CONSTRUCTION AND VALIDATION OF AN INSTRUMENT TO MEASURE PROBLEM-SOLVING SKILLS OF SUBURBAN HIGH SCHOOL PHYSICAL SCIENCE STUDENTS

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Patrick James Herak, M.S.

Graduate Program in Education

The Ohio State University

2010

Dissertation Committee:

Arthur L. White, Advisor

Donna F. Berlin

David J. Horn
Copyright by

Patrick James Herak

2010
Abstract

The purpose of this study was to develop a problem-solving instrument that could easily be used by a classroom teacher. The research questions were (1) can the Problem-Solving Skills Assessments (PSSAs) differentiate between students with varying levels of selected problem-solving skills? (2) Can the PSSAs measure student growth due to problem-solving events and problem-solving units? and (3) Does varying types of the assessment activities provide similar or consistent results?

This study used multivariate analysis of variance with repeated measures to detect significant changes over time and a to look for significant differences between cohorts and interaction effects. There were significant differences ($p=0.000$) in assessment detected by the PSSA for overall problem-solving ability as well as the four problem-solving skills: problem definition, problem planning, plan execution, and problem evaluation. Additionally, there were significant interactions for assessment by cohort for three of the problem-solving skills: problem definition, plan execution, and problem evaluation. Findings also
suggest that in many instances, the PSSA is able to detect significant differences over time, and significant differences between cohorts.
Acknowledgments

A great deal has gone through my mind as I finish this 20-year journey through higher education at The Ohio State University (OSU) that includes 299-hours of undergraduate credit and 184-hours of graduate credit (not to mention the few miscellaneous hours from other universities). The first thing that comes to mind is, what a journey it has been. I have grown so much personally and professionally over these 20-years due to family, friends (many of which have become family), and staff at The Ohio State University.

The other thing that comes to mind is that I am not overly emotional; many times the appreciation I have for others does not come across. I want to assure you that if your name comes across in these acknowledgements, I don’t really have enough ways to express my thanks for you individually. Also, if you are affiliated with any of the groups mentioned (including friend, family, etc.) and are not mentioned individually, you were still part of the incredible support systems I am very fortunate to have had that made this degree possible, and you are not overlooked!

-iv-
Initially, I was somewhat reluctant to attend The Ohio State University because my dad had attended there. Many teenagers tend to rebel from their parents (I don’t believe I did too much), but my attempt to attend other universities was arguably my greatest defiance from the sound parental guidance I had.

Fortunately, I did decide to come to OSU and have gained an understanding of what an incredible educational resource I have had all these years. I was very fortunate my freshman year to become a member of The Ohio State University Marching Band. I can’t begin to tell you how appreciative I am of this organization and their Alumni (especially Dr. Woods, John Waters, and Dr. Paul Droste, whom I will talk more about later). It not only gave me a sense of belonging in a very large campus, but also provided me with friends, family, and a support system that I will surely enjoy into the future.

As an undergraduate, I spent 2.3-years in the College of Engineering before switching to the College of Education (as a side note, my grade-point average actually dropped as I went into education, a somewhat notable fact for those who look down on teachers). It was at that point that Dr. White took me in as one of his advisees. I remember him observing me student teaching like it was yesterday. It seems hard to believe that it is some 18 or so years later and we are still sharing some good discussions about education, teaching, and learning (I hope these continue even into the future). The mentorship Dr. White has given me has been incredible, and I thank him for that.
As a graduate student, I chose the path less taken by most science teachers; I chose to get a Master’s of Science degree. I don’t think that I foresaw obtaining a PhD, but I did know that down the road, some employers/universities looked down upon students who received back-to-back degrees from the same college. I had some great teachers throughout my master’s program (Dr. Krissek, Dr. Fortner, and Dr. Roth come to mind) including a great advisor in Dr. Horn. It was with the various moth sorts and field studies with Dr. Horn (and one of his assistants Foster Purrington) that I felt I really learned “science” (and statistics) for the first time. I felt very fortunate to have the collegial environment, and the candidness of the lab on the top of the 12th avenue parking garage.

When I enrolled in the PhD program, it truly was not my idea. Our school district does not offer any pay incentives for a PhD but does for a Masters +30 semester hours. I realized 30-semester hours would get me another Master’s Degree and talked to Dr. White about a Master’s in Education. I was quickly talked out of another Master’s Degree and enrolled in the PhD program at OSU. I can’t begin to tell you how many times I wish I had taken the easy way out and obtained another Master’s Degree, but I can tell you how thankful I am to Dr. White and Dr. Berlin for pushing me along.

And why have I not mentioned Dr. Berlin until now? Perhaps because of the fogginess of my undergraduate switch to education. I think I must have had her as one of my professors (especially with how closely she has worked with Dr. White over the years), but I can’t find a record of it. Of course, had she been one of my teachers, wouldn’t I most certainly have remembered her distinct, New
York accent? Regardless, Dr. Berlin (and for my Master’s degree Dr. Boerner) provided me with one of the most important things any graduate student needs … accountability. My Master’s defense and PhD candidacy exam were both very difficult experiences for me because I was held accountable. It is difficult to describe how much more accomplished I feel because of the standards I have been held to by my committee members and I would like to thank all of them for that.

Since my dissertation work took place over the past four plus years I would be remiss if I did not thank the students, staff, and administrators at Westerville North High School. Without the cooperation and support of Kurt Yancey, Laura Ferguson, Jen Kirk, and Damon Mollenkompf, this would not have been possible. Also, I feel incredibly privileged to work with an incredible science department that includes one of my dearest friends and professional colleagues, Kyle Campbell.

As for the last and greatest group of people to thank, it must be my family. Obviously I’ve had my genetic family my entire life! Over the years, I’ve learned a great deal from both my mother and father that have made me both a better scientist and a better teacher. I think my father taught me a great deal about responsibility, pragmatism, and accountability (in fact, I distinctly remember “pragmatic” being a vocabulary word in high school and being the only one who knew what it meant). He always showed a great deal of respect for university education and has supported me to no end in that regard. In fact, as I’ve neared the conclusion of this degree, my dad’s brothers and sister (Jim, Rick, and Janis)
have shown an incredible level of support and value for the degree I am about to be awarded.

