (each step justified with evidence not by gut feeling)
		Remember previously noted facts
		Remember what s/he has calculated or reasoned.

For a description specific behavior associated with each category see Rubric in Appendix B

to converge. In the end, I found 44 items are required to fully describe the problem solving process involved when solving the problem contained in the evaluation tool.

During all interviews, I wrote down skills as they emerged, either as a strength or a weakness. I did not have a list of skills in front of me that I attempted to score with each student. This was done for two reasons. First, I did not want to artificially limit the skills that I could observe. The motivation for the second reason was more practical, not having a list of skills in front of me allowed my complete focus to be on observation of the subject rather than trying to watch for the skills on a list in front of me. Recording skills, as they became evident, had the additional benefit of providing interesting data about at what point in the solution process the various skills could be confidently scored. This data is useful because I found that certain skills, such as ‘keeping track of information’, took over an hour to score.

I have interviewed a total of 30 people while working through the evaluation tool and analyzed an additional 16 people through their written responses. The problem solving skills used by interview subjects were compiled and later a grading rubric (Appendix B) for identifying and rating these skills was created. The interview subjects have a broad range of backgrounds including physics PhD’s, physics students, business owners, high school drop outs, elementary education majors, humanities graduates etc… This range of subjects has provided a wide variety of approaches to solving the problem. This has helped clarify the rubric since the observation of weaknesses has been one of the easiest ways to identify specific skills that are important for solving problems. This set consists of 44 distinct skills that completely describe a person’s behavior while solving the problem within the evaluation tool. In subsequent tests, these 44 skills were also found to be adequate descriptors of the solving processes used while solving a mechanics problem and quantum mechanics problems with the exception of specific content knowledge such as geometry.

In Table 1, I include all 44 distinct items necessary to solve the problem contained in the evaluation tool divided into the three categories described in my literature review – 1. Knowledge (have), 2. Beliefs, Expectations and Motivation, and 3. Processes (do). My results have revealed many identifiable distinct items within these broad divisions with no indication for the necessity of a hierarchy of skills. The interview subjects have had different combinations of strengths and weaknesses in each area without any particular items appearing to require strength in another as a basis.

Note there are many more items listed under Processes than Knowledge or Beliefs. This result was not unexpected. The evaluation instrument was designed to probe how students solve problems rather than to probe the specific knowledge they have. Most evaluations in physics probe the knowledge students have rather than what they do when solving problems. This may not be intentional but it does make for incomplete evaluation of student's skills. I feel that the list of items under Processes is adequate for describing problem solving in many different areas; however, the listing of Knowledge and Beliefs may not be complete for certain types of problems. It is easy to imagine a wide range of knowledge that has not been listed. This was clear when evaluating students undertaking the mechanics problem (ie. geometry) or quantum mechanics (i.e. Eigenstates). The Beliefs, Expectations and Motivation category is much closer than the knowledge category to being complete; but, also has other aspects depending on the specific problem. A student who is solving a physics (or math, or biology etc…) problem will likely bring with them ideas about what practices are important and/or necessary for solving a physics problem. Some of these ideas would not apply to the Jasper story.

Some of the items in Table I, such as ‘Enjoyed analyzing the interns’, are not commonly thought of as traditional problem solving skills. My criteria for determining whether something is an skill, is: Anything that can affect the subject’s ability to solve the problem. In the earlier stages of creating the listing of skills, there were items that I found useful for my internal tracking of students but did not necessarily sound like problem solving skills. For example, ‘Wants to find the best solution for self’. I did not pass these few items on to the independent interviewer who scored the quantum mechanics interviews. However, after studying the rubric and beginning to score her students, the interviewer independently asked me if I had a way to describe certain characteristics that she had seen as important but did not seem to be in the rubric I initially provided her. It turned out that all of the characteristics that she was concerned about were in my own internal tracking system; but, I had not included them, because they didn’t sound like skills. Because these characteristics, that I found important, were necessary for me and for the quantum mechanics interviewer to fully describe a student, I refined my definition of skill to include anything that the subject *has* or *does* that can affect the subject’s solution.

Each problem solving skill has be broken down into individual component skills necessary to solve this particular problem. When used for evaluating quantum mechanics problem solving, each skill was gradable using the definition given in the rubric and did not appear to need broken into smaller pieces. However, one could imagine that a problem with the focus of identifying more detail about a person's ability in a specific area such as their ability to analyze others may require more specific component skills that would make up the overall skill identified here as 'ability to analyze the interns'. With most of the skills, this would not be the case but there may be some that this problem does not completely break out.

Written version

The next step in the design and validation process was to create a version of the evaluation tool that can be administered without an interview. A total of 16 written responses have been analyzed. Seven written responses were compared to independent instructor evaluation of the student's skills and nine were compared to my interactions with the students. When the current version is graded using the listing of skills from the interviews, I can confidently evaluate the subject in 90% of problem solving skills listed in Table I based only on their written responses. The written version takes people an hour to complete.

The first round of written only evaluation was undertaken at the end of the semester in a capstone course for Elementary Education majors. I had the good fortune of being able to teach two sections of the course that had 30 students enrolled per section. This course was taught with interactive engagement and most class periods involved group problem solving so that I had a fair amount of time getting to know each student’s skills over the course of the semester. In addition to interviewing seven students using the evaluation instrument, written responses without an accompanying interview were acquired for eight students. These written responses were evaluated using the same listing of skills that were used for the interviews that, at that time, consisted of 21 distinct problem solving skills. When evaluating the written responses, the name of the student was kept separately so that I could grade objectively. These results were then compared to the students’ performance in the course, throughout the semester. While the interviews provided a rating of the students’ strengths in 100% of the problem solving skills, the written results provided measures for 75% of the skills. The skills that could be evaluated, were done so less accurately than with interviews. With written results, it is possible to identify strengths and weaknesses in most categories but with limited precision. Both interview and written results agreed with the evaluation of each student’s course performance.

During the evaluation of the written version, the initial version of the grading rubric for the evaluation tool was developed. This initial rubric identified the specific questions that allow the evaluator to grade the subject on the various items listed in Table 1. With each version of the instrument, this rubric underwent further refinement.

When looking at data from the written version only, it was clear that many important skills could not be evaluated from just the subject’s written responses. The think-aloud format interviews provided a lot more insight into the students problem solving process. During a few more iterations, more questions were added to the evaluation tool in places where this information was imperative and more written responses were collected. The data from the most recent set of responses was more complete and I was able to grade nearly all of the 36 identifiable skills (at that time). This result depended on the student and how much they wrote down. Out of five written responses I was able to grade 89% to 97% of the categories for four of the subjects; however, with the fifth subject, I was only able to grade 81% of the categories.

The skills that were not measured for these subjects varied with the exception of metacognitive processing. This is a very difficult or impossible skill to grade on the written only version. Analysis of metacognitive processing requires specific leading questions. Without these questions, it is difficult to observe metacognitive skills in an interview and they are completely invisible with written data. The other categories that were difficult to score varied. Every once in awhile a subject gave a response that could indicate a couple different possible reasons. Without hearing the subject's stream of consciousness, it's not always possible to correctly identify which reason caused the answer.

Further Data Collection and Validation

Comparisons with students' performance in various environments were made to validate the skills that the evaluation tool measures. Since the goal of the evaluation tool is to untangle general problem solving skills from specific content skills, it does not contain any physics. For this reason it is necessary to verify that the skills being measured by this content free evaluation tool are the same skills that a person uses when solving physics problems. To accomplish this goal a series of studies were undertaken comparing results of the evaluation tool to problem solving in various environments and different types of physics problems. Rigor of these studies increased as each successfully demonstrated that the evaluation tool consistently identified the same strengths and weaknesses in an individual as other measures such as course work, the work place, solving mechanics problems and finally solving quantum mechanics problems.

INITIAL CONCURRENT VALIDATION STUDIES

Initial studies involved comparing a characteristic summary of a subject's strengths and weaknesses to the results of an interview with the subject working through the evaluation tool. These studies were done with earlier versions of the evaluation tool as it was being refined (described in Design, Interviews and Face validation section above) so the complete list of 44 skills had not yet been identified. The subject's characteristic summary from their interview was compared to either teacher and/or employer evaluations of the subject. In each case I found total agreement amongst interviewer evaluations and those of the teacher and/or employer

Comparison of interview results to course performance

Interviews were conducted with fifteen different physics students ranging from elementary education majors, both sophomore and senior level, to physics majors. Some of these students were in my courses and others were not. In the cases where the students were not in my courses, I would interview the students using the evaluation tool and create a summary of their strengths and weaknesses and characterize their overall problem solving ability. The faculty who had these students in their course independently described their observations of the student and success in the course. All students were enrolled at the University of Northern Colorado where classes are small. I worked with two different professors and in both cases they worked closely with their students and they graded all the homework and exams. This allowed them to become quite familiar with the individual students by the end of a semester. The two sets of results were then compared. In each case there was complete agreement between their descriptions of student strengths and weaknesses and what I measured with my assessment interview.

Quite often the detail from the results of my 1 ½ hour interview helped the faculty understand the behavior they'd been seeing with the student throughout the semester. As an example, I interviewed a student who had very strong math and number skills and thought of a lot of details that may affect the solution. He remembered all the details that he'd acquired and numbers he'd calculated, never getting lost in his solution. He was also very critical of the interns. However, he had very weak reading skills, was not actually skeptical of the information offered by the interns (he took what they said for granted and never checked), and was unable to identify which details were important to consider for the problem and which would make very little difference. The student was aware that he was weak in some areas of problem solving so added extra cushion to his answers just in case he'd made a mistake. The professor had identified the student as weak but knew he had strong math skills and was very conscientious. He knew the student worked hard and appeared skeptical in group environments yet never thought through what was offered by others, and so was often being led astray. The one feature the professor had not identified was the student's weak reading ability. The professor commented that this knowledge could have helped him address the student's weaknesses. Also most of what the professor knew of the student took the better part of a semester to identify. If he'd known at the beginning of the semester where the student would struggle, he could have addressed these issues early on.

As mentioned all interview results matched up to faculty assessment of the students as well as my observations of my own students. I had the same experience where the interview results helped me understand why my students had struggled and succeeded in various activities throughout the semester. My course involved many group activities and covered topics in physics, earth science, chemistry and biology. The students exhibited the same behaviors regardless of the topic. However, depending on the type of activity, the same student may struggle or work through with ease. As an example, one young lady (Janet) would always do quite well in almost any group activity, especially those involving calculations and/or working through data. She was a much sought-after group member. When interacting with her groups, it was clear that she knew how they'd arrived at their solution while the other students were still working on understanding all the detail to their solution. However, when working on an individual project involving the synthesis of ideas between two books, Janet struggled and came in for help several times a week. I was quite surprised by this considering her performance throughout the semester. At the end of the semester I interviewed her with the evaluation tool and found that she could not come up with a question. She literally struggled for half an hour trying to organize the ideas in the story enough to come up with one thing that she'd like to find. She would start a sentence or begin to write something down and then stop. After 30 minutes I asked her what she was thinking and she said she was confused, it was just too much and she didn't know where to start. I encouraged her to move past the 1st question to see what the interns would do. From there she went through the script answering the questions easily. If the interns had any ideas about what might be useful, Janet was able to figure out how to find that information and then to actually carry out whatever calculation needed to be done to find it. Throughout the interview she demonstrated an inability to ask her own questions – what might be needed for the next step in the solution; but, was extremely good at finding the answers to any question posed and evaluating its usefulness. Hence her very strong group skills as displayed in class.

It is often claimed that elementary education majors struggle with science and/or math because they have low confidence in their abilities to do science or math. With this set of eight that I interviewed, there were only two who had low confidence in their math/science abilities. In fact, there was not a general set of characteristics that described this group. Janet was the only student without the ability to form a question out of both groups - the students I interviewed and those who completed the evaluation tool in written format, without an interview. Other weaknesses that were observed from this group of elementary education majors included: A couple with weak reading ability; Some had limited number sense or ability to do basic math (yet were still confident in themselves); and, Others did things superficially while not paying careful attention to information or what they had previously done. This lack of processing (thinking carefully) was actually more common than any other weakness. These students would go through information quickly – but not in a hurry - without attempting to organize or make sense of what they were doing. This was not in an effort to finish quickly but rather an avoidance of tasks based on difficulty. They were more than willing to take the time to do the simple work in each problem. This lack of acquiring the information that was needed, created an impossible barrier. In the course this was also quite evident. These students were willing to do projects or extra credit to make up for a lack of performance on exams because they avoided anything that required cognitive processing – assimilation of information. One other common characteristic of these students is that they are content with their abilities and were not interested in working to strengthen their weaknesses. Janet was an incredibly valuable group member, because the other elementary education majors were not particularly strong problem solvers, but were usually good at asking questions. Even though she was quite weak when given a problem that does not have a distinct question posed, Janet graduated with straight A's in a major that is taught almost 100% in group format.

In the future, I would like to give the students the evaluation tool at the beginning of the semester so that I can tailor my course accordingly. In many cases, after the interview I've told the students where they are weak and strong and some feel that they'll be able to use this information to improve their skills. An interesting further research study would be to use this tool to evaluate students' strengths and weaknesses and then follow up later to see which students, if any, are able to assimilate this information from such an evaluation to improve their skills.

Comparison of interview results to performance in the work place.

I also had the opportunity to interview six people with the evaluation instrument that work for my property management company. The results of these comparisons were quite similar to those described in the course comparison section above. The interview produced information about the subjects' strengths and weaknesses that took me months, sometimes years to figure out while working with these people. Their weaknesses were especially apparent during the interview with the evaluation tool. Weaknesses the subjects had learned to overcome with other skills or often explained away or covered with various reasoning (excuses) making it difficult to identify during the course of a days work. I definitely was not able to discern these weaknesses during a typical job interview.

By observing these people in the work environment and evaluating their skills, I also saw that certain skills can be weak and the person can still be quite effective at their job. As with the different activities within a course, certain skills are more important for different types of problems. For example, some of the problems encountered at work involve keeping track of many varied, small tasks that need to be completed. Thinking about everything that needs to be done and keeping track of how they relate, makes completion of the tasks go faster and makes it less likely to forget something. Other responsibilities include balancing a checkbook or figuring out why a tenant has a balance on their rental account. These tasks require some of the same skills such as keeping track of what you've already done but also require math skills that are not useful for the first job responsibility described above. So an employee can do the first task well, but fail in the second.