My mother, on the other hand, has been a great resource in the behaviorist side of education. I can’t imagine being a teacher (or a scientist for that matter) without being able to look at things from other perspectives. Likewise, I feel fortunate to be able to look at classroom behavior problems from the perspective of a behaviorist (positive/negative reinforcement, punishment, reward, etc.) rather than simply a teacher. Additionally, in 1982, my mother married Dr. Melvin Arnoff. I can’t even begin to describe the impact of knowing somebody well that has earned an advanced degree. Mel would end up being one of many in my life that had earned a PhD, but will always be the first. Many people in this country do not know anybody who has earned an advanced degree, so I feel quite fortunate.

Lastly, I wish to thank my “in-laws.” Between the six of us (my wife, her two parents, her brother, and our sister-in-law), we will (at the acceptance of this document) have collectively, six Bachelor’s Degrees, six Master’s Degrees and two PhDs (and we may not be done). I don’t know anybody who has appreciated teachers more than this family. Many times, the love and concern they have shared could be mistaken for “nagging,” but as I tell my students, “your family wouldn’t care unless they were on your back.”

My wife Diana also deserves a special mention. To make great strides, I often needed her to either pick up additional household responsibilities or just
“get out of my way.” She has been very patient, and may very well be happier than I am that this is done.

To summarize, I feel quite fortunate by the people I have been surrounded by over these past 20-years (and the family that has been there my entire life). As a tribute to you all, I will do my best to keep on being a scientist and teacher.
Vita

September 27, 1973 ...................................... Born – Dover, New Hampshire

June 1991 ................................................ Graded from Orange High School
Pepper Pike, Ohio

1991-1996 ................................................ B. S. Ed. Science Education,
The Ohio State University,
Columbus, Ohio

1996-1997 ................................................ Science Teacher, Bloom Carroll
High School, Carroll, Ohio

1997-2004 ................................................ Science Teacher, Westerville
City Schools, Westerville, Ohio

August 2001 ................................................ M.S. Environmental Science,
The Ohio State University,
Columbus, Ohio

2004-2005 ................................................ Physics Teacher, Aldenham
School, Elstree, Hertfordshire,
United Kingdom

2005-2006...........................................................Graduate Teaching Assistant, The Ohio State University, Columbus, Ohio

2006-present.....................................................Science Teacher, Westerville City Schools, Westerville, Ohio

Publications


Fields of Study

Major Field: Education

Special Area: Science Education
# Table of Contents

Abstract ............................................................................................................. ii  
Acknowledgments ............................................................................................ iv  
Vita ................................................................................................................... x  
Publications ..................................................................................................... xi  
Fields of Study ................................................................................................. xi  
Table of Contents ............................................................................................. xii  
List of Tables .................................................................................................... xviii  
List of Figures .................................................................................................. xx  
1. Introduction ................................................................................................. 1  
   Rationale ...................................................................................................... 2  
      A fundamental need .............................................................................. 2  
      Maturation and independence ......................................................... 3  
      A global economy ............................................................................. 5  
      Technology .......................................................................................... 6  
   What Are Problems? ............................................................................... 7  
   Lack of Research .................................................................................... 7
Definition of Terms ................................................................. 8
Measurement of Variables ....................................................... 11
  Independent variables ....................................................... 11
  Dependent variables ......................................................... 12
Problem Statement ............................................................... 12
Hypotheses ........................................................................... 12
Delimitations, Limitations, and Assumptions .......................... 13
2. Literature Review ............................................................ 14
Problem Solving ............................................................... 14
Expert Versus Novice Problem Solvers ................................. 17
  Knowledge and organization .............................................. 19
  Strategies ........................................................................... 19
  Self-justification and self-reflection ................................... 20
  Heuristics and problem-solving steps ................................. 20
Aiding the Novice ............................................................... 21
Social Constructivism .......................................................... 21
Scaffolding ........................................................................... 23
Problem Solving and Distance Learning ................................. 24
Literature-Based Decisions .................................................... 31
  Social constructivism ....................................................... 31
  Scaffolding instruction .................................................... 31
  Student maturation .......................................................... 33
  Problem solving outside the classroom ............................... 33
3. Materials and Methods........................................................................................................ 35

Pilot Study......................................................................................................................... 35

Likert-type instrument................................................................................................. 36

Authentic assessment ................................................................................................. 37

Sample Population....................................................................................................... 39

School data .................................................................................................................... 39

High School Transition............................................................................................... 40

Social............................................................................................................................. 40

Academic....................................................................................................................... 41

Smaller Learning Communities.................................................................................. 42

Goals.............................................................................................................................. 42

Concerns......................................................................................................................... 43

Team data....................................................................................................................... 44

Class selection............................................................................................................... 46

Participation................................................................................................................... 48

Cross-teamed students................................................................................................. 49

Data Collection in the Current Study......................................................................... 50

Measurement of Variables......................................................................................... 51

Independent variables.............................................................................................. 51

Cohort............................................................................................................................. 51

Problem-solving activities and units........................................................................... 52

Instructional procedures and instructional content.................................................... 53

Dependent Variables................................................................................................. 57
6. References......................................................................................................................... 109

7. Appendix

   A. Classical Assessment................................................................................................. 115
   B. Supplemental Assessment........................................................................................ 117
   C. Pearson Correlation Coefficients to Determine Inter-Rater Reliability..................... 119
   D. Repeated Measures for Problem-Solving Ability .................................................. 121
List of Tables