My most interesting subject to date is a high school drop out who owns his own plumbing business - Tony. In addition to his plumbing business, he helped answer phones and show houses for a year but refused to do anything in the office. By limiting what he did, he appeared to be a very strong all around problem solver. He continually demonstrated an ability to handle many tasks at once and to quickly and accurately trouble shoot problems with properties. In his business, he is known as the best at what he does. He has the uncommon ability to accurately visualize how the plumbing is run throughout a house and underground to the street. He has a very solid understanding of water flow. When evaluated by the problem solving interview, Tony, was very strong in many areas and extremely weak in others. He demonstrated outstanding spatial ability, number sense, creativity, planning (what and how), self reflection and ability to analyze the interns. However, he had extremely low self confidence, had weak math skills, weak skepticism and poor adaptability. The most interesting behavior was that he was completely unable to confine the problem to the abstract framework set up by the story. He continued to treat it as a real scenario and would forget the simplifications added to problem such as no wind. He did not like the idea of Emily flying because she may know how but certainly didn't have the experience for this sort of trip. (Not necessarily a weakness in the real world.)

I also had the unique situation of hiring one of my past students – Buffy. During the course, she got an A on the first exam and then low C's on the next two and a B on the final. It was a large course – 175 students, so I did not have personal interaction with her. In fact, when she came for the job interview, I knew the name but had no idea what she looked like. Almost all of Buffy's skills were quite strong. She had very strong knowledge and processing skills. However, her attribution (who's to blame for failure) was the worst I'd ever seen. Buffy appeared to have supreme confidence in everything that she did. She was quite good at most things that she tried. However, once she made an error, or had to work harder than usual, she would latch on to whatever outside problem may exist. Either someone was being noisy or she was late etc… When this would happen, she would lose track of everything that she was doing. She'd continue to try but once she got worked up, she quit processing and just floundered. She also had quite weak metacognitive knowledge and metacognitive processing abilities. These weaknesses allowed this unproductive behavior to continue unnoticed by her. The following two questions are right next to each other on the evaluation tool and explicate her weak beliefs and metacognitive skills. 1. Did you personally do a good job of solving the problem? "YES" 2. Do you think there's a better solution? "Probably, I didn't really come up with a solution."

RIGOROUS COMPARISON WHILE SOLVING A MECHANICS PROBLEM

The first rigorous comparison to physics problem solving was studied using an involved mechanics problem that came from Maryland's Alternative Homework

Assignments (AHA) (University of Maryland Physics Education Research Group, 2002). This problem requires the use of Newton's Laws as well as conservation of energy. It takes a couple of hours to solve and is designed to require a fuller range of skills than typical back of the chapter text-book problems. Five students attended two or three one-hour interviews where they solved the mechanics problem. Results from these interviews were then compared to their anonymous results from the written form of the evaluation tool. All five subjects were easily identifiable from their written results.

Study Design

Five student volunteers, two men and three women, were recruited from a lab that contained both algebra based and calculus based introductory physics students who were currently enrolled in the 2nd semester of the two semester series. These students and I had never met before the day I went into their lab to recruit them. The student volunteers were asked to participate in a series of physics problem solving interviews. The students were each given a copy of the written evaluation instrument and asked to complete it on their own and bring it to their first interview. When the students brought in their written evaluation tool, an identifier was added so that it could be matched up to the student later. The students then participated in a series of interviews where they worked through the AHA problem about building the Pyramid of Giza. Most students were able to complete the problem and a few follow up questions during two one-hour interviews. However, one of the students did not return for her second interview and another required a third interview to complete the problem. During the interviews, students were scored on their problem solving skills that were used while solving the physic problem.

Problem

The Pyramid of Giza problem (Appendix C) was written by the Physics Education Group at the University of Maryland. These problems were written to address many problem solving skills that are not required while solving typical back of the chapter textbook problems. The group describes the problems as follows:

> “Alternative Homework Assignments (AHA's) are a new approach to introductory physics homework. Traditional textbook problems generally fail to develop students' observational and mathematical modeling skills. Such problems are frequently abstract and do not require the students to think critically about how the concepts are related to the mathematical equations required to solve the problem. AHAs attempt to address this difficulty by combining conceptual, reasoning, and traditional problem solving elements with observational, "equation reading," and dynamic modeling elements. Each assignment includes a series of questions about a single context or situation. Typically only one of these assignments would be given between class periods, and no single assignment attempts to embody all of the elements mentioned above.”

These problems are written with a series of questions designed with the idea that they could guide the students through the modeling. I removed these questions since they limited the number of skills students must use to solve the problem and made it quite difficult to score some of the other skills. The structure provided by the questions removed the need to plan ideas (figure out what the solver needs to find), connect steps and pieces, and to keep the problem framework in mind. Other skills would have been quite difficult to observe but might still be used by a solver including metacognitive skills, visualization, judgment of reasonable issues and creativity.

Without having observed students using this structure it's hard to say what else may have been affected.

Interview Protocol

When the students first arrived to the interview, discussion was purposefully limited. Students were advised about the consent form and that they would be video taped. Their written evaluation tool was collected and filed. Students were then handed the pyramid of Giza problem, some blank sheets of paper, and a pencil. They were asked to work on the problem while thinking aloud. I asked them to tell me what they were thinking about as they wrote things down or contemplated their next step. I told them I was willing to provide any physics information that they requested; however, I could not offer any solution ideas. I was only able to provide information such as the equation for work or to show them how to find the component of gravity along a ramp etc… Partway through the interview students would pull out their calculator and one borrowed mine. If students progressed to a point where they considered the problem complete but it was not, I asked them to reread the question to make sure they had answered everything. All students found that they had not and began working on the next section. After an hour of working, all students were exhausted and their productivity was visibly reduced. They continued to work on the solution at a second interview two weeks later. At the end of the last interview, students were asked a few follow up questions about their solving process.

Pyramid Interview Protocol

Day 1:
Do not ‘break the ice’
Solve for number of blocks
Begin Working on number of men problem

Day 2:
Finish number of men to build the pyramid
Ask if the problem was a realistic situation.
Are there more variables that would affect the real answer? (fishing to see if students think of quarrying the rocks, planning the structure etc…)
Follow up:
What was your favorite part
Least favorite
Most fun
What are your strengths/weaknesses
Did you have a plan?
Did it evolve as you went based on results?
What did you think of the online/written survey you brought in with you?

Results

When interviewing the students, I did not use a set list of skills, rather I used a similar protocol to the evaluation tool interviews where I noted problem solving strengths and weaknesses as they surfaced during the interviews. Below is a listing of the problem solving categories that are scored during interviews with the problem solving evaluation tool but that were not scored during the interviews using the pyramid problem.

- ♦ Reading Comprehension – Four of five students did not read through the story completely. The story for the Pyramid was also only one page and read more like instructions than a story.
- ♦ Acquires Information 1st time through – Because students didn’t read all the way through, this could not be scored.

- Outside knowledge – Pyramid problem did not provide a good setting for bringing in outside facts to the problem.
- Keeps problem framework in mind – I was only able to score two of the five Pyramid subjects in this category. Without solving the problem from beginning to end in one sitting, it's really difficult to score this category.
- Aware of how girls helped – No interns in Pyramid problem.
- Skepticism – Information was not provided by an authority so there was no opportunity to demonstrate this skill.
- Adaptability – There was no change of direction offered since no scaffolding was provided so there was no opportunity to demonstrate this skill.
- Can throw out useless information – When students read the pyramid problem the 1st time, they scan it and start working. Only one student actually read through it. In each case, they went and retrieved facts as needed. With this solving method there were no indications that any of the students noticed that excess information was provided removing any opportunity to observe their ability to filter information.
- Ties in info from other people – With the pyramid problem the only other person available to offer information was me as the interviewer. Two of five students were scored in this category because they asked for a fair amount of information from me.
- Scientific Process (each step justified by evidence rather than a gut feeling) – This category never surfaced while observing the pyramid interviews.
- Enjoyed Analyzing Interns - No interns in Pyramid problem.

- ♦ Enjoyed complete survey – This category becomes the same as Enjoyed the problem since there is no additional idea of enjoying analyzing the interns with the Pyramid problem.

There were also a few categories that were needed to describe a students' problem solving during the Pyramid interviews that were not used with the evaluation tool.

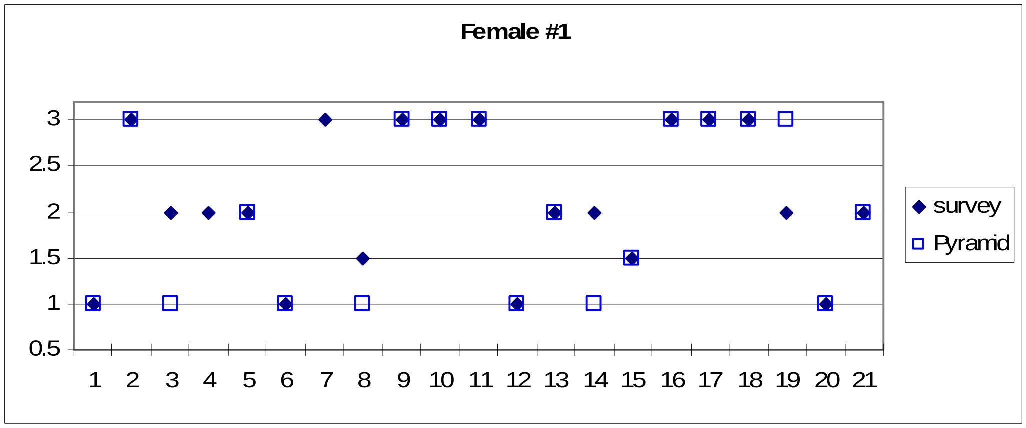
- ♦ Geometry – Very important for the Pyramid interviews and definitely different than other math skills used including algebra.
- ♦ Rounding off numbers – There were lots of calculations and some students rounded numbers prematurely and others went to great lengths to keep a needless number of digits.
- ♦ What should know inhibits effective solving – With the Pyramid problem students quite often felt they should know a simple formula for determining the number of blocks in the pyramid and spent their time futilely searching for the illusive formula rather than just working through a straightforward but tedious calculation.

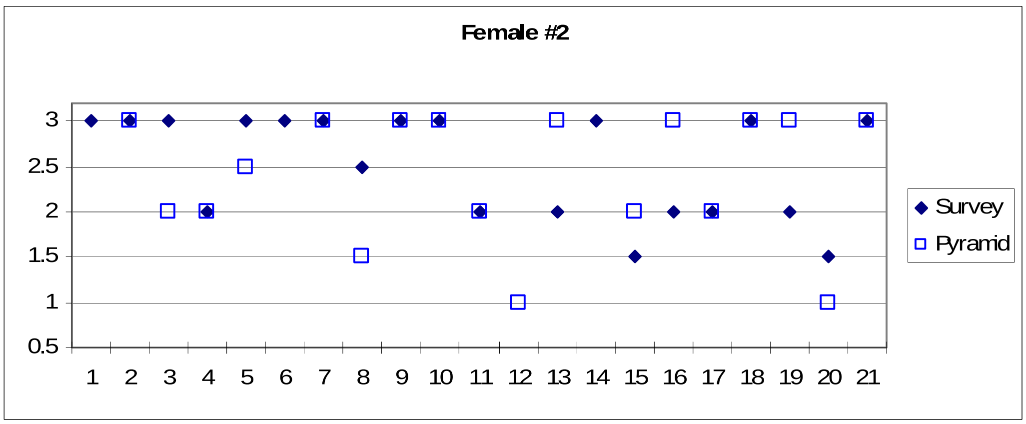
After all interviews were complete, the written problem solving assessment tool responses were graded without identifying information. Grading took place nearly two months after all interviews were complete. This provided an additional level of control since it's very difficult to remember details of student's skills after this amount of time. The results from the written responses were then compared to the results of the Pyramid interviews. The five sets of data were immediately matched up without question. The strengths and weaknesses of each student were consistent across problem situations. The identifiers were checked and the written results were correctly matched to the Pyramid results.

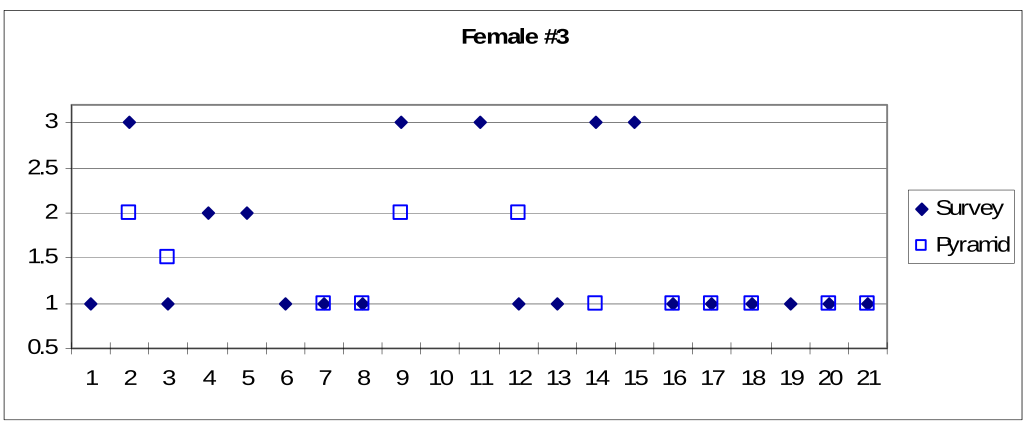
Figure II shows 21 categories that were scored for both the Pyramid interviews and the written results of the evaluation instrument. The scores are displayed for each subject in the graphs below. The written version is scored on a 3-point scale because there is not enough detail in student answers to use a 5-point scale. Pyramid interview results were collapsed to a 3-point scale for comparison purposes.