Table 3.1. Problem-solving activities for each subskill and course ............... 57
Table 4.1. Multivariate analysis of variance by cohort by assessment with repeated measures-between-subject effects .......................................................... 72
Table 4.2. Multivariate analysis of variance by cohort by assessment with repeated measures-within-subject contrasts ..................................................... 72
Table 4.3. Means (M) and standard deviations (SD) for problem-solving ability by assessment by cohort and overall across assessments and cohorts ........................................................................................................ 73
Table 4.4. Multiple comparisons for problem-solving ability mean scores by cohort ............................................................................................................. 76
Table 4.5. Multivariate analysis of variance of problem-solving skills: problem definition, problem planning, plan execution, and problem evaluation assessments with repeated measures by cohort ...................... 78
Table 4.6. Univariate analyses of variance of problem definition, problem planning, plan execution, and problem evaluation problem-solving subskill by cohort between-subject effects .................................................. 78
Table 4.7. Univariate analyses of variance of problem definition, problem planning, plan execution, and problem evaluation problem-solving subskill by cohort within-subject effects........................................................................ 79
Table 4.8. Mean (M) and standard deviation (SD) for problem-definition skills by assessment by cohort and overall across assessments and cohorts............................................................................................................. 80
Table 4.9. P-values for T-tests, with corrections for multiple comparisons, comparing problem-definition mean scores by cohort by assessment......... 82
Table 4.10. Mean (M) and standard deviation (SD) for problem-planning skills by assessment by cohort and overall across assessments and cohorts............................................................................................................. 84
Table 4.11. Mean (M) and standard deviation (SD) for plan-execution skills by assessment by cohort and overall across assessments and cohorts........ 86
Table 4.12. P-values for T-tests, with corrections for multiple comparisons, comparing plan-execution mean scores by cohort by assessment......... 89
Table 4.13. Mean (M) and standard deviation (SD) for problem-evaluation skills by assessment by cohort and overall across assessments and cohorts............................................................................................................. 90
Table 4.14. P-values for T-tests, with corrections for multiple comparisons, comparing problem-evaluation mean scores by cohort by assessment......... 92
List of Figures

Figure 1.1. Problem-solving skills/subskills..................................................... 10
Figure 2.1. Various models of problem-solving skills. ..................................... 16
Figure 2.2. Problem-solving flow chart............................................................ 18
Figure 3.1. The core courses which students were enrolled in as part of Team B’s four cohorts...................................................................................... 36
Figure 3.2. Problem-solving skills and subskills from preliminary study........ 38
Figure 3.3. Distribution of class sections among four teams at a Midwestern suburban high school … .................................................................................. 45
Figure 3.4. Organizational hierarchy ............................................................... 46
Figure 3.5. Time series design........................................................................ 50
Figure 3.6. Problem-solving activities and the corresponding skills and subskills addressed in the team-taught Global Studies/ Language Arts class . 54
Figure 3.7. Problem-solving units and activities and the corresponding skills and subskills addressed in Mathematics courses ............................................ 55
Figure 3.8. Problem-solving units and activities and the corresponding skills and subskills addressed in Physical Science II........................................... 56
Figure 3.9. Independent and dependent variables in the main study ............. 62

Figure 3.10. An example of independent and dependent variables in the physical science course ................................................................. 62

Figure 4.1. Change in mean problem-solving ability scores by assessment by cohort and cohorts combined ......................................................... 73

Figure 4.2. Display of overall mean assessment differences ....................... 74

Figure 4.3. Display of overall mean assessment differences by cohort .......... 75

Figure 4.4. Comparing means and standard deviations for the overall problem-solving ability scores across the four assessments by cohort .......... 75

Figure 4.5. Change in mean problem-definition scores by assessment by cohort and cohorts combined ............................................................. 81

Figure 4.6. Display of mean problem-definition skill score differences by assessment by cohort ................................................................. 82

Figure 4.7. Change in mean problem-planning scores by assessment by cohort and cohorts combined .............................................................. 84

Figure 4.8. Comparing mean scores and standard deviations across the four assessments for problem-planning skill scores by cohort ............. 85

Figure 4.9. Change in mean plan-execution scores by assessment by cohort and cohorts combined ............................................................... 87

Figure 4.10. Display of mean plan-execution skill score differences by assessment by cohort ................................................................. 89
Figure 4.11. Change in mean problem-evaluation scores by assessment by cohort and cohorts combined.......................... 91

Figure 4.12. Display of mean problem-evaluation skill score differences by assessment by cohort................................................................. 92

Figure 5.1. Subskill scoring rubric with criteria from current study .......... 99

Figure 5.2. Subskill scoring rubric with recommendations implemented........ 102
Chapter 1: Introduction

Following World War II, the United Nations (UN) was formed to prevent further wars between countries and provide a center for dialogue in solving world problems (United Nations, 2009a). Shortly thereafter, the United Nations wrote The Universal Declaration of Human Rights, 30 rights and freedoms that should be available to all human beings (United Nations, 2009b). Numbers 18 and 19 in the declaration included the rights to free thought, opinion, and expression. These are three very powerful entitlements that carry with them great responsibility.

What are the rights to thought, opinion, and expression without the abilities to think critically, evaluate thought, justify opinions, express one’s thoughts in a logical manner, and the ability to problem solve? One might say it is nothing more than ignorance. It is very doubtful that when writing The Universal Declaration of Human Rights, the United Nations (2009b) felt everybody had the right to be ignorant.

In fact, later in the charter, Article 26 (United Nations, 2009b) states that all humans have the right to a basic education, so clearly ignorance was not on
the agenda. Furthermore it states that humans have the right to higher
education, with the stipulation “based on merit.” This goal of higher education
cannot be possible, and merit can not be achieved, without problem-solving
skills.

Rationale

There are many reasons to study problem-solving skills; however, the
fundamental principle is that students should be informed. Students need to be
assisted in making decisions that will allow them to mature into autonomous
adults and become productive members of society.

A fundamental need. Having knowledge in a specific content area is not
enough to be successful. Even with a solid background in domain knowledge,
without good problem-solving strategies, students may not be able to draw upon
this knowledge (Chi, Bassok, Lewis, Reimann, & Glaser, 1989) leaving all of this
information isolated and inert.

Furthermore, Linn (1995) states that without problem-solving skills, any
new knowledge gained may be temporary and subject to prior conceptions. That
is, if a student lacks the problem-solving skills to evaluate and justify a new claim,
they will soon forget it and revert back to prior, incorrect facts. Student ability to
manage this information in order to solve problems is essential (Chinnapan &
Lawson, 1996).
Reflecting back on the UN charter, there is a certain amount of empowerment that can be given to people by teaching them to solve problems. Our national standards in both mathematics and science reflect this fundamental need.

The National Council of Teachers in Mathematics (NCTM, 1989, 2000) has stated for many years that all students should be able to receive an education with an emphasis on problem solving. In the 1996 National Science Education Standards, the National Research Council (NRC, 1996) got even more detailed, citing the need for scientific inquiry, exploration, the ability to analyze and synthesize data, and the ability to communicate this information. Even more recently the American College Testing Program (ACT, 2010) has emphasized many problem-solving subskills including devising procedures and evaluating data in their college readiness standards. The need for all of these skills is ultimately so that students can understand and act rationally on personal and social issues.

**Maturation and independence.** Problem solving has a much greater importance than just gaining content and process knowledge in a particular subject area. Many of the skills needed to be a successful problem solver are the same as those needed to become a successful student and responsible adult.

Perhaps one of the marks of a truly successful student is autonomy. Autonomy can be thought of as the ability to be a self-sustaining or self-sufficient
learner. Students can receive help, but an autonomous student is one who seeks the help and knows which questions to ask.

As students progress through secondary school and higher education, Master’s and Doctoral degrees are based upon theses and dissertations representing true autonomy. The ability to take one’s own topic and see it through to a point of new knowledge gained represents autonomy at the university graduate level.

Two key problem-solving skills that can lead to autonomy are the ability to subdivide large goals into smaller parts, and the ability to evaluate and monitor progress related to these parts. Linn (1996) found that computers can help students by organizing information and teachers can help students by administering frequent guidance. Aspy, Aspy, and Quimby (1993) found that students that were in an environment emphasizing problem-solving skills compared to just “covering the material” not only mastered content equally well, but also became more autonomous learners.

Autonomy need not apply to merely an academic setting. Becoming a self-sufficient adult and productive member of society can also represent autonomy. Elias and Tobias (1990) state that how children become adults depends on their ability to judge critically and make decisions, both of which are skills associated with problem solving. They believe it is worthwhile to teach social decision making to children in order to improve student thinking, their chance of becoming more successful members of society, and of reaching important life goals. Eccles, Midgley, Wigfield, Buchanan, Reuman, Flanagan,
and Iver (1993) also recognize that if students learn skills that promote autonomy they will become more independent and reach higher levels of social maturity.

**A global economy.** Producing autonomous adults should be a minimum required by the educational system. On the higher-achieving end of the spectrum, educators must produce the world’s next set of leaders to deal with the world’s problems. The future of our planet depends on students that can deal with the multitude of problems left by their adult predecessors.

Today’s problems (global climate change, energy and natural resource depletion, world poverty, global pandemics, etc.) are incredibly complex and may be “unsolvable” given limited knowledge, funding, and time, although sometimes the best or most cost-effective solution may be to abandon the problem until a later date. In order to even approach these problems, society requires leaders that can look at not only the fine details, but also the big picture. There is a great deal of concern that schools today are not teaching students how to interrelate multiple content areas and see the “big picture” (Ward & Lee, 2002).

With the international explosion in communications due to satellites, the internet, and mobile phones, nations have not only had to adapt to a global economy for natural resources, but now the service industries as well. The United Kingdom is shifting away from traditional curriculum to help prepare students for adult life in an ever-global economy (Garner, 2007). In countries that are based upon service economies there has been a great shift in curriculum towards thinking critically, communicating, and problem solving (Kozma, 2005).
Not only does the literature show that there is an increased demand for workers with higher-order thinking and problem-solving skills (Meier, Hovde, & Meier, 1996), but on the reality show, The Apprentice (Burnett, 2007), Donald Trump stated that problem-solving skills is one of the primary skills that the next apprentice should have.

**Technology.** Since the early 1990s, there has been a steady increase in the use of computers in the classroom. From word processing to spreadsheets to the use of the internet and smartboards, technology is becoming an ever increasing part of the classroom environment.

The use of computers can increase student problem-solving skills. More specifically, computers can aid students in organizing information and becoming more reflective (Linn, 1996). Additionally, the internet can provide a collaborative environment in which students can construct knowledge together by problem solving (Holt, Kleiber, Swenson, Rees, & Milton, 1998).

Not only can technology be used to improve problem-solving skills, but the reverse is also true. Problem-solving abilities are critical to achieve technological literacy (Custer, Valesey, & Burke, 2001). It becomes quite evident that without a solid foundation in problem-solving skills, students will not only be left out of the technological revolution, but will be unable to reap the rewards of it as well.
What Are Problems?

One of the more accepted definitions of a ‘problem’ comes from Hayes: “Whenever there is a gap between where you are now and where you want to be, and you don’t know how to find a way to cross that gap, you have a problem” (Hayes, 1989, p. xii). One of the key concepts in Hayes’s definition is that it requires the solver to be initially ignorant of the method to solve the problem.

There are examples in the literature of situations whereby the known path to solution is quite simple and straight-forward, at least for experts in that content area. These situations are often referred to as algorithms or exercises, rather than problems (Arcavi & Friedlander, 2007; Boero & Dapueto, 2007; Maloney, 1994; Taconis, Ferguson-Hessler & Broekkamp, 2001). In this research the term problem will not refer to exercises or algorithms, but rather situations for which the path is not clear and the answer is not set.

Lack of Research

As a general topic, much has been written about problem solving since Polya’s groundbreaking work in 1945. Even 30-years later Heppner & Peterson (1978) noted a deficiency involving the empirical testing of problem-solving skills. Furthermore, in Maloney’s (1994) review of a more recent review of empirical studies, most all the studies examined were based on exercises and algorithms rather than true problems. Without substantial literature in how to measure student problem-solving skills, it is quite difficult to develop the best methods of intervention to remedy the situation. This identification cannot solely be for the
educational research, but there must be a way for the classroom teacher to
document student problem-solving skills so that student needs can be addressed
in the classroom.

Research in problem solving was relatively stagnant until this decade.
However, even with a recent surge in research of problem-solving skills (Boero &
Dapueto, 2007; Curtis & Denton, 2003; Hovardas & Konstantinos, 2006), none of
these studies have looked at implementing assessment from the perspective of
the classroom teacher. Classroom teachers have limited resources; they only
have one assessor, and they have limited time.