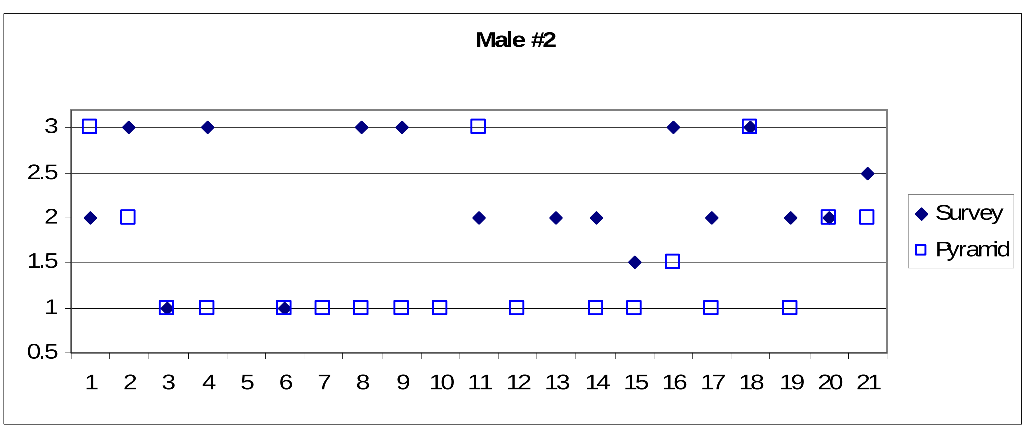
Figure II – Common categories scored for both the Evaluation Tool (survey) and the Interviews.

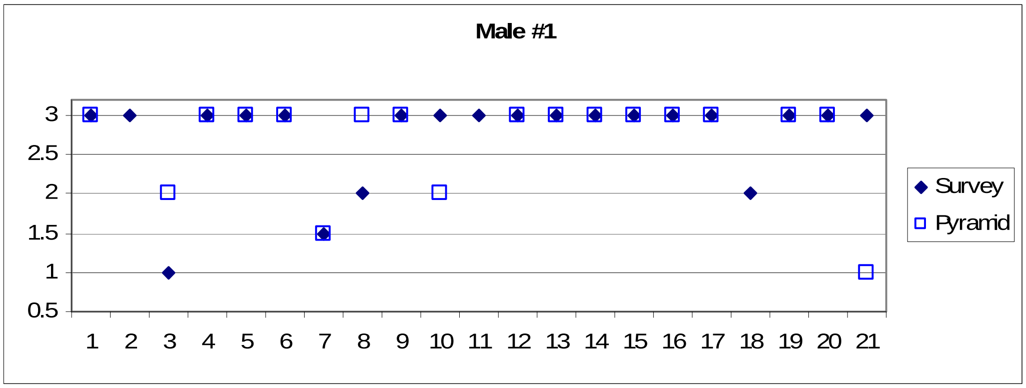
1. Real World Experiences
2. Math – Basic add/sub/mult/div
3. Spatial – Mapping
4. Estimation
5. Number Sense
6. Ties in Personal Experiences
7. Planning What – Question formation
8. Planning Big Picture
9. Planning How – Way to get answers
10. Connects steps and pieces
11. Checks calculations
12. Monitors own progress.
13. Knowledge of own strengths and weaknesses
14. Creativity
15. Judgment of reasonable issues
16. Confidence
17. Enjoyed solving the problem
18. Wanted to succeed on ‘test’.
19. Attribution (takes responsibility)
20. Real life vs. Student (1 = Student, 3 = Real life)
21. Overall Score


Female #1
3
2.5
2
1.5
1
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
survey
Pyramid


Female #2
3
2.5
2
1.5
1
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Survey
Pyramid


Female #3
3
2.5
2
1.5
1
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Survey
Pyramid


Male #2
3
2.5
2
1.5
1
0.5
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Survey
Pyramid



Evaluation tool (survey) was graded from written data only. Pyramid problem data acquired from 2-3 interviews.

This study goes along way towards validating the content free evaluation tool and demonstrating that a majority of the skills measured by this evaluation tool are indeed the same skills used by a person solving physics problems. When a person is weak at a particular skill, they are weak no matter what the context. This study was designed to keep the interviewer from imposing ideas from the pyramid problem interviews onto the grading of the evaluation tool by keeping all results anonymous; however, using only written results of the evaluation tool created a slight weakness in the study. Grading the written evaluation tool provides scores for approximately 90% of the categories with about 2/3 of the accuracy, at best, in comparison to an interview with the same instrument. This study would have been much stronger if the problem solving skills had been determined by an interview rather than through the grading of written results only. However, with only one researcher, it would be impossible to have a blind comparison with interviews on both the pyramid problem and the evaluation tool. Another difficulty lies with the structure of the pyramid problem. It did not require many of the problem solving skills that are required to solve the realistic problem contained in the evaluation tool. For these reasons, another, more rigorous, study was undertaken with a second researcher conducting one set of the interviews.

RIGOROUS COMPARISON WHILE SOLVING QUANTUM MECHANICS PROBLEMS

Sam McKagan, a Ph.D. in physics doing a post-doc in physics education research spent a semester interviewing students while they solved quantum mechanics problems relating to the material in a course that they were currently enrolled. After the semester was over I approached Sam about grading her students' ability to solve problems using the 36 skills that I had identified during my previous studies. In addition I interviewed these students during one 2-hour session using the content-free evaluation tool. We each independently scored the students and the results were compared. This comparison showed that skills matched in nearly all categories as discussed below.

Study Design

Sam interviewed six students during the semester. Of these six I was able to interview five (the sixth moved to New Zealand). The students who participated in both types of interviews were three male and two female physics and engineering majors. During the semester these students each met with Sam eleven times for half an hour each week. After the semester was over, she contacted her students and asked if they'd be interested in participating in an interview with me. After following up with them a few times I was able to coordinate an interview with five of the six students within the year.

Interview Protocol

Quantum Mechanics Interviews were designed to learn about the students in the course and how they understood quantum mechanics. The first interview consisted of a series of questions designed to learn about the student's background, future plans, how they view their own learning, reasons for taking the course and reflections on a particular content question from that week's lecture. The remaining ten weeks consisted of a short bit of reflection on that week's class and then specific conceptual questions relating to that week's material. The following is an example of a mid semester interview protocol.

<u>McKagan Quantum Mechanics Interview Protocol – 2/5/06:</u>

1. Do you feel that you have a solid understanding of what light is and how it works?
2. Explain what you know about light and how it works.
3. `What aspects of light are you still confused about?
4. How has your understanding of light changed since starting this class? How would you have described light before this class?
5. Work through question 4 on homework 4 and explain your reasoning.
6. Suppose you are shooting photons at a screen one at a time and you see a dot appear on the screen as in the picture below. Where was that photon the instant before it hit the screen?

7. What does the blob in the Quantum Wave Interference simulation represent?
8. What is your understanding of the wave-particle duality?
9. When does light behave like a wave and when does it behave like a particle?
10. What evidence do we have that light behaves like a wave, and what evidence do we have that it behaves like a particle?
11. Do you have any specific questions about the discussion of photons in the lectures on Wednesday and Friday?
12. Do you remember any of the questions that other students asked during these lectures?
13. Can you remember one question that another student asked that you knew the answer to? Tell me the question and the answer.
14. Can you remember one question that another student asked that you didn't know the answer to? Did you understand the response from the professor? Tell me the question and the answer.
15. One question that came up several times in lecture was the question of whether a photon is really spread out in space or actually has a definite position that we just don't know about. Before discussing the answer, can you

tell me whether you understand the question? What is the fundamental issue that this question is getting at?

16. How would you answer this question?

As mentioned earlier, Sam was not shown the listing of 36 skills until after she had completed her interviews. This meant she was unable to ask any follow up questions with her students to help determine scores in any of the categories. She was limited to the interviews she had. Upon careful review of the video tapes, Sam felt confident scoring the students in 31 separate categories.

When I interviewed these same students with the content-free evaluation tool, I structured my interviews in the same way as past evaluation tool interviews. I briefly covered the consent form and then immediately handed the students the evaluation tool. I tried very hard not to discuss anything with the students before they began working with the problem solving instrument so that I would not gain any information about them outside of their responses to the instrument. This meant that the beginning of the interviews was a bit awkward; but, I just tried to be relaxed and friendly without saying anymore than was absolutely necessary.

I listened and took notes while the students worked through the instrument. I did not have the listing of 36 skills in front of me, rather as necessary skills surfaced, I scored them, revising if needed. However, immediately after the interview, I sat down with the complete set of skills and scored all categories that I felt I had acquired enough evidence from the interview to score accurately.

Rubric Development

The use of the problem solving categories by an independent party required a detailed grading rubric that clearly explains what is meant by a 1 through 5 score in each category. This rubric can be found in Appendix B. I began with the initial version of the grading rubric that I'd previously developed, that before these interviews consisted of the categories and a very short definition of each. This exercise helped clarify a few categories that had been somewhat imprecise in their previous use. As an example, I had a category called Planning – Big Picture. After creating a written description of its definition, several discussions about how to grade this category and review of previous notes on scoring, I realized that it really should be two separate categories - those of visualization and keeping the problem framework in mind. There were also a few categories relating to the student's motivation to solve the problem that were necessary to completely describe their behavior. When the exercise of creating the rubric and training Sam how to use it was complete, the list of skills expanded from 36 to 42.

It is clear that a robust rubric that people can be trained to use reliably would be difficult. This exercise showed me that it is very hard to identify which skill is causing a particular behavior. Training requires working through many hypothetical scenarios. Once Sam and I had discussed each scenario and the possible categories that could be the cause of the behavior, we always agreed on which categories were causing difficulty for the student; however, it took this exercise, of exploring the possible categories, to learn how to use the rubric.

Results

After Sam and I each independently scored the students, the results were compared for a list of 31 skills. Our grading was quite consistent, even without collapsing the grading to a 3-point scale. We ended up with 31 skills because there were a few that Sam found could not be graded by watching the videos of the students. These skills were

- ♦ Previously known facts – The material discussed during the quantum mechanics interviews was new to the students and rarely could they incorporate previously known information.
- ♦ Spatial / Mapping – Not used in the quantum mechanics interviews.
- ♦ Estimation – Not applicable to the specific quantum mechanics questions.
- ♦ Number Sense – Not applicable to the specific quantum mechanics questions.
- ♦ Enjoyed analyzing the interns – Specific to the evaluation tool scenario.

In addition, there were a couple of categories where our scores differed due to the difference in the problems:

- ♦ Reading Comprehension – The quantum mechanics interviews did not require extensive reading only brief problem statements while the Jasper story and intern script do require a fair amount of reading.
- ♦ Acquires Info 1st time through – The quantum mechanics interviews did not require extensive reading so this skill was used infrequently while it's imperative in the Jasper story.

Table II: Scores from Evaluation tool interviews and Quantum Mechanics Interviews

	Rick		CC		HS		Guy		Raj	
	ET	QM	ET	QM	ET	QM	ET	QM	ET	QM
math - basic add/sub/mult/div	3	NA	4	5	4	5	4	5	3	5
math - equation formation	4	3	4	5	4	5	5	4	4	5
Reading comprehension	2	1	3	3	3.5	4	3.5	3	3	3
acquires info 1st time through		1	2	3	1	4	2.5	4	2	5
remember previously noted facts	4	1	3	5	5	5	4	5	4	5
remember what he/she has calculated or figured out	4	1	4	5	5	4	4	5	3	5
planning ideas (what – ask questions)	4	1	5	3	2	4	4	3	4	4
Planning - Big Picture (Visualization)	1	1	4	3	4	4	3	4	4	4
planning way to get answer (how)	4	1	4	5	5	5	5	5	4.5	5
connect steps and pieces		1	4	4	5	5	4	3	4	5
Check calculations	4	1	2	3	5	4	4	2	4	4
aware of how girls/interviewer helped	2	4	4	5	4	5	4	4	4	4
Monitored own progress	1	2	4	4	2	4	2	2		4
keeps problem framework in mind (remembering plan)	4	1	4	4	4	4	3	4	4	5
knowledge of own strengths and weaknesses	3	3	4	2	4	4	4	4	4	3
Skepticism	2	1	2	3	4	3	2	3	4	5
Creativity	1	1	3	3	5	4	4	3	3	4
Adaptability (shift direction easily)	1	2	5	5	5	5	3.5	3	3	4
can throw out useless info	1	1		4	4.5	5	4	4	3	3
Judgment of reasonable issues	1	1		4	5	5	3.5	4	3	1
ties in personal experience	4	3		2	4	5	5	4	4	5
tie in info from another person with what they are working on	**1**	**3**	4	5	5	5	4	4	3	4
Scientific Process (logically dominated thinker)	3	1	4	3	5	4	2	2	4	5
Confidence	**3**	**1**	3	3	4	4	3.5	3	3.5	2
Enjoyed Solving the problem	4	1	5	3	4	5	3.5	4	4	4
Enjoyed complete interview		5		4	4	5	4		4	4
Wanted to Succeed on 'test'		5		3	3	3	5	5	5	5
Attribution (who's to blame)	4	5	4	5	4	5	N/A	5	4	5
Wanting to find the best solution - for self	4	5	5	3	5	5	4	4	5	5
Wanting to find the best solution - for interviewer		5		3	3	3		5		3

Table II contains the 31 skills that were graded by both Sam and I for each student on a 5-point scale. Our scores are quite consistent for each student. What is even more striking is a comparison of summaries written independently before discussing the students. Here are excerpts from each interviewer about the same student:

Quote from QM interview:

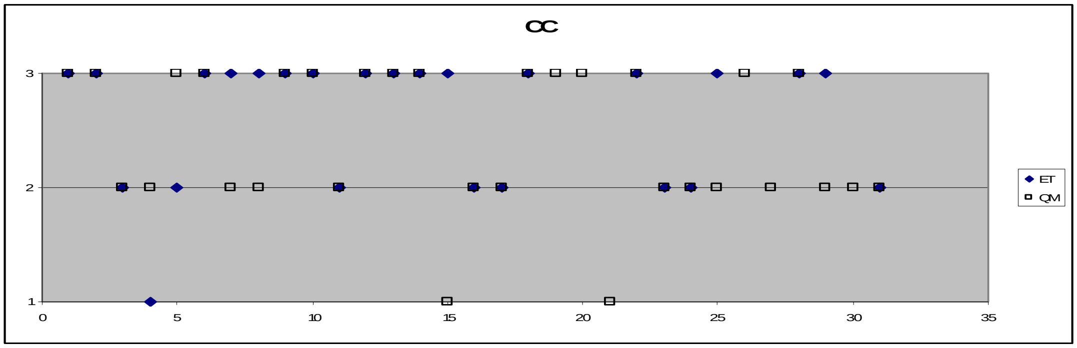
> *"she seemed to view learning [as] how to accept every weird thing we told her… she thought the first step was to accept things and the second step was to try to understand them. She always rethought her ideas when another student suggested something, although she maintained enough skepticism to recognize that other students were often wrong."*