This study hopes to evaluate a method that could facilitate assessment by
the everyday classroom teacher. This study also addresses both the issue of
authentic identification of proficiency in specific problem-solving subskills and the
manner with which the classroom teacher can easily assess deficiencies in these
areas.

**Definition of Terms**

Although there have been several different ideas as to what constitutes
authentic assessment (Darling-Hammond & Snyder, 2000; Rennert-Ariev, 2005;
Williams, 1999), the following key points will be used for this study:

1. Assessment must be conducted in the context of experience-based
   work.

2. Assessment provides for student reflection (justification of their
decision making).
Problem-solving skills and subskills are the processes used to reach the solution of a problem. Polya (1957) identifies four major problem-solving skills. These skills will be used as the basis for this study: however the wording has been modified:

1. Problem definition – identify the task at hand
2. Problem planning – devise a method to solve the problem
3. Plan execution – observe the method in action and collect data
4. Problem evaluation – review the results and look for generalizations

Although Polya (1957) was one of the first to produce significant writings in the field of problem solving, there has been no consensus in the literature as to what constitutes good problem-solving skills and subskills. In an attempt to fit the variety of problem-solving skills and subskills found in the literature into one convergent model, the researcher has taken these various subskills and merged them into Polya’s four skill areas (a summary of these skills and subskills is provided in Figure 1.1).

1. Problem definition – a subject that demonstrates these skills should be able to show evidence that they have identified possible topics, selected their topic, and can justify why they selected their topic.
2. Problem planning – a subject that demonstrates these skills should have the ability to identify variables, isolate variables, brainstorm and then choose appropriate materials, and lastly, develop questions and envision the end product.
<table>
<thead>
<tr>
<th>Problem Definition</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1-Brainstorm possible topics</td>
<td>A2-Identify topic choice</td>
</tr>
<tr>
<td>A3*-Justify topic choice</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem Planning</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B1-Identify variables</td>
<td>B2*-Isolate variables</td>
</tr>
<tr>
<td>B3-Brainstorm/ identify appropriate materials</td>
<td>B4-Choose appropriate materials</td>
</tr>
<tr>
<td>B5-Brainstorm/ identify appropriate procedures</td>
<td>B6*-Choose appropriate procedures</td>
</tr>
<tr>
<td>B7*-Questions developed regarding end product</td>
<td>B8*-End product envisioned</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plan Execution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-Appropriate data collected</td>
<td></td>
</tr>
<tr>
<td>C2*-Appropriate method of data presentation selected</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem Evaluation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D1*-Find errors in data</td>
<td>D2*-Find cause for errors in data</td>
</tr>
<tr>
<td>D3*-Find solution for errors in data</td>
<td></td>
</tr>
<tr>
<td>D4*-Evidence used to support results</td>
<td></td>
</tr>
<tr>
<td>D5*-Validity of results questioned</td>
<td>D6*-Validity of results supported or refuted</td>
</tr>
<tr>
<td>D7*-Plan next step</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.1. Problem-solving skills/subskills (novice in white, intermediate in light gray, and expert in black). *Evaluated in this study
3. Plan execution - a subject that demonstrates these skills should be able to show evidence that they have collected appropriate data and selected appropriate data presentation techniques.

4. Problem evaluation – a subject that demonstrates these skills should be able to show evidence that they have found errors in data, their causes, and possible solutions. They should be able to use evidence to support their conclusion. The validity of their conclusion should be questioned and either supported or refuted. Lastly, there should be evidence that they are planning their next step.

In this study, the term problem-solving ability is used to describe student scores in problem-solving subskills and therefore problem-solving skills. The problem-solving ability score will be determined by averaging the scores for the 12 higher-level and 1 intermediate problem-solving subskills.

Measurement of Variables

Independent variables. The independent variables in this study involve two areas, instruction and student classification. Instruction of students with short term problem-solving activities and longer term problem-solving units were expected to positively affect problem-solving skills. Additionally, students were grouped into four cohorts. These cohorts were expected to demonstrate different levels of problem-solving skills.
Dependent variables. Thirteen problem-solving subskills were examined during the study to evaluate 4 problem-solving skills. Additionally, overall problem-solving ability, determined from these skills and subskills were examined. Two different types of assessments were used to collect information regarding ability, skills, and subskills.

Problem Statement

This study examined problem-solving skills of suburban ninth-grade students who were members of a smaller learning community (SLC). This study addressed the following questions:

1. Can the Problem-Solving Skills Assessments (PSSAs) differentiate between students with varying levels of selected problem-solving skills?
2. Can the PSSAs measure student growth due to problem-solving activities and problem-solving units?
3. Does varying the types of assessment activities provide similar or consistent results?

Hypotheses

The purpose of this study is to develop an appropriate instrument to assess student problem-solving skills. Therefore, the following hypotheses are in order:
1. The PSSA will be able to differentiate between varying levels of student problem-solving skills. In this study, that included honors, regular, and remedial students.

2. The PSSA will show significant differences over time, due to increased problem-solving ability.

3. Because all of the problems were novel, with no known/obvious solution, it is expected that results would be consistent across the four problem-solving assessments. That is there should be no evidence of significant drops in problem-solving skills over time.

**Delimitations, Limitations, and Assumptions**

In addition to the independent variables addressed in the study, there were several other variables that could affect problem-solving skills. Student maturation over the course of the study could improve problem-solving skills; however, due to normal development level and the short time of the study, this variable was not addressed. The transition to high school and a new building was of some concern; however, the school in this study implemented Smaller Learning Communities (SLCs) to help minimize this issue. Distance learning and student work outside of class could potentially reduce the evidence of problem-solving skills the teacher could observe. An attempt to control this variable was made by providing an online forum for students to collaborate outside of class. Lastly, difficulty of assessment tasks could vary; However, the researcher attempted to make them equal in difficulty and operated under this assumption.
Chapter 2: Literature Review

Problem Solving

Writers on the topic of problem solving are like most pundits, quick to criticize and offer their point of view, but very few are actually willing to do something about it. In other words, there is a great deal of literature regarding what makes good problem solving, what steps should be taken, etc., but very little on how to actually measure problem-solving skills.

The ground-breaking moment in problem solving came when George Polya (1957) wrote his book *How To Solve It*. Polya not only outlines various steps involved in mathematical problem solving, but gives various examples and solutions to problems and various strategies for solving mathematical problems. In his book, Polya notes that some of the problems and strategies are specific to mathematics, what we now call domain specific. However, given how long it had taken for significant writing regarding problem solving to come along it was obvious he was well ahead of his time.

Since then there have been many different models and opinions as to what constitutes basic problem-solving skills (see Figure 2.1). The goal of the
researcher was to try to synthesize all of these models into one model that covered the main points found in the literature, and everything comes back to a model that mirrors Polya’s: Problem definition (understanding the problem), problem planning (devising a plan), plan execution (carrying out the plan), and problem evaluation (looking back).

Heppner and Peterson (1978) tried to assess problem-solving skills using a Likert-type instrument. In their model, there were five main steps: General orientation, problem definition, generation of alternatives, decision-making, and evaluation. After conducting the study, a factor analysis was computed to see if similarities in answers matched the five hypothesized categories. The results did not, but were found to match attitudes (systematic approach behavior, impulsive behaviors, and self-confidence) toward problem-solving more than skills.

The premise of their factor analysis was somewhat flawed. For example, there were several questions regarding “problem definition.” Their hypothesis was that those taking the survey would answer all of these questions similarly. However, what they did not take into account was that there are different levels of understanding when it comes to problem definition and that differences in answers do not necessarily suggest error in the original categories, but differences in levels of understanding (novice, intermediate, and expert).
<table>
<thead>
<tr>
<th>Field</th>
<th>Researchers current model</th>
<th>Researchers previous model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychology</td>
<td>Heppner &amp; Peterson (1978)</td>
<td>Knowledge recall</td>
</tr>
<tr>
<td>Psychology</td>
<td><em>PRE</em></td>
<td>Problem identification</td>
</tr>
<tr>
<td>Special Education</td>
<td>Heppner &amp; Peterson (1978)</td>
<td>Strong content knowledge</td>
</tr>
<tr>
<td>Social Sciences</td>
<td><em>POST</em></td>
<td>and justification skills</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>Elias &amp; Tobias (1990)</td>
<td></td>
</tr>
<tr>
<td>Astronomy</td>
<td>Anderson from</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anderson &amp; Garrison (1995)</td>
<td></td>
</tr>
<tr>
<td>Food Chemistry</td>
<td>Garrison from</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anderson &amp; Garrison (1995)</td>
<td></td>
</tr>
<tr>
<td>Physical Science</td>
<td>Diederen, Gruppen, Hartog, &amp; Voragen (2005)</td>
<td></td>
</tr>
</tbody>
</table>

**Problem Definition**
- Understanding the problem
- General orientation
- Feelings are your cue to problem solve
- Problem definition
- Identify the issue
- Guide yourself with a goal

**Problem Planning**
- Devising a plan
- Generation of alternatives
- Systematic approach behaviors
- Think of as many possible things to do
- Clarification of course requirements
- Exploration of explanations
- Envision end results for each option
- Select your best solution
- Plan the procedure and anticipate roadblocks

**Plan Execution**
- Carrying out the plan
- Critical evaluation and exploration
- Testing of outcomes
- Solve Equations
- Systematic analysis of information

**Problem Evaluation (Review & Extension)**
- Looking Back
- Evaluation
- notice what happened and remember it for next time
- Integration and application to everyday living
- Reflective phase
- Critical reflection
- problem evaluation

<table>
<thead>
<tr>
<th>Not Applicable</th>
<th>Self confidence</th>
<th>Good Attitude</th>
<th>Make answers fit into practical solutions</th>
</tr>
</thead>
</table>

**Figure 2.1. Various models of problem-solving skills.**
An analogy would be hypothesizing that 2+4, 25+84, and 108+265 all fit under the category of addition. However, after giving a test to fourth graders one might discover that those who got 2+4 right also got 8-4 and 2*3 correct. Deciding to group the last three problems simply because fourth graders got all three correct and giving up on the original category does not illustrate sound reasoning. Even though it appeared that evidence from their research found three general problem-solving skills, a review of later research suggests that they found differences between novice, intermediate, and expert problem solvers, which will be further discussed later.

A review of other models of problem solving (Anderson & Garrison, 1995; Ding, 2006; Elias & Tobias, 1990; Shin, et al., 2003) showed a variety of skills numbering from three to nine. In an attempt to identify a common, simple thread, the idea to group these different skills by similarities emerged. The main headings for these categories were labeled problem-solving skills and the individual skills were labeled subskills.

**Expert Versus Novice Problem Solvers**

In order to observe and record the transformation from novice to expert problem solvers, key subskills associated with the expert problem solver must be identified. As you view Figure 2.2, you can follow a path from bold box to bold box. An expert problem solver would complete most all the steps in the problem-solving flow chart. The dashed lines represent the impulsive and incomplete steps of a novice problem solver.
Figure 2.2. Problem-solving flow chart. Problem-solving skills are in gray, sub-skills in white. Shortcuts often taken by novices are represented by bold boxes and dashed lines. * represents subskills assessed in this study
**Knowledge and organization.** Experts organize information better than novices. Much like content knowledge, which can be very deeply understood or only be superficially understood, expert problem solvers have a much deeper and integrated organization to their knowledge structure than novices (Mestre, Dufresne, Gerace, Hardiman, & Touger, 1993).

Additionally, experts tend to have more domain-specific knowledge that allows them to internalize information in a well-organized manner (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981). Expert problem solvers identify major ideas and relate that information to their own knowledge base; thus all knowledge becomes well connected (Lester & Kehle, 2003). This skill allows expert problem solvers to be much more efficient by classifying the problem and thus pulling only relevant information from their domain knowledge (Mestre et al., 1993; Sutherland, 2002).

**Strategies.** One difficulty that novices have is that they tend to be fixated on details rather than on learning the larger concepts. They rely on outsiders to provide hints and make connections for them.