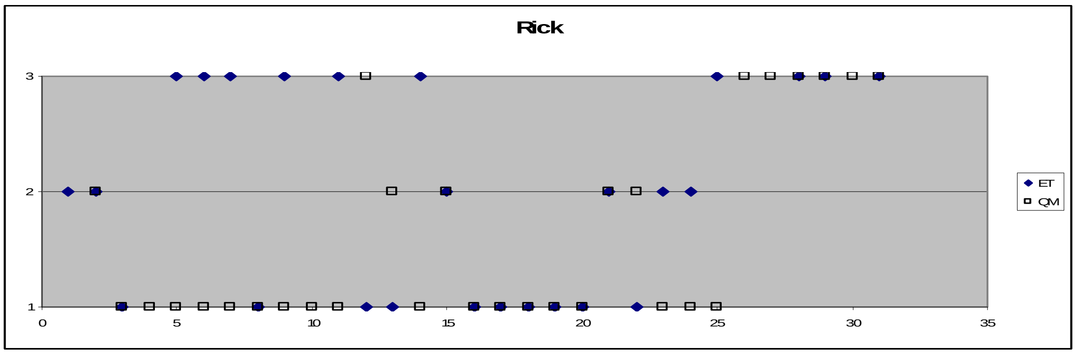
Quote from content-free evaluation tool interview:

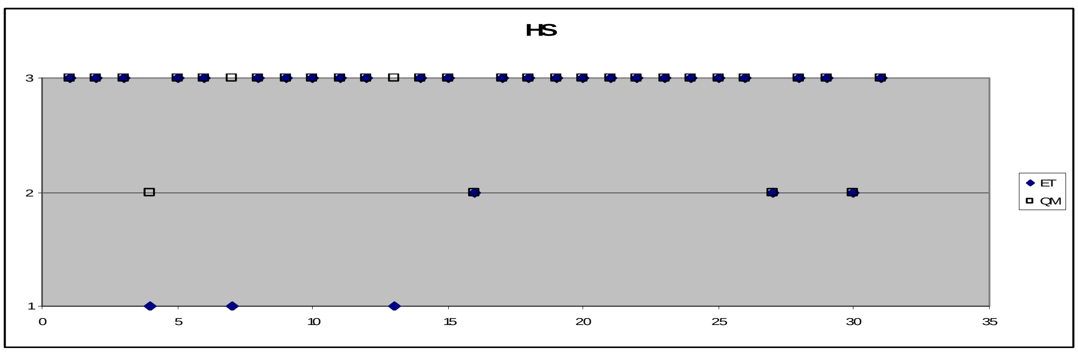
> *"Always had a knee jerk response which was not always good but then on her own she considers carefully and comes up with the right answer. …she'll consider whatever is thrown out there, decisions are based on the most logical answer. If a suggestion does not make sense after careful consideration, she holds onto her beliefs. Very reliable"*

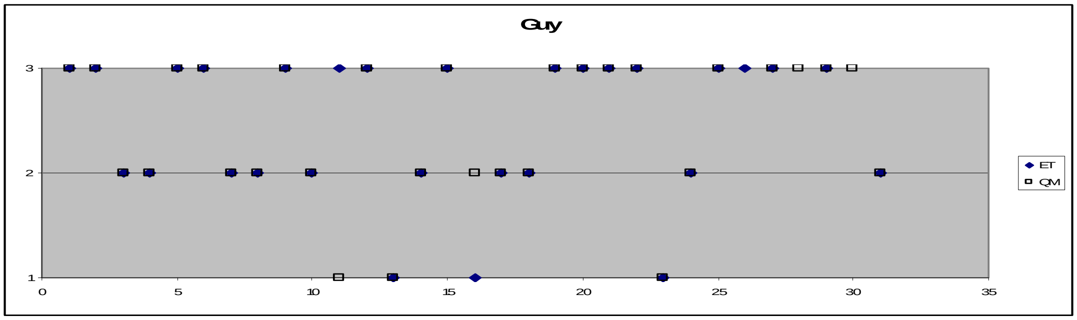
This is just an example from one student; however, the summaries from each student were just as consistent for every student. Even with as much as a year between the quantum mechanics interviews and the content-free evaluation tool interview. We were also each able to predict the level of success that each student had in the other domain, ie. Sam accurately predicted their ability to solve the problem in the evaluation tool and I was able to predict how the students did in quantum mechanics.

Figure III shows graphs of the quantum mechanics results compared to the evaluation tool results for each student collapsed to a three point scale. It is easy to see that these comparisons are much closer than those of the pyramid interviews. This demonstrates the greater precision of an interview with the evaluation tool rather than written results. It also is reflective of the work that was done refining the definitions of the categories and the rubric.


CC
3
2
1
0
5
10
15
20
25
30
35
ET
QM


Rick
3
2
1
0
5
10
15
20
25
30
35
ET
QM


HS
3
2
1
0
5
10
15
20
25
30
35
ET
QM



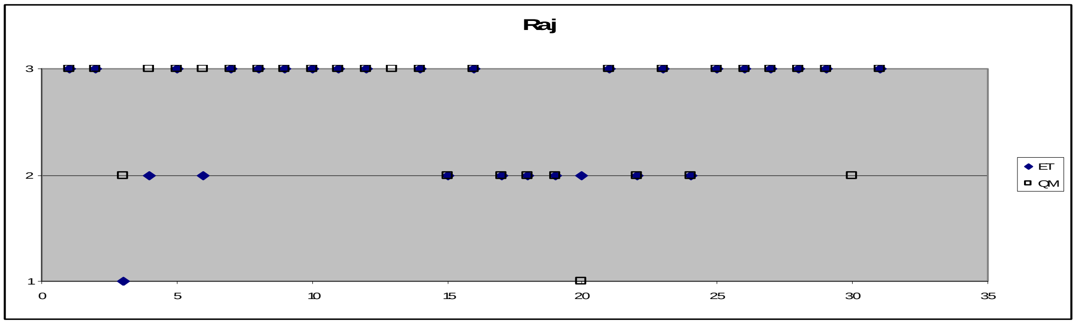


Figure III – Graphs of quantum mechanics interview results compared to evaluation tool interview results on a 3-point scale.

Four of the five subjects have scores that agree in better than 90% of the categories. The subject we will call Rick did have several categories that were scored differently by each interviewer. Sam stated that in all of their interviews, Rick was never able to successfully solve one problem. He would jump from idea to idea often not really sure where to go next. After discussing this person's behavior, we decided that during the quantum mechanics interviews, Rick often was not engaged in problem solving. He wanted to engage and tried very hard; however, the problems were beyond his skill level and unsolvable for him. While working on the Jasper problem, Rick did a fair job. He did get quite stuck on some very unnecessary details but did engage in problem solving during most of the interview. It appears that Sam and I were not scoring Rick engaging in the same activity – problem solving. This example emphasizes the importance of having a student engaged in problem solving to measure their skills.

Problem Solving

Component skills' impact on successful problem solving

Using various combinations of skills

When interviewing students, I found that weaker solvers were actually the most informative. It is difficult to analyze good students because weaknesses are much easier to identify than strengths. When a person is successful, the solving process is a smooth demonstration of solving. Many things that the solver does and knowledge they have are not independently observable. By studying a variety of subjects, it was possible to tease apart specific skills that are needed to solve problems by observing different solvers with different weaknesses. When a person is unable to use a specific skill it becomes clear that it is missing. Physics students were generally capable of solving the Jasper problem and do not often require the assistance of the interns; however, elementary education majors have a much narrower set of skills often struggling with the Jasper problem and leaning heavily on the interns.

Results of interviewing 30 subjects and analyzing written results of another 16 show that not all skills are necessary to be successful. Often people can use various combinations of skills in place of others. It is also the case that different problems require different combinations of skills. Thus a person can put themselves into a field of study that fits with their particular strengths. However, there are certain skills that if they are weak, are debilitating. Weaknesses in content knowledge can stop all solving in the artificial environment of a school type problem; however, in real life a strong problem solver can figure out how to get the missing information unless the problem is time sensitive. Weaknesses in the beliefs/motivation category can be particularly debilitating. Weaknesses in metacognitive skills can be overcome by other combinations of skills and conversely, strong metacognitive skills can overcome other weak skills.

As mentioned above, the physics majors were able to find solutions to the problem using strong math, planning, and judgment skills; however, these students did not demonstrate strong metacognitive skills or engage in meta-processing. I believe the reason for this is that their personal set of skills with which they are proficient is nearly complete and has been adequate for learning in their major so far. When physics students decide to go to graduate school and work in a research lab, they are faced with more complicated problems, than are typically seen in the classroom, that require a broader range of skills. When this happens, the missing skill(s) become readily apparent. To work through the process of determining where they are weak, having metacognitive skills to identify their own weaknesses helps facilitate the process of correcting any weaknesses. This means that quite often, when faced with this new challenge, the metacognitive skills are now more important and students, at this stage, work on building this ability out of necessity. In addition, many of these problems are sufficiently challenging that meta-processing becomes crucial to finding the solution to the sorts of unique problems that researchers encounter. As Schoenfeld says "The research indicates that the presence of such behavior [meta-processing] has a positive impact on intellectual performance. That its absence can have a strong negative effect – when access to the right knowledge is not automatic…”

Many elementary education majors pose the counter example to physics majors. The elementary education majors that were interviewed had many more weaknesses. The weaknesses varied but none had a strong, broad skill set resembling a physics major. To compensate, almost all of these students had strong metacognitive skills (reflection on their own skills), some were careful and thorough when working on a problem and a few even engaged in meta-processing. These skills helped them compensate for their many other weaknesses in knowledge and processes.

Can they do it vs. do they do it

The thorough literature review coupled with the exercise of writing this dissertation has helped me understand a few of the weaknesses in my rubric as it stood after the latest set of interviews comparing the content-free evaluation tool to quantum mechanics problem solving. There were a couple of categories that I was not able to satisfactorily describe for Sam when we were collaborating. The exercise of sorting my problem solving skills into the three categories that I settled on for the literature review - knowledge, Beliefs/motivation and processes - has shed light on the difficulties that I was having when working with Sam on these categories and a few other items that were unsettling when scoring students during an interview.

As I described in Chapter 1, knowledge is something a person has and brings with them to the problem, processes are something they do when they are solving the problem while beliefs, expectations and motivation are the ideas that a person has that facilitate how they use their knowledge and what processes they do engage in. The idea of “Can a solver do something?” versus “Will they do that thing?”

The category of estimation is an example. In the script between the interns they discuss Emily’s weight. The solver is asked to give what they believe is Emily’s weight. Then further discussion between the interns points out that there’s a picture of her standing on a scale and the solver is asked if they can see what she weighs. On the following page one of the interns carefully observes the picture, uses the tick marks on the scale and counts out what Emily must weigh. When people work through these pages, some make the estimation on their own and count out the tick marks. Many others look at the picture and say it’s impossible to tell but after being prompted by the interns’ description of how she did it, then proceed to do it themselves. Then there are a further group of people who even after being given directions on how to do this, still are unable.

Another similar example occurs with the category of scientifically dominated meaning does a person need a reason for each step and do they follow step-by-step process when solving. During the pyramid interviews and during observations of students solving physics problems, students expect that they must do this in order to solve a scientific/math problem. In those types of problem settings the measurement is then simply one of *can* they follow through in a step by step logical progression. With the content-free evaluation tool, the scenario of saving a wounded eagle does not automatically put students into this scientific frame of mind. In this case, I had two things to measure, do they find it useful to approach the problem in this manner and then if they do, can they do it.

Because the processes that a person engages in depend on their knowledge (what they can do) and on their beliefs/motivations (do they think it's useful, do they think they can do it, do they want to do it) there are skills that are quite difficult to tease apart and determine exactly which category they would most appropriately fall. This structure is still quite informative however, because it does require this understanding of what a solver has and why they engage in particular processes. It's not just as simple as they know this so they should do it.

Disengagement

During the compilation of the data for this dissertation, there appears to be another category/behavior that is crucial for success when solving difficult problems. There are some people who are no longer able to productively process after certain events or series of events. I'll term it 'disengagement'. In most cases it's involuntary. I've seen this happen with several interview students and in the classroom. The trigger is different depending on the person. It is similar to but the opposite of engaged exploration described in Chapter 2. During the simulation interviews students had to be in engaged exploration to gain anything new from the simulations (Adams et al, 2008a). There were times that they would begin by describing what they were seeing or what they'd learned previously in class or elsewhere, but were not engaged in problem solving or sense-making. Once the students began interacting and exploring – asking their own questions – they were engaged. At times it'd take a little prompting but it would always happen. There was one exception that I believe is an example of this disengagement phenomena and is discussed in more detail below. While doing the problem solving interviews, I had one student with similar hesitation to engage. She considered the need to save an eagle from the wilderness by a woman who had only flown an ultra-light one time to be a completely unreasonable life risking feat. She did not think carefully about the first several questions and answered them quickly and superficially. Finally the questions asked by the interns required more careful thought and at that point she finally became engaged. When this transition occurred, her skills were quite strong; however, she never gained any sort of belief in the problem scenario. This lack of belief was actually evidence of her strong real world thoughtfulness about the problem.

Disengagement is when the person is unable to successfully process. Something in the solving process happens such as the problem scenario contains a difficulty that the student is unable to cope with and they are (involuntarily) able to continue engaging. Many times the person continues to try to think about the problem but are unable to. Below I will discuss a few cases during problem solving interviews that I saw this. In addition, I have come to realize that I personally have this happen when trying to solve a problem in front of group of people such as when I'm teaching.

The first student that demonstrated this instant lack of processing was a student, we will call Mike, in calculus based introductory physics at the University of Northern Colorado. The professor of the course asked me to interview Mike because she was trying to figure out why he was having difficulty. Mike was involved in the pyramid problem study so was solving a mechanics problem rather than taking the problem solving survey. Mike did quite well the first half of each hour long interview and then reached a point in each where he could not figure something out and then lost track of what he had already known and accomplished. He'd get frustrated with himself when this would happen and continue to work but at a much more superficial level accomplishing very little. Mike had many very strong skills such as visualizing, asking questions, creativity and judging useful information; however, when he got to his capacity of new information, that was it. All progress stopped and he fell backwards. I could see it coming in each interview session because it was like he started walking up a steep hill, progressing painfully slowly. Then suddenly he'd lose it completely and slide back almost to the beginning. He was not able to recover in either interview session.

The next example is Buffy, who is described in the workplace comparison section above. Buffy would hit a difficult spot in the problem and immediately become very agitated. When she'd try to find the information she needed or to calculate what she needed after hitting this difficulty, she became very frustrated and tried to quickly flip back through the problem scenario without really looking. She also started in with a whole series of comments about external noises, not enough time, how stupid the interns were etc… Buffy did the same thing in her job and similar to Mike, she would not recover without going home and coming back the next day.

During the simulation interviews there was a situation with the specific simulation that was studied one week that caused three of the four interview students to exhibit this same disengagement behavior. These three students had previously worked with the simulation and remembered that they ‘knew’ how it worked and what was going on. The fourth had never seen it and engaged quite productively. Four months before this set of interviews the three students had worked with the simulation in class and done homework with it. The students had forgotten a few things and the question being asked was actually quite challenging. Rather than using the simulation to help them connect some ideas and create an explanation, they kept trying to remember the answer. All got frustrated and came up with a variety of excuses for their inability to answer the question. Despite my encouragement and prompts, I could not get any of them to engage; although, all had done so productively when fist working with this and other simulations before.

When I’m teaching, I always make sure my notes are meticulously prepared and that I’ve thought about all the ins and outs of a problem conceptually before stepping in front of the students. The reason for this is that I know that if I come across a new question that I have not previously thought through, I will screw it up in front of the class. Then two minutes after I leave the room, the correct answer will become quite obvious to me. This is because I have this same problem where I’m not able to engage in productive thinking when under the pressure of looking smart in front of my students. I’ll do the same thing in a large group. I’ve learned to cope with this by buying time. I simply refuse to offer a solution while in the group setting. Then I can follow up once I’ve had a few minutes alone to process. Such behavior is also seen routinely in graduate oral exams, where some students find the environment and format so stressful they are incapable of any productive problem solving.

There are a few studies that describe students that exhibit disengagement (Dweck, 1999) but describe only very specific cases. Dweck speaks of students exhibiting helpless behavior. This helpless behavior only happens with a certain subset of the population. This group is part of the group that also believes that one's intelligence is a fixed quantity rather than a malleable thing that can be improved with work. In her specific studies she has not seen this behavior with students who do believe intelligence is malleable. In her studies, students have successfully solved problems that fit their level and then are given problems that are beyond what they are currently capable of. Some of those in the fixed intelligence belief group struggled with the problem a short time and then disengaged. At this point they began making up answers and telling stories of other unrelated personal accomplishments. After becoming disengaged (helpless), these students couldn't even begin to solve the problems that they'd previously successfully solved.

After careful review of my interview subjects and the literature it seems that disengagement is a sort of emotional response that causes a chemical change in the brain. When this happens the person is unable to engage in the problem and will also lose track of information that they previously knew or could handle. The only way to recover is to remove themselves from the situation and wait for things to return to normal. At which point they can become productively engaged once again.

Implications for Teaching

Improving Problem Solving

Problem solving is teachable but not all at once. There are many specific skills which must be addressed, one at a time, similar to teaching physics. Instruction in physics focuses on one aspect at a time, slowly building a knowledge base over several years. With problem solving, knowing that there are specific component skills can help an instructor focus activities on individual skills helping students slowly build up a strong set of problem solving skills. Quite often problem solving instruction tries to teach the art of problem solving all at one time with a step by step strategy. These attempts have met with, at best, limited success. (Bunce, D. M. and Heikkinen, H., 1986; Heller, Keith, Anderson, 1992; Heller and Hollabaugh, 1992; Heller and Reif, 1984; DiLisi, Eulberg, Lanese and Padovan, 2006; Huffman, 1997; Leonard, Dufresne and Mestre, 1996) When one sees that problem solving requires a set of specific component skills, it is not surprising that problem solving can not be taught all at once.

Anderson and Schunn apply the ACT-R learning theory to teaching (2000). How it applies depends on whether education's motivation is long term knowledge to create a better public or short term success as indicated by in class assessments. They discuss how a lot of research and testing of learning looks at quick-to-learn items such as mnemonics as indicators of success; however, if a person practiced language on a regular basis, these mnemonics cease to be relevant. If a person only learns mnemonics, then tests have shown that over time a person forgets more. What they've found is that declarative knowledge is easier to forget than procedures;

however, procedures take more time to build and are very specific. A person must undergo extensive practice, in slightly different contexts, to build a broader firmer procedure. This is demonstrated in Anderson, Simon and Reder (1996) and consistent with DiSessa's ideas about how to build a complete coordination class (2005) as well as Ericsson's work on becoming an expert (2006).

Practice is key; however, it is not helpful if the correct things are not being practiced. Various studies have shown that students who do poorly in different subjects are missing some basic information to support their understanding of a subject. For example, Palinscar and Brown (1984) produced dramatic improvements in student's reading comprehension by teaching students about and having them practice asking questions, summarizing and clarifying difficulties. Apparently these students were not there (physically or mentally) when these skills were taught in previous years. This is why tutoring can be so effective, a good tutor can spot individual deficiencies and proved specific feedback on how to address those deficiencies.

Anderson and Schunn say

> "This implies that there is a real value for an effort that takes a target domain, analyzes it into its underlying knowledge components, find examples that utilize these components, communicates these components, and monitors their learning. Unfortunately, cognitive task analysis receives relatively little institutional support. In psychology, there is little professional reward for such efforts beyond those concerned with basic reading and mathematics. The argument (which has been received from many a journal editor) is that such task analyses are studies of characteristics of specific task domains and not of psychological interest. For experts in the various target domains (e.g., mathematics), the reward is for doing advanced work in that domain, not for analyzing the cognitive structures underlying beginning competence. In education, such componential analyses have come to have a bad name based on the mistaken belief that it is not possible to identify the components of a complex skill. In part, this is a mistaken generalization from the failures of

> behaviorist efforts to analyze competences into a set of behavior objectives. Thus, there is a situation today where detailed cognitive analyses of various critical educational domains are largely ignored by psychologists, domain experts and educators."

My problem solving evaluation tool has analyzed the underlying knowledge components used in problem solving beyond declarative knowledge and has the ability to identify student deficiencies before the semester begins. This provides the information needed to develop appropriate in-class activities target student deficiencies. An example of how to focus on some of these often overlooked skills is discussed in Chapter 1, where Schwartz, Bransford and Sears (2003) consider two types of problem solving practice: Innovation and Efficiency. Both are useful however, most teaching only focuses on efficiency training - teaching students how to solve one type of problem at a time by showing them the method and then having them practice. Innovation training involves students devising their own solution strategy to a new problem situation. Students do not succeed at solving these types of problems; however, this research consistently demonstrates that this unsuccessful problem solving provides training that makes the students better prepared for future learning. It stands to reason that this different task of attempting to devise a solution method to an unfamiliar problem is accomplishing two things: 1. It provides practice of skills that aren't needed when shown a specific solution method; and, 2. After the innovation task, students start thinking about the new type of problem and form their own questions about how to handle it and are ready to find the missing pieces.

When thinking about how to teach problem solving, it is useful to consider that not all skills are required to solve all problems. Back of the chapter text-book problems are short and usually well-defined so require only a handful of the skills that have been catalogued during this study. There are many problem solving skills that are important in the workplace or the laboratory that are not encouraged in the typical classroom.

- ♦ Picking up information the first time through – quite important in real life but almost never required in school.
- ♦ Planning ideas (Creating questions) - This is a debilitating weakness in any work environment that requires leadership or independent work; however, it is not required with most problems where the problem is clearly laid out. Pretz, Naples and Sternberg (2003) discuss the important skill of recognizing the problem and express concern that students rarely get to practice this skill. Emphasis on group work also allows a person who has poor skills in this area to do well in school without ever using this skill.
- ♦ Visualize the Problem – Some problems do require this but those are rare. Most are not involved enough to require this process for solution. Although students are often told they should make a diagram in solving physics problem, it is a meaningless exercise to them when this is unnecessary for finding a solution.
- ♦ Remember previously noted facts – Most problems are short enough or broken into pieces for the student so this is not necessary.
- ♦ Keep problem framework in mind – Classroom problems are rarely complicated enough for this to be necessary.
- ♦ Can throw out useless information – Some problems give information that is not needed but it is not commonplace. Students have learned that the information that is provided can be a excellent clue in determining the correct solution process.

- ♦ Judgment of reasonable issues – Problems in most courses do not have multiple facets that require the student to consider different factors that may affect the outcome; so, students never have to use the skill of determining if a factor is important enough to affect the problem solution; however, in the real world this is a very common requirement.
- ♦ Creativity – Most school environments discourage this skill, though usually not intentionally.
- ♦ Skepticism – Also discouraged by the typical environment because it competes with another important lesson, do what your teacher tells you.
- ♦ Ties in personal experiences – Most problems do not have a real world context that requires the student to assimilate information from their previous experiences outside of the classroom; but a typical job requires this on a daily basis.
- ♦ Student vs. Real world – Typical classroom problems normally have all sorts of assumptions and unrealistic limitations so that one specific bit of information can be focused on; but, that also means opportunities to practice other skills such as considering all sorts of other factors that could affect a problem situation are never offered. In the real world, a person does not get the luxury of a clean simple problem.

One problem solving skill that is encouraged in the classroom but not intentionally is:

- ♦ Ability to analyze interns – As a student the ability to analyze your instructor goes along way toward success in the classroom. Quite often learning all the material in a class is an insurmountable task; however, learning the material that is important to your instructor is not. Once a student learns enough about the instructor to determine what to expect on the exams, their attention can be focused accordingly. Similarly, as an employee the ability to analyze your boss can be crucial.

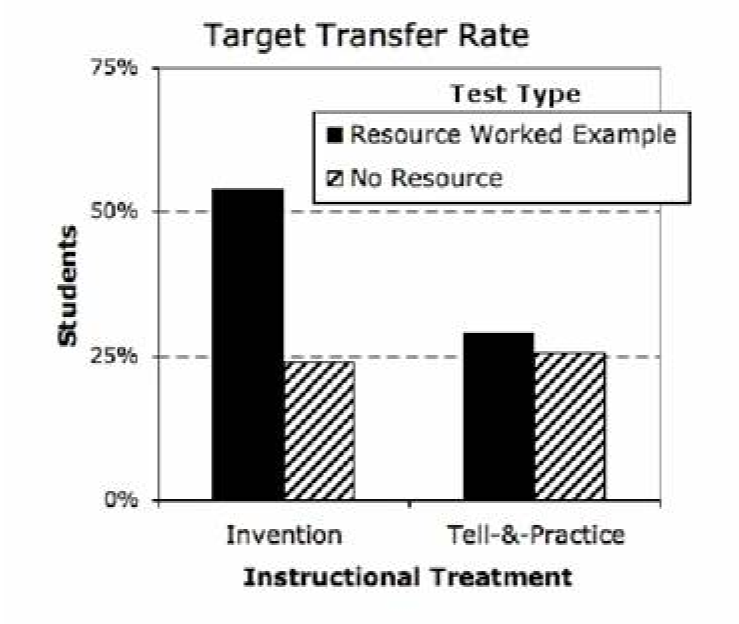
With the awareness that there are different specific component problem solving skills where certain combinations are effective for different types of problems, instructors can focus on specific skills by giving students appropriate problems that will require the use of the skill(s) that the teacher is currently focusing on teaching. Considering the specific skills that are being focused on can also inform new more useful evaluation methods to probe these skills.

Evaluation

Standard evaluation methods are good at evaluating student content knowledge; however, they rarely evaluate or even require the use of most of the skills that a person uses when solving new problems. If the focus of a curriculum is redesigned to include specific problem solving skills, then the evaluation methods will have to be reconsidered so that these skills can be measured. Considering problem solving as a set of component skills can help a person think about how to evaluate different sorts of teaching methods.

Schwartz, Bransford and Sears for example, advocate a new type of assessment called preparation for future learning (PFL). They term traditional evaluation as sequestered problem solving (SPS) which “makes us look dumb” while PFL “make us look smart. This is nicely illustrated with many examples of studies such as the Burgess' Eagle Challenge. 5th grade students and college students each had to create statewide recovery plans to increase the number of bald eagles in their state. The solutions developed by each group were equally poor making it appear that the education one receives from 6th grade to college is not useful for this sort of real-life situation. This assessment is considered an SPS type. An extension to this study involved telling the participants to create a list of questions they would ask to help them work out the problem. The questions revealed quite a difference in the two groups. The 5th grade questions were somewhat superficial but still useful while the college students' questions were much deeper and more useful. This not only showed a difference between the groups, but that neither were completely incapable as the SPS type evaluation had shown. The Challenge was later used with K-12 principals who showed even better problem solving abilities with respect to their use of resources and their willingness to let go of their initial ideas. This "tentativeness" was not seen in the college students who blissfully muddled on with their original ideas not even considering they could be based on incorrect assumptions.

Three further studies illustrate the differences between instructional technique and evaluation methods. In all three cases SPS type assessment showed equivalent disappointing results for transfer; however, PFL assessment revealed instructional techniques that included innovation and efficiency were far superior to other forms of instruction that used innovation or efficiency but not both. For example, 15 classes of 9th grade students received two weeks of equivalent instruction and then on the last day they received different treatments.



Invention Condition: Invent own solution for comparing high scores. No one succeeded.

Tell and practice: Shown how, then practiced while teachers answered questions and corrected.

Post test had a worked example that included summary measures (which had not been seen by the students previously) for only half the students.

Food for Thought

Is problem solving different from learning?

Researches who offer models of learning do so with the assumption that their model holds across discipline. Mayer offers the SOI (selecting, organizing and integrating information) model of learning (2006), Anderson and Schunn (2000) use the ACT-R learning theory and apply this model of learning to all domains. Many of the problem solving skills that I identify are characteristics of these theories of learning. So it is not surprising that student's use of these skills is consistent across domain. One could argue that learning is solving the problem of "I don't know how to do this, how do I figure it out?" This process requires asking questions, exploring, making connections, putting ideas together in new ways and then checking your conclusions to see if they fit with other things that you know. During the whole process, things go much smoother if you have the metacognitive ability to can step outside of the process every once in awhile to make sure you're moving in a positive direction and that what you're doing to learn is helping you.