One way to help novice problem solvers is to aid them in identifying the important information in the problem (Sutherland, 2002). Having novices verbalize their problem-solving strategies, a behavior which experts are quite good at (Mestre et al., 1993), can help them make these connections.

Another big difference between novices and experts is that novices tend to look for the single answer to a problem (e.g., the one they think the teacher
wants, the one in the back of the book). Experts tend to look for a solution (an answer if you will), realizing there may be more out there and that sometimes the best solution is based upon justification of their answer (Smith & Good, 1984).

**Self-justification and self-reflection.** Good problem solvers are much more aware of their own strengths and weakness, leading to productive self-regulatory behaviors (Lester & Kehle, 2003). These behaviors can then create a system of checks and balances to ensure that a correct solution is reached, rather than simply accepting the easy first answer as correct.

In fact, in one study (Smith & Good, 1984), expert problem solvers did not obtain more correct answers than the novices, but were able to justify their answers better. While the “expert” may not show a higher degree of accuracy, they are showing a higher degree of comprehension.

Linn (1995) suggests that students need to be encouraged to evaluate evidence, rather than depending on teachers. However, without being taught self-evaluative skills, students may struggle.

**Heuristics and problem-solving steps.** When looking at heuristics or the steps expert problem solvers take in solving a problem, it is important to realize that expert problem solvers keep moving in a forward direction (step by step), while novices will tend to work backward (Larkin, McDermott, Simon, & Simon, 1980). Novices use weak problem-solving heuristics (e.g., means-end
analysis, hill climbing) that limit big picture thinking (Chase & Simon, 1973; Chi et al., 1981).

**Aiding the Novice**

Linn (1996) would likely classify most novice problem solvers as passive learners, that is those looking for outside direction and unable to connect ideas. One way to aid these novice problem solvers is to break down a large problem into several smaller, achievable problems (Diederen et al., 2005).

Linn (1995,1996) wrote extensively about this breakdown of steps, thus making thinking more visible and labeling these steps as scaffolded instruction. Scaffolded instruction is designed to aid the passive learner (novice problem solver). This deconstruction would help the novice keep the problem moving in the forward direction. In the case of this study we use the terms skills, and then subskills to describe the breakdown of steps in the entire problem-solving process that would aid in scaffolding.

Additionally, students need to be taught self-evaluative skills (Linn, 1995). Students have much more experience with processing information outside of school. If self-evaluative skills are not taught, students are likely to revert back to prior conceptions.

**Social Constructivism**

This study adopted a social constructivist approach as describe by Ernest (1996). Social constructivism uses persons in conversation as a metaphor for
the mind. It can be effective to use the metaphor of the mind as a computer, or an evolutionary metaphor, whereby schemes evolve and the best adapted “survive.” These best-adapted schemes will serve as the solutions to the problems for which students are asked to present solutions.

Part of the social constructivist theory outlined by Ernest (1996) is that through shared experiences, humans are constantly modifying their own views to match a reality for which we can never get a true picture. Much of this shaping takes place through the medium of language. Additionally, part of Ernest’s classification of social constructivism is that unlike the other forms of constructivism, there is neither a “right answer” nor absolute position that students must be driven towards. We cannot “regard the world as something that can be known with any certainty” (p. 344).

Students approach science with different existing prior knowledge. Hewson and Hewson (1983) describe this prior knowledge as providing some indication as to the scientific conceptions and “alternate conceptions” that the students possess. Hewson and Hewson suggest three ways new concepts can be integrated with existing concepts: (a) differentiation, (b) exchange, and (c) conceptual bridging. Differentiation occurs when a student replaces a prior concept with a newer one because the new concept is more plausible in a new situation than the previous concept. Exchange occurs when the newer concept is chosen over the prior one because they conflict with each other and the learner determines the newer one to be more valid. Lastly, conceptual bridging
occurs when abstract concepts are linked together by meaningful common experiences.

Following the model of conceptual change presented by Posner, Strike, Hewson, and Gertzog (1982), before any of these conceptual changes can occur, the teacher has to ensure that students find the new concept to be intelligible, plausible, and fruitful. New conceptual understanding can only be achieved by considering prior knowledge. Although different students are likely to have different prior experiences, given the rather homogenous nature of the population being studied, some of the conflicting experiences may be smaller than in other studies. For example, if the problem studied involved quality of drinking water, and almost all students receive city water with only very few relying on well water, there will be little conflict due to prior experience.

**Scaffolding**

The enormity of the task of problem solving can be quite overwhelming for students. In order to breakdown such a grand task, this study will be modeled after the “scaffolding knowledge integration framework” as described by Linn (1995). The scaffolding knowledge integration framework is a method to help students use their ideas as building blocks to develop and understand more complex domains. Students are encouraged to reconcile conflicts with their knowledge rather than memorize isolated facts. The scaffolding knowledge integration framework has four main criteria: (a) identifying new goals for
learning, (b) making thinking visible, (c) encouraging autonomous learning, and (d) providing social supports.

**Problem Solving and Distance Learning**

Distance learning is the idea that at least some instruction, such as student/student and student/facilitator interaction, will take place outside of the confines of the classroom. In the past, this has been done with simple technologies, such as audio teleconferencing (Anderson & Garrison, 1995), video teleconferencing, E-mail, and listserves. As technology has advanced, new software has become available to aid in the sharing of information and ideas such as Blackboard (Ruberg, Taylor, & Moore, 1996); FacilitatePro (Holt, et al., 1998; Williams, Watkins, Daley, Courtney, Davis, & Dymock, 2001); Webtec; and Carmen (which has been used by The Ohio State University). Additionally some software has been written specifically for programs like the online entry software written for GLOBE (Becker, Congalton, Budd, & Fried 1998; Finarelli, 1998; Means, 1998) developed by the National Oceanic and Atmospheric Administration (NOAA) and National Aeronautic and Space Administration (NASA).

Why should distance learning even be considered if it potentially takes away from face-to-face time? Distance learning allows for collaborations that can take place under a much more flexible schedule than the traditional school schedule (Holt et al., 1998; Means, 1998). Additionally, there are no constraints
for distance learning as long as a computer with an Internet connection can be accessed.