To learn is to solve a problem. However, does solving a problem always constitute learning? Yes, if you are an adaptive expert problem solver. Wineburg (1998) asks

> "an important question in any study of expertise is how experts get to be that way. ….Snapshot studies such as this offer little…. But the study of expertise must also address a second key question: How is it that experts keep learning? Why do they continue to get smarter from encounters with materials and situations that leave other problem solvers unfazed (Holyoak, 12991; Perkins & Saloman, 1989)"

Some of the students that I interviewed learned many things from completing the evaluation tool. When reflecting on what they'd done many were excited to see they had figured out new ways to do some things. Others commented that if they'd done any problems like this before, they would have tackled this one differently. Still others wanted to know why there were not more experiences like this in school because this seemed so much more authentic and useful. It made them think about so many things they'd never considered before such as basic ideas of calculating how long it takes to travel from one town to the next using the speed limit and the distance or more complicated ideas of balancing different problem solutions when there is not an absolutely correct answer. The problem required thinking about the entire situation instead of an isolated unrealistic bit of information and how to use it. I believe when a person is *engaged* in problem solving, they are thinking about new ideas and putting them together in a way they have not previously done. This is learning. Wineburg says many people are unaffected by solving a problem. I also saw similar behavior with the interviews. Some students solved the problem on a very superficial level and engaged very little. Due to the nature of the script and series of questions, all did eventually engage; but, only because they were forced to. In these cases, much less was learned; but, I can't say they learned nothing because I'm sure these students could tell someone later about the problem scenario so in the strictest sense they did learn.

Schwartz says, "The fact that I achieve a correct answer in problem solving (when I could not before), does not entail I have learned." (2007). It may be that problem solving is a subset of learning; however, I am not convinced that true problem solving does not involve some learning every time. Schwartz also says that learning means "Something has changed." DiSessa and Wagner (2005) speak of Class A, B and C transfer. Their description of Class A seems very close if not identical to solving an exercise - it assumes that knowledge is *well prepared* and does not require further learning to apply - while his definition of Class B is the same as I've defined problem solving – it may not occur in short periods but presuming persistent effort, knowledge is *sufficiently prepared* to be applied (in a few hours or days in real-world problems) using learning and other resources that might be available. Class C is more typical learning - how *relatively unprepared* subjects use prior knowledge in early work - but Class B also entails learning.

Future Plans

Maximizing Widespread Usability

For the survey to be used on a widespread basis, particularly in the classroom, it must be very straightforward to grade. The exercise of training another person to use the rubric demonstrated that parsing a person's actions into specific skills requires extensive training and evaluation would likely be unreliable and/or unstable without such training. For this reason the evaluation tool has to have an unambiguous grading method. The most common way to do this is to create a multiple choice version of the survey. However, after careful consideration, it is clear that this method will not work to evaluate a number of the skills, such as creativity or metacognitive skills. The different options to each question provide ideas to the students and change their actions.

For this reason, I plan to create a computer administered and graded version of the evaluation tool. Preliminary work has shown that many of the questions elicit similar types of responses from the people being evaluated, and so I believe that it will be straight forward to carry out this conversion to a multiple choice format for some questions and to have a database of common responses for the computer to grade. Since the typical responses are similar, basic string recognition software should do a reasonably good job at scoring these questions. In addition to scoring open questions, the online administration provides the benefits of not allowing students to look ahead and could provide a tree structure that lets the students choose their own path. This would also make the evaluation tool that much more convenient for a teacher to use in their course.

Another concern for educators who want to use assessment tools in their courses is the amount of time required for administration. I have found in interviews that many problem solving skills can be determined within 30 minutes (i.e. planning what, picks up information the 1^{st} time through, judgment of reasonable issues, has to solve the problem); however, there are other skills (i.e. keeping track of information, connects steps and pieces, keeps problem framework in mind) where it is necessary for the student to be problem solving for 45 minutes to an hour before their ability in these areas is clear. For this reason, I am considering designing the evaluation tool so that it can be administered in either a short or long form. With the short form, the instructor can gain information on approximately 70% of the student's problem solving skills while the remaining 30% can be measured as desired using the long version.

Pre and Post Versions

A necessary component of the evaluation tool will be its ability to identify student improvement after instruction. To do this, it must be administered before and after instruction. Due to the nature of the Jasper story, it is not expected to be effective if given twice to the same student, since the story and solution will be familiar. To deal with this problem, once all the validation on the basic method is confirmed, a second story will be created that includes parallel questions to the first. Again the storyline will come from the Jasper series to maintain as much consistency as possible.

These two instruments will be validated by administering both of them at the beginning of the semester to a large enrollment class. Half the students will receive version One and the other half version Two. At the end of the semester students will take the version of the evaluation tool they did not take in the beginning of the semester. In this way a level of comparison can be determined between the two versions. Once the two versions have been completed, they will be administered in several different courses with version One used pre and version Two used post to evaluate what problem solving skills students are learning in different science classes.

When the evaluation tools have reached this stage, they will both need validated with students at the middle school and high school level. Interviews will be conducted with these students to first determine face validity and then concurrent validity following the previous procedures which were used with the college students. The Jasper problems were designed for and are used in 5th and 6th grade classrooms. We have targeted our reading level to be similar with the intent of using the survey from 5th grade to the workplace. Hence we expect that by basing a problem solving evaluation instrument on these Jasper problems, it should prove suitable for a wide range of ages.

Conclusion

I have created a content-free evaluation tool that successfully measures cross discipline problem solving skills in an interview situation and can also be used in a written format. In the process of developing this tool, I identified a distinct list of 44 individual skills that are used when solving problems. Understanding this set of problem solving skills will help teachers and researchers address problem solving. For example interviews showed that different types of problems require different combinations of these skills. Interview subjects often had different strong combinations of skills that were adequate in many situations to overcome certain weak skills; however, there are many skills that if weak, were debilitating. This listing of skills can also be used to determine the focus of a curriculum and identify many areas that are important but overlooked in the classroom.

This research has also shown that a person must be productively engaged in problem solving in order to measure their skills. Rick, who made little progress on the quantum mechanics problems, demonstrates this point. Recognizing that measurement of problem solving skills requires the subject to be productively engaged in problem solving shows that previous research on expert and novice problem solving provides limited information to the field of problem solving since the experts and sometimes the novices are not engaged in solving problems. The experts are extremely familiar with the area, and so the tasks presented are simply exercises for them. In most studies the tasks used are problems that are similar to text book problems and require only a limited selection of the problem solving skills for even the novices. Many studies do not even ask for the subject to find a solution and are focused specifically on one aspect of the problem such as grouping different problem statements together. There are however, many studies, as noted in chapter one, that do study people while they are engaged in solving problems. These studies and the research described in this dissertation have found consistent ideas about skills that are needed to solve problems and have also shown that there is still much to learn about how to teach problem solving. The work in this dissertation has provided an important step by elucidating the nearly complete set of problem solving skills and establishing a way to measure them.

Appendix A – Problem Solving Evaluation Tool

Physics Problem Solving Evaluation Tool

Introduction

Meet Jasmine and Sara, two interns working for your company. Both have been working for you all summer and done a great job. You have been informed by the higher ups that you can hire one of them permanently but not both. This summer they've both done a terrific job. Both are very personable and reliable. You know that a permanent position will be more challenging than the internship positions they've held so far. In an attempt to determine the stronger problem solver, you ask them to complete the Rescue at Boone's meadow challenge. The challenge is a 15 minute video about three people who must work together to save a wounded Eagle. In an attempt to determine the stronger problem solver, you listen to the Sara and Jasmine's dialog and analyze how they're progressing as they discuss the details of the problem and some possible solutions.

Important: You won't be able to watch the video, but lots of pictures have been provided along with the story. There is information necessary for solving the problem embedded within these pictures.

About You

First, please answer a few questions about yourself.

Name:_______________________________________

Year in School:

Declared Major:

If planning to change your major change, intended major:

Gender:

Race/Ethnicity:

Age:

# of credits:

Do you work? If so, how many hours per week:

Current GPA:

Please read the online story summary and then proceed to the next page of this packet to analyze Jasmine and Sara's problem solving.

http://cosmos.colorado.edu/CPPSS/ Choose Rescue at Boone.doc for the story or Survey 2_28.doc for an electronic copy of this survey.

Analysis

Please answer the questions in bold as you go. Answer each question in order before reading ahead to see what Jasmine and Sara think of next. Feel free to refer back to the story summary as often as you like. Remember you must choose which intern to hire so keep this in mind while critiquing their solution process.

After watching the video, Jasmine and Sara sit down and start thinking about the problem.

Jasmine thinks awhile and says, "I just don't know what to do about Boone's meadow. Remember at the vet's? Dr. Ramirez said that the field is only half as long as you need to land a plane. However long that was, I don't remember. Anyway, I was thinking that maybe someone could hike in halfway and meet Jasper..."

1. **Do you have any ideas about the fastest plan to rescue the Eagle? What would you try to figure out first?** (No need to work out the whole problem right now.)

2. **How do you feel about this problem?**
    a. **Could solve it on my own right now.**
    b. **Would rather solve it in a group setting.**
    c. **Need more information before I could say.**
    d. **It's completely over my head.**

    Please explain further:

I'm sure they expect the ultra light to land in the field at Boone's meadow, that's sort of the whole point of the video. Don't you think? They also mention payload a lot so it must be important too.

Sure. I guess that's probably true. Why else would they have shown us the plane if it couldn't land there?

3. What do you think?

After a little discussion they agree on a plan.

Ok, so we'll have Larry fly straight to Boone's Meadow, pick up the Eagle and then come back to Doc's? There aren't any extra stops so it should be the fastest way to save the Eagle.

4. Is this a reasonable plan? Is there any reason that their plan might not work?

5. Is there something you want to do next or would you rather see what the girls are planning?

I think Larry weighs 180 pounds, plus we need the box for the Eagle.

Yeah, that sounds right. I think they said the payload that the plane could carry was 220 pounds.

So unless it's a really heavy box we should be fine. I guess we should watch that part again just to be sure.

6. What is the safe payload capacity and the weight of the box?

7. How confident are you about your answer?
   a. Positive
   b. Pretty sure
   c. Think it's close
   d. Not sure at all

Do you think that means we have to count the gas in the tanks as part of the payload or maybe it only counts extra fuel that you take along?

Sara considers this for awhile. That seems strange but he did say the payload includes fuel so I think we should count that. What does a full tank of gas weigh?

How should I know? It held 5 gallons, does that help?

8. Do you know what the gas weighs or how they could figure this out?

Well…. I remember learning that water weighs 8 pounds per gallon. What's 5 gallons times 8 pounds per gallon?

They both said "40 pounds" in unison.

Jasmine writes this down and says, "Ok. I have 180 pound Larry, 40 pounds of fuel plus the 10 pound box. That's too much. It comes to 230 pounds!"

9. What are you thinking right now?

Let me think… Gas floats so it must be lighter than water.

Really?

At this point it is 5:00 so Jasmine and Sara pack up for the day.

10. Does the 5 gallons of gas weigh 40 pounds, or is Sara right that it weighs less since it floats?

11. Is the weight of the gas important?

12. After the end of the first day, what are your initial impressions of Sara's and Jasmine's problem solving skills?

13. Do they both seem to know what they are talking about? Do you trust the facts and calculations that they've provided so far?

14. Who would you hire if you had to pick right now? Why?

The second day:

Where were we yesterday when we quit? Let me look at my notes… Ok, it looks like you were thinking that gas is lighter than water. So far it seems like all the information has been in the video, maybe we should start watching it again to see if they tell us the weight of the gas.

I'm sure they didn't say anything about the weight of the fuel. I would have remembered. You can look if you want.

While Jasmine started the video, Sara sat and thought about everything they'd done the day before, trying to recall the weights and payload capacity. Thinking about anything else they might need to calculate or a way they could stop and get more fuel since 5 gallons was too heavy.

Sara, look at this. There's a number on the gas tank. It looks like it says 30 pounds above the 5 gallons!"

Nice. I missed that. That means we're right at 220 pounds, just right!"

6

Jasmine consults her notebook, writes down the new value for the weight of the fuel and rechecks the calculation.

Let me see… 180 pounds plus 30 pounds plus 10 pounds….Ok, I get 220 pounds also.
So is that it? Are we done?

15. Are they done? Is this a good plan to get the Eagle? If not, why not?

16. How confident are you about your answer?

- **e. Positive**
- **f. Pretty sure**
- **g. Think it's close**
- **h. Not sure at all**

7

It can't be that simple. We'd better check to see if 5 gallons is enough gas. What kind of mileage does the plane get?

How should I know the mileage for an ultra light plane?

I remember Emily said something about 1 ½ gallons gone in each tank and Larry said he'd flown somewhere that morning. Do you remember how far?

No. He said he'd gone to Hattonville, or somewhere like that, and back… I think. We need that map on Doc Ramirez's wall!

17. Do you remember how far and how much gas it was?

18. How many miles per gallon can the ultra-light go on a calm day? How did you figure this?

19. How confident are you about your answer?

- **a. Positive**
- **b. Pretty sure**
- **c. Think it's close**
- **d. Not sure at all**

8

They watched the appropriate bit of video and found that he'd gone to Headlyville which was 30 miles round trip and it took 2 gallons to make the complete trip. Sara does a quick calculation in her head while Jasmine wrote it all down in her notebook.

So what's that… 15 miles to the gallon? So how far is it to Boone's meadow? I remember Jasper said something about 60 miles but… I think that was to the service station. The vet measured it on the map didn't he?

They watched the video again and saw that the vet did measure the distance to Boone's meadow with his calipers to be about 65 miles. Jasmine writes down the 65 miles and pulls out her calculator.

Hmmm… So 65 miles to Boone's Meadow and 60 miles to Hilda's so it must be 5 miles between Hilda's and Boone's Meadow but we don't need that. Anyway, let's see…. 5 gallons is enough for 75 miles. That gets us to Boone's meadow!

Wait! You're going too fast. Give me a second to work this all out.

20. Now that they've checked the plane's mileage does this confirm that their plan to have Larry fly straight to Boone's Meadow, pick up the Eagle and fly back work? Why?

21. How far does Sara say it is between Boone's Meadow and Hilda's? Do you agree with her?

22. At this point is there anything on your mind that you would like to check or try or are you happy to wait for the girls?

9

Oh No! You know what we forgot? We have to have enough gas to get back again!

Wait! I think I saw an extra gas can when Larry showed Emily the box. If we fill that up, Larry can get the Eagle, refill the tank from the extra can and fly back.

That sounds alright as long as the gas smell in the cargo box doesn't hurt the Eagle. That would be sad

23. Is there any reason it won't work to take along enough extra gas in the gas can that Larry pulls out of the cargo box?

What about money? They made a point of showing how much money each person had left after supper. How much did they each have? I couldn't really see but it looked like Jasper had the most. Since we need more gas than the one tank, maybe we're supposed to go to the gas station after getting the Eagle and when we get the Eagle we get some more money from Jasper. I bet that's why they showed all that stuff with the guy driving the Mustang. We need to know how much per gallon gas costs at Hilda's. It was cheap. Remember?"

Ahhhh, does this problem never end? Oh, look at the time. We have to go to lunch. Shucks, I really wanted to work on the money thing.

24. Do you think money is a part of the problem that they must consider?

25. This is the end of their 2nd session. If you had to choose now, who would you hire permanently? Why?

You know what? I was thinking about this problem during lunch and just don't think we should worry about the money. The plane is fueled up right now and they didn't ever say that was all the money they each had. Let's just skip this part and worry about the rest.

Do you think they showed us the money and price of gas for no reason?

Maybe they want to rub it in how much gas prices have gone up since the mid 90's.

I guess how to get a lemon Jell-O cone out of a stubborn old waitress doesn't really matter either.

26. Are they overlooking an important detail with the money?

So where were we? Let me see what I have written here…. Ok, I have that we have enough gas to get to Boone's meadow since the plane is fueled up. 5 gallons = 75 miles. The vet measured Boone's meadow to be 65 miles from his clinic. 10 miles to spare but we still have to get the Eagle back to Doc's so we were talking about using the gas can to carry extra fuel. 65 miles to Boone's Meadow and back again is 130 miles.

So how many gallons do we need? How do I do this? Is it 130 miles divided by 75 miles….?

27. What does Jasmine need to do to figure out how many gallons they need to go the 130 miles?

So 4 gallons in the spare container. (Looking up at the ceiling while she does a quick calculation.) The payload with Larry, the box, a full tank plus the extra fuel is 25 or so pounds too much not to mention the eagle…

What about Emily? What did she weigh?

28. What about Emily? Is this a good idea?

- **a. No, Larry is supposed to fly the plane.**
- **b. No, Emily does not know how to fly the plane.**
- **c. No, it'll be fine using Larry.**
- **d. I didn't think of this but it just might work.**
- **e. Yes, I've been waiting for them to figure this out.**

Stop it! You're confusing me. I need to work this out on my own. (Takes a few minutes to work everything out and confirm that Sara did her math right) Actually it's not 4 gallons. It's 3.67 gallons. That's still more than the plane can carry because it adds another 22 pounds. Emily huh….? Ok, let's watch the video and see what the scale said she weighs.

The girls start watching parts of the video again. They look closely in the beginning at Emily standing next to Larry.

Look at her here. She's a little heavy set and she's got strong arms and shoulders. She probably came from the dense side of the family too! I'd guess, even though she's pretty short, that she weighs at least 150 pounds. Maybe more.

Let's look at the part where Larry weighs himself. I think they showed her weight in that part.

29. Does it look like they'll be able to get what they need from the scale? (What you see in the pictures is as much detail as Sara and Jasmine were able to see on the video.)

30. What do you think Emily's weight is?

31. How did you decide this?

They don't show Emily's weight. There's no way to read the numbers on the scale when the camera is behind her like that. We'll just have to guess.

Can I see? (Jasmine looks again at Larry's weight and spends some time going back and forth between Larry and Emily on the scale.) I think Emily weighs 120 pounds.

How can you see that?

See in the picture of Larry's weight. The needle points almost to where half past 2 o'clock would be. I looked at the picture of Emily and it looks like the needle has to be pointing almost straight up, but not quite. That would be about the 120 pound mark."

No Way! Just look at her!

32. Does Jasmine have a good idea here?

33. What do you think Emily weighs?

34. How did you decide this?

150 pound Emily plus 9 gallons of gas, plus the 10 pound cargo box should be just fine. I think I figured 6 pounds for each gallon. So 9 gallons times 6 pounds per gallon is 54 pounds. Plus the box plus Emily. *Emily is 150 pounds*. No problem, we have 6 pounds to spare!"

Hold on… I need to get this all down… Ok. Now, what about the Eagle? The vet said how much it weighed – I think?

35. Will this put them over the maximum payload? What should they do?

Uh oh. I didn't think of that. Ummm…….. Oh, we'll have used up half of the gas getting there so we'll have room for the Eagle! So Emily's going to fill up the extra gas can with 4 more gallons of gas, fly the plane to Boone's meadow, fill the plane back up with the four gallons from the extra gas container, pick up the Eagle and fly back. Sounds good ?

Sure. That works. Are we done, *yet*?

Yes, I think so.

36. Are they done?

37. How confident are you about your answer?

- **a. Positive**
- **b. Pretty sure**
- **c. Think it's close**
- **d. Not sure at all**

15

The instructor comes by and asks if they're finished. They said yes and told her the plan. The instructor asks them if they answered both questions.

It looks like we only answered the first question. The quickest way to get the Eagle to Cumberland City. Now we have to figure out how long that will take. I guess that means we need to know how fast the plane flies.

Well…. in the beginning when Larry talks about how much gas he used to go to Headlyville he said "2 minutes for every mile is my rule of thumb"

38. What is this in miles per hour? How did you figure this?

5 minutes for Emily to get to the flying field.
5 minutes for her to fill the extra gas can with 4 gallons of fuel.
5 minutes to load the plane and take off.
65 miles times 2 minutes per mile gives us 130 minutes to Boone's Meadow.
5 minutes for Jasper to load the Eagle while Emily refuels using the extra gas can.
130 minutes to fly back to the airfield.
5 minutes to the vet's.

So a grand total of 270 minutes. Right?

39. Does this look reasonable?

Let me check…. Yeah, I get the same answer. It seems like a long time. What is that in hours? 270 divided by 60 minutes per hour gives us exactly 5 hours.

Well, it would take six hours for Jasper to hike out and then he'd still have to get to Doc's. So our plan is better than that.

40. Are they set?

41. Now that they are done, would you like to suggest any changes or does this look good to you?

After presenting their plan, Sara and Jasmine see that other teams who've solved this problem in the past have not gotten the same times. One team's rescue took a little longer and there are a couple other teams with times quite a bit faster than theirs. The slowest time was 5 hours and 25 minutes while the fastest time was 3 hours and 55 minutes.

42. How do you think the other teams got different times?

43. **Were Sara and Jasmine both good problem solvers? Was one of them better at it than the other? Please rank each of them on a scale from 1 to 5 where** *1 is poor* **and** *5 is excellent.* **Also please fill in the last column ranking how important the skill is for solving this particular problem for anyone.** *1 is not useful* **and** *5 is critical* **for solving the problem.**

	Jasmine	Sara	Useful for this problem.
Could pick out important facts and details.			
Remembered useful information			
Math skills			
Figured out what to do next			
Stopped to check on progress every so often			
Made sense of each thing as they progressed		`	
Was able to bring facts together to understand the big picture.			
Wanted to solve the problem because interested in it.			
Wanted to solve the problem for the job.			
Confidence in problem solving ability			
Worked hard to solve the problem			
Overall problem solving ability			

44. **Who's the better problem solver? Who would you hire and why?**

45. **Were you as the solver interested in the problem of saving the Eagle?**

46. **Would you like to solve a problem like this one again?**

47. **Did you enjoy analyzing the two interns?**

48. Would you have preferred to solve it on your own?

49. Do you work better on your own or in groups? Why?

50. Did you find yourself looking ahead at what Jasmine and Sara were doing before answering the questions about their conversation?

51. How many ideas or facts did they point out that you had overlooked or not thought of yet?

    a. What were the items that they pointed out?

52. Did you think about your own solution as you evaluated the interns?
    a. Yes, I couldn't resist
    b. Oh, I could resist.

Explain:

53. Did you feel that you personally did a good job solving this problem?

54. Were you in a hurry and/or felt that you could have done a better job solving this problem in a different situation?

55. What parts of problem solving do you consider yourself good at and what parts do you enjoy?

56. Which parts of problem solving do you not like and/or are not as good at?

57. Do you think there's a better solution than what you've come up with?

58. Do you think the author's of the Rescue at Boone's Meadow have a correct solution in mind or is this the type of problem that has more than one correct answer? Is it realistic or did they miss some important ideas?

Appendix B - Grading Rubric

Problem Solving Rubric (12/4/07)			
	1	**3**	**5**
math - basic add/sub/mult/div	struggles with computations. Works out by hand slowly.		breezes through any calcs that require basic math.
Math - equation formation	Has a terrible time writing down a formula to figure out something like miles per hour if given 15 miles in 1/2 an hour.		easily (seemingly without thought) writes down and uses formulae
Reading comprehension	Answered a different question than what the survey asked. Got different information from the story than was written.		Read easily and responded appropriately without any misunderstandings or misinterpretations.
Spatial – mapping	Considered locations to be in a straight line. Do not realize blocks have three dimensions. Only think of height and length.		Realized locations on survey were in a triangle. When doing pyramid interviews much easier to see with Geometry of pyramid. Most did calculations for a triangle or pyramid shell w/out interior
Previously known facts	Stated knowledge learned elsewhere incorrectly and with confidence several times.		Every piece of knowledge they brought to the interview was accurate.
Real World knowledge	Appears to have never stepped foot out of the building. Does not have any knowledge of factors that could affect a real everyday problem.		Has general knowledge of how things work and are aware that wind resistance could slow things down or the ground is bumpy or gas smells, etc…
Knowledge of own strengths	States that they are good at something that they are not!	Can identify one or two dominant strengthses but not a lot of them.	Is able to accurately identify several of their strengths not just one or two. Gives a fairly cohesive picture of their own PS skills
Knowledge of own weaknesses	States that they are not strong in an area that they are.	Can identify one or two dominant weaknesses but not a lot	Is able to accurately identify several of their weaknesses not just one or two. Gives a fairly cohesive picture of their own PS skills

		of them.	
Number sense (Numbers have meaning)	Will state two numbers in the same sentence and clearly indicate that the numbers have no meaning for them. The plane goes 30 mi/hr so… Planes are faster than cars so it makes more sense to fly.		When they see a number or calculate a number it immediately translates into useful physical meaning that guides their planning.
Estimation (can they do it - knowledge)	When trying to estimate, they could not do it.		Could do estimates quickly and easily as they worked through other more difficult tasks or if prompted.
Ability to analyze interns	Analysis of interns has very little to do with their actual actions		Can accurately describe each intern and her strengths and weaknesses.
Acquires info 1st time through	Keeps looking back at story. Has to check for all facts. Does not remember entire scenario		Reads through once and remembers scenario as well as what facts have been told and specific values of facts in most cases.
Plan ideas (What – ask questions)	Does not know where to start. Can't even put info together into enough coherent thought to create a single question to start with or item to find.		Can think of all kinds of things that one must figure out before solving the problem.
Plan - way to get answer (How)	May have figured out a plan or may see a specific question but have no idea how to get the answer. Doesn't even know what facts or ideas might apply.		Once they are told (or figure out on their own) what is needed, they can easily determine a plan and carry it out (formula etc) to find the answer to that specific question.
Plan - big picture (Visualization)	Looks at problem as bits and pieces with no evidence of time passing by. Thinks the setting is the same before take off as after 2 hours of flying. Can't 'see' past what is specifically stated in the problem and doesn't tie it together into one coherent story.		Visualizes the problem in motion like a movie in their head. This helps them think of important details and to coordinate things in space and time. Tied to 'ties in personal experience'

Keep problem framework in mind	Has a plan on how to solve and what needs solved. Then after working out some bits, forgets the things they needed to get and thinks they are done		Remembers all the parts of the question and even with many calculations needed, keeps moving and does not believe they are done until they actually have answered all the questions
Connects steps and pieces	Can solve for specific items but cannot figure out where or how to use the pieces to find a solution		Puts ideas and solutions to parts of the problem together without effort
Check calculations of others	Does not check any of their own or supplied calculations		Checks all calculations supplied and those of their own.
aware of how others helped	Thinks they thought of everything on their own. Does not give credit at the end after important facts that they used were provided by the interns (or maybe you the interviewer) even if they are asked specifically - what about the gas tank?	Remembers a few things or remembers after being reminded "what about the gas tank" "oh yea, they did point that out to me."	Knows exactly what they forgot or hadn't noticed or hadn't thought of until the problem supplied it.
Meta-process	Had no idea what they had done and never looked back to consider if they were on the right track		Took time at regular intervals to stop and say, am I headed the right direction still? IS what I'm doing useful for solution
Skepticism (pertains to info being delivered to the solver – trust in source) Do they evaluate info given	Believes whatever is told to them. If the character says something different from what they've already figured out, will immediately question themselves and assume they must be wrong and the outside person correct. Will really struggle to find why the character is right and question self before	Selective about what information they trust; however, base this decision on source	Always checks new information based on its reasonableness as well as looking for another source to corroborate the information.

	anything written even from a made up intern in evaluation tool.	rather than content.	
Estimation (Do they do it - process)	Never tried to estimate anything		Estimated values before calculating anything or to check that they're approximately on the right track.
Creativity	Follows one obvious path and cannot think of any other ways or sometimes not even one way without guidance		Can think of other possible solutions or routes to explore. Even outside of the box. The entire trip has been planned around using a plane that can land in a field but thinks of flying to a nearby road and using a car to return seems to require rare creativity.
Adaptability (shift gears if someone else or problem requires a new direction)	A new scenario is presented halfway through problem solution and the student will not consider it even if it makes logical sense. Keeps coming back to question it or outright rejects the change		Student hears new scenario, thinks a bit about facts to see if possible and then quickly integrates new idea into their current plan.
Can throw out useless info	Trys to use everythign just because its' there. Gets lost in detail. Can't identify level of importance		Will find what is needed and ignore residual info.
Judgment of reasonable issues/info - is it material? (Pertains to value of actual info either given or calculated) can they evaluate it effectively?	Will think of a possible detail and dwell on it. Will actually say this could affect the payload of 120 lbs by half a pound. That puts them over and will ruin everything.		When evaluating a possible difficulty with a problem or possible factor that may impact the problem, is able to decide if something will have a material impact

Judgment of importance of number values (is it material?)	Will calculate or find information about a factor that could affect the problem but has no ability to see that a difference in weight of 1/2 pound does not matter when the maximum payload is 210 lbs. Considers this factor that adds ½ a pound to the payload enough to ruin that solution plan.	Is able to judge whether a factor is material. Does it impact the situation enough to be noticed?
Tie in personal experiences	Someone who appears to have no life outside of the classroom. Only uses what has been taught in science or math class. Doesn't even think about the outdoors while doing survey. If asked directly the person has outside experience but is not comfortable using it because they don't feel that they can use anything that was not stated.	Thinks of all sorts of factors that could influence this scenario. This category is closely related to big picture.
Tie in info provided by another	When someone offers a new idea or piece of info, they either outright reject it or just ignore it. Everything must come from the solver on their own. Usually they just don't know how to use info.	When they hear something new they immediately evaluate it's usefulness and figure out how it affects their plans.
Scientific Process (each step justified with evidence not by gut feeling)	Will follow a plan or use an answer because it seems right. Even if it contradicts something else they have determined.	Must have a basis for what they use and each step must follow a path. If something has a problem, they fix it before moving onto the next step.
Remember previously noted facts	Will comment on a fact from the story and later need it and not remember what it is and at times that they even knew that bit of information.	Once they comment on a fact or bit of info they remember it
Remember what s/he has calculated or reasoned	Will calculate a value and need it later but does not remember what it was. Other example is figure out part of a scenario or plan and later need this and not even remember that they had it.	Whatever info they calculate, they know they know and remember what it was.

Confidence	Not sure of themselves or their answers. Always asking for reassurance to the point of interfering with progress.		Very confident in what they do (even when wrong) Never second guessing. Totally comfortable with stating facts or answers
Attribution (who's to blame for their failure)	Everyone else messed them up. Dust in their eye etc..		Takes responsibility for their own shortcomings.
Judgment of information based on the source	Will not believe anything that comes from someone s/he's determined to be stupid. To the point that s/he will refuse to use an idea (even a good one) if it comes from someone he's decided is stupid.		Looks at all new information with an open mind. May consider the value of the source but still evaluates info on its own merit.
Wanting to find the best solution to the problem for self	Not bothered one bit if they can't or don't find the answer. Could be worried about what interviewer thinks but the idea of a solution is not of interest to them at all.	May want to find the solution but doesn't NEED to solve it.	Challenged by the problem either because of the problem scenario or simply the idea of solving a challenge. It will bother them after they leave if not finished
Wanting to find the best solution to the problem for interviewer	Completely uninterested in what the interviewer thought of them.		Really does not want to let the interviewer down. Wanted to demonstrate their abilities for the interviewer
Wanted to Succeed on 'test'	Wanted to leave or just passing time.		Wanted to prove themselves capable
Interested in the context of the problem	Considers the scenario within the problem to be far fetched and/or not important to them personally so not worth their time to think about it.		Finds the scenario within the problem worthwhile.
Enjoyed Solving the problem	Did not like having to solve the problem but could have enjoyed interview just not the problem.		Loved solving it but maybe not analyzing the girls
Enjoyed analyzing interns	Very annoyed by the questions requiring them to think about how the two interns were progressing		Really liked discussing what was going on between the two interns and how they were doing.
Enjoyed complete experience	Some aspect of the interview, either problem, interns or interview made the subject rather not be there.		Enjoyed solving the problem, meeting with the interviewer and analyzing the interns.

Real Life vs. Student	Trapped in the confines of school ideas and the specifics of the problem, observes from the outside and doesn't even consider if or how it could be a real situation doesn't even consider the idea that it could be real and how they would handle it.	Aspects of both	Thinks about things in the big picture as if the problem were real and it matters to someone's life
Careful/Thorough	Answers each question with first thought and does not spend much time considering the ins and outs of the problem.		Takes time with every answer, writes things down and considers each step carefully

Appendix C - Pyramid of Giza Problem - Building an Egyptian Pyramid

Over the years there has been lots of speculation and discussion about the construction of the ancient Egyptian pyramids, including some claims that it would be impossible for the ancient Egyptians or Mayans to do this. Some claim that interstellar visitors must have done it for them. There is also more plausible speculation that the amount of labor required to build them, also led to the downfall of the Mayan civilization, but this depends on guesses as to how much labor was actually required. As you will see, a little physics tells you a great deal about:

1) how to build the pyramids;

2) how many people are required;

3) and how long it will take.

This is information you would have a great deal of trouble figuring out any other way but is surprisingly easy to get using your physics knowledge. The key idea is that you look at the energy issues involved. Knowing that the work is coming from humans and the energy is only the energy lost to friction and the gravitational energy involved in stacking up big chunks of rock can go a long ways.

The Great Pyramid at Giza in Egypt is nearly 153 m tall and 230 m wide at the base. The average block of stone used in building the pyramid has dimensions 0.66 m x 1.00 m x 1.50 m. These blocks were quarried at a site on the Giza Plateau that was probably about 5 km away. The density of the sandstone in this area was about 2000 kg/m^3. So the mass of each block is 2000 kg/m^3 x volume = 2000 kg/m^3 x 1m^3 =2000 kg. The ancient Egyptians transported these blocks from the quarry to the pyramid on wooden sleds. They lubricated the ground in front of the sled's skids to make it easier to pull the blocks. With this lubrication, the frictional force opposing the forward motion of the loaded sled was about 0.3 times its weight. Once they reached the pyramid location, they needed to raise the blocks up to the particular level of the pyramid they were working on at that time. In order to move the blocks to the required heights, the Egyptians built earthen ramps that greatly reduced the slope of the pyramid side.

If we observe how the pyramid is constructed, we find that the blocks are stacked with the 1 m side vertical, and so each block occupies a horizontal area of 1 m^2. The top layer has 4 blocks. As a reminder, the volume of a pyramid is 1/3 x area of base x height, and the volume of a rectangular block is length x width x height.

In this fashion, Egyptian's working for the glorification of their gods and king built the pyramid in 20 years. They were mostly farmers and needed to tend their crops for most of the year so they could only work for 3 months a year when there was nothing to be done on their farms. These workers were typically athletic males who weighed about 140 lbs and had a density of equivalent to water. Most pyramid data obtained or calculated from Encyclopedia Smithsonian (www.si.edu/welcome/faq/pyramid.htm).

Say one person could sustain pushing a box across the floor for ten hours of work at ~ ½ normal walking speed (0.5 m/s). A good average number for an athletic male (as the Egyptian workers were) is 67 lbs of force (or equivalently 67 lbs * 4.45 N / lb = 300 Newtons of force). This gives a power output of 150 Watts (Power = Energy per second; Watts = Joules/second, so 150 watts = 150 J/s). To build a pyramid requires two steps: 1) sliding the blocks from quarry to site of pyramid, and 2) sliding them up the ramp to stack on the pyramid.

Please determine how many workers were required to build a pyramid.

Bibliography

Anderson, L. W., Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching and Assessing*. New York: Longman.

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed), *Advances in instructional psychology: 5*. NJ: Lawrence Erlbaum Associates. (1-34)

Anderson, J. R., Simon, H. A., and Reder, L. M. (1996). Situated learning and education. *Educational Researcher, 25,* 5-11.

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Pertrosino, A., Zech, L., Bransford, J. D. and The Cognition and Technology Group at Vanderbilt (1998). Doing With Understanding: Lessons From Research on Problem- and Project-Based Learning. *The Journal of the Learning Sciences*, 7(3&4), 271-311.

Bunce, D. M. and Heikkinen, H., (1986). The effects of an explicit problem-solving approach on mathematical chemistry achievement. *Ournal of Research in Science Teaching, 23,* 11-20.

Bunce, D. M., Gabel, D. L., & Samuel, J. V., (1991). Enhancing chemistry problem-solving achievement using problem categorization. *Journal of Research in Science Teaching, 28,* 505-521.

Chi, Michelene T.H. (2006) Two Approaches to the Study of Experts' Characteristics in The Cambridge Handbook of Expertise and Expert Performance edited by K. Anders Ericsson, Neil Charness, Paul J. Feltovich and Robert R. Hoffman. (21-30)

Chipman, S. F., Segal, J. W., and Glaser, R. (1985). *Thinking and learning sklls,* Volume 2. Hillsdale, NJ: Erlbaum.

Cognition and Technology Group at Vanderbilt. (1997) The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

The Cognition and Technology Group at Vanderbilt (retrieved 9/20/04) Website http://peabody.vanderbilt.edu/projects/funded/jasper/Jasperhome.html.

Crouch, C.H. and Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics, 69,* 970-977.

Cummings, K. (2005). Personal conversation regarding efforts towards creating a physics problem solving evaluation.

DiLisi, G., Eulberg, J., Lanese, J., and Padovan, P. (2006). Establishing Problem-Solving Habits in Introductory Science Courses. Journal of College Science Teaching, 35(5) 42-47.

diSessa, A. A., Wagner, J. F. (2005) What coordination has to say about transfer. Transfer of Learning from a Modern Multidisciplinary Perspective edited by Jose Mestre. Information Age Publishing 121-154.

Duschl, R. A. and Hamilton, R. J. (1992). *Anchored instruction in science and mathematics: Theoretical basis, developmental projects, and initial research findings.* Philosophy of Science, cognitive psychology and educational theory and practice by The Cognition and Technology Group at Vanderbilt (pgs 244-273). Albany, New York; State University of New York Press.

Dweck, Carol S. (1999). Essays in Social Psychology. Self-Theories; Their role in motivation, personality and development. Psychology Press: Philadelphia, PA.

Ebly, A. (1999). The Idea Behind EBAPS, Retrieved from http://www2.physics.umd/edu/~elby/EBAPS/idea.htm on 11/2/05.

Ericsson, K.A., Charness, N., Feltovich, P.J., Hoffman, R. (2006). The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press; New York.

Ferguson-Hessler, M. G. M. and de Jong, T. (1987). On the quality of knowledge in the field of electricity and magnetism. *American Journal of Physics*, 55(6) (492-497).

Ferguson-Hessler, M. G. M. and de Jong, T. (1986). Cognitive Structures of Good and Poor Novice Problem Solvers in Physics. *Journal of Educational Psychology*, 78 (4) 279-288.

Ferguson-Hessler, M.G.M. and de Jong, T. (1990). Studying Physics Texts: Differences in Study Processes Between Good and Poor Performers *Cognition and Instruction*, 7(1) 41-54.

Finney, Roxi (2003). Research in Problem Solving: Improving the progression from Novice to Expert.

Halpern, D. (1992). *Enhancing thinking skills in the sciences and mathematics.* Hillsdale, NJ: Erlbaum.

Hammer, D. (1989). Two approaches to Learning Physics. *The Physics Teacher, Dec*

Heller, P., Keith, R., Anderson, S. (1992). *American Journal of Physics.* Vol. 60, No. 7. 627-636.

Heller, J. I. and Reif, F. (1984). Prescribing Effective Human Problem-Solving Processes: Problem Description in Physics. *Cognition and Instruction* 1 (2) 177-216.

Hestenes, D., Wells, M. and Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141-158.

Hestenes, D. and Wells, M. (1992). A Mechanics Baseline Test. *The Physics Teacher, 30*, 159-166.

Huffman, D. (1997). Effect of Explicit Problem Solving Instruction on High School Students' Problem-Solving Performance and Conceptual Understanding of Physics. *Journal of Research in Science Teaching*, Vol. 34, No. 6. 551-570.

Kohl, P. (2004) *Research into Teaching Problem Solving Skills in Physics.* Comps II Exam.

Leonard, W. J., Dufresne, R. J., and Mestre, J. P. (1996). Using qualitative problem-solving strategies to highlight the role of conceptual knowledge in solving problems. *American Journal of Physics*, 64, 1495-1503.

Mayer, R. E. (2003). *Learning and Instruction*. New Jersey: Merrill Prentice Hall.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science* 26, 49-63.

Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2^{nd} ed.). New York: Freeman.

Mayer, R. E. and Whitrock, M. (2006). *Ch. 13 Problem Solving.* Handbook of Educational Psychology, Second Edition. Edited by Patricia A. Alexander and Philip H. Winne. Erlbaum; New Jersey (287-303)

Mayer, R. E. and Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner and R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 47-62). New York: Macmillan.

Maloney, D. P. (1993) Research on Problem Solving in Physics. *Handbook on Teaching and Learning*

Palanscar and Brown (1984). Reference came from Anderson and Schunn book

Pressley, M. and Burkell, J. (1990). *Cognitive strategy instruction that really improves children's academic performance.* Brookline Books; Cambridge, MA.

Pretz, J. E., Naples, A. J. and Sternberg, R. J. (2003). Recognizing, Defining, and Representing Problems in *The Psychology of Problem Solving* edited by Davidson, J. E. & Sternberg, R. J. Cambridge University Press; New York (3-30).

Ross, B. (2007) Problem Solving and Learning for Physics Education. Presented at the 2007 Physics Education Research Conference.

Sabella, M. S. (1999). *Using the Context of Physics Problem Solving to Evaluate the Coherence of Student Knowledge.* Retrieved April 4, 2003, from University of Maryland, Physics Education Web site: http://www.physics.umd.edu/perg/dissertations/Sabella/

Schoenfeld, A. H. (1987). What's All the Fuss About Metacognition? *Cognitive Science and Methametics Education.* Hillsdale, NJ: Erlbaum.

Sherwood, R. D., Kinzer, C. K., Bransford, J. D. and Franks, J. J. (1987). Some Benefits of creating macro-contexts for science instruction : Initial findings. *Journal of Research in Science Teaching*, 24(5) 417-435.

Sternberg, R.J. (1995). Conception of expertise in complex problem solving: A comparison of alternative conceptions. In P.A. Frensch & J. Funke (Eds.), *Complex problem Solving: The European Perspective* (pp. 295-321). Lawrence Erlbaum Associates: Hillsdale, NJ.

Sternberg, R.J. (1985). *Human abilities: An information-processing approach.* W.H. Freeman; New York.

Schwartz, D. L., Bransford, J. D. and Sears, D. (2005). Efficiency and Innovation in *Transfer Transfer of Learning from a Modern Multidisciplinary Perspective* edited by Jose Mestre. Information Age Publishing; North Carolina (1-52).

Thacker, B., Kim E., and Trefz, K. (1994). Comparing problem solving performance of physics students in inquiry-based and traditional introductory physics courses. *American Journal of Physics, 62*, 627-633.

Thornton, R. K. and Sokoloff, D. R. (1998). Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics, 66,* 338-352.

University of Maryland Physics Education Research Group (2002). Activity-Based Physics (ABP) Alternative Homework Assignments. retrieved from: http://www.physics.umd.edu/perg/abp/aha/index.html

Wineburg, S. (1998). Reading Abraham Lincoln: An Expert/Expert Study in the Interpretation of Historical Texts. *Cognitive Science*, *22* (319-346).