Whilst there is currently limited research, distance learning has been found to facilitate collaboration, problem solving, and questioning skills (Williams et al., 2001). Distance learning can be a very powerful tool for developing critical-thinking skills (Williams et al.) as well. Distance learning can also help develop a community of learners (Anderson & Garrison, 1995; Holt et al., 1998) by creating more student-student interactions. Typically, student-teacher contributions in a classroom are asymmetrical with the teacher doing the majority of the talking (Ruberg et al., 1996). Distance learning can change this model from 1-many or 1-1 to many-many. By taking advantage of this social environment, distance learning can support social cognition (Anderson & Garrison).

Distance learning isn’t automatically the “yellow brick road” leading to the “Oz” of instruction and learning. It can share many positive and negative similarities with a “traditional classroom.” Unless done properly, both types of instruction can be misused as simple cognitive delivery systems rather than collaborative conversations (Ruberg et al., 1996). It is important to create a community of learners by facilitating group activities, projects presented, and creating new knowledge. Both formats, unless done properly can be no more than a didactic approach forcing institutional demands upon its students providing no more than independent learner support (Anderson & Garrison, 1995).
Another similarity is that independent students are more likely to succeed and therefore, teachers must provide structure to help all students become more independent. Linn (1996) classifies students as passive, active, and autonomous. Passive learners leave all of the responsibility for selecting course goals to the course designer. Passive learners typically fail to connect ideas and rarely remember what they learn. They tend to memorize details rather than concepts. Active learners respond to hints, follow instructions but do not internalize their learning. They are dependent on others to lead and monitor their learning. They need guidance to become autonomous. This guidance can be delivered with scaffolded instruction. The four key elements of scaffolded instruction that must be incorporated in distance learning are: (a) make course goals accessible, (b) make thinking visible, (c) encourage autonomous learning, and (d) encourage the social nature of learning. Autonomous learners take initiative (Linn). They critique their own understanding, diagnose weaknesses, and seek help. Autonomous learners know their own learning habits. Autonomous learners set goals and meet them or adjust them as necessary. Many students resist becoming autonomous learners. Teachers often reinforce this resistance by discouraging students from adapting coursework to meet their needs and/or by not encouraging the development of cognitive skills. Courses that help students become autonomous learners help explain to students what progress is, help them make connections among examples, and encourage them to critique others’ work. In many “discovery environments,” only autonomous learners will succeed.
Electronic distance-education courses typically require students to be more autonomous because there is less interaction between “classmates.” However, flashy lectures, videos, and multimedia presentations may encourage passive learning. By collecting their own data/information and sharing it online, students can learn necessary science skills through a distance-learning course (Becker et al., 1998), just like a “traditional science” course. Additionally, students can implement expert protocols developed by experts and see their value by interacting with the experts (Finarelli, 1998).

Just as when using cooperative learning in a “traditional” science course, teachers and learners in a distance-learning course must learn group dynamics (Williams et al., 2001). Also, in order to aid in learning, teachers must be fluent in framing questions as well (Williams et al.).

Whilst there are some similarities between distance learning and traditional instruction, there are also many differences. Some of these differences are neither positive nor negative. For instance, students, teachers, and facilitators, will need to adapt to expectations of each other that are very different from a traditional classroom. One of these paradigm shifts is that in creating new knowledge each participant’s contribution is typically weighted as important as that of facilitators (Williams et al. 2001).

The variety of ideas that can be shared in the online format can help students better imagine other points of view (Ruberg et al., 1996). In the GLOBE program, the addition of resident experts to these discussion groups was found to lead to more sophisticated discussions about issues like procedures and
simple statistics such as variance and outliers (Means, 1998). All in all, online forums can develop more in-depth discussions and aid cooperative problem solving (Anderson & Garrison, 1995).

One of the reasons that more information and opinions might be shared online as compared to a traditional classroom is because of the increased feeling of privacy that participants have, leading to more candidness (Ruberg et al. 1996). Also, there is a greater feeling of inclusion (Anderson & Garrison, 1995) among students. Female students as well as less-able, poorer students participate more than they would in a traditional setting (Linn, 1996; Ruberg et al.). Even those that participate infrequently can have a larger influence on a discussion than they would in a traditional classroom (Ruberg, et al.).

Moreover, online forums don’t have constraints of time and place as it is possible to revisit old issues at a later date and bring them back to the fore-front (Ruberg et al., 1996; Williams et al., 2001). The pace can be much more flexible for different learners and can be easily adjusted because of some of the lack of time constraints. Unlike classroom discussions, it is much easier to track what students are thinking because the dialogue can always be reviewed. Additionally, the dialogue can often be of greater detail and precision than that in a regular class as students have time to invest in their words more and have longer, more crafted comments than they would in a traditional class.

Distance learning is not without its flaws. When trying to collect and share data with expert scientists, collecting data for several weeks may work best for a classroom, but this discontinuous data set is not good for the researchers. This
can be overcome if an entire district decides to adopt the program and share the
data collection responsibilities among its schools to have a continuous data set;
however, the logistics of this can be very difficult (Means, 1998).

Another troubling issue can be Information overload, too much unwanted
information and consequently not enough time to read all messages (Ruberg et
al., 1996). These problems need to be dealt with as software improves making
wanted information more easily accessible. The switch from a mailing list format
to a forum format has been found to decrease student frustration and increase
student reflection (Holt et al., 1998). Also, as software changes and gets
upgraded some do not like the time required to get familiar with the software
(Ruberg et al.). Participants also have demonstrated some frustration when
waiting extended periods of time for answers (Ruberg et al.) or if their questions
go unanswered (Means, 1998). This can even lead to misunderstandings and
misinformation remaining visible for an extended period of time (Ruberg et al.).
This issue can only be dealt with by making sure all participants are committed to
the project.

Williams et al., (2001) found online discussion groups can lead to a
greater anxiety among facilitators. This problem suggests that there must be
proper training and support for facilitators. Lastly, many misunderstandings can
occur through online dialogue. Because of a lack of face-to-face contact, the
potential for cross-cultural discussions and decreased social context clues may
lead to a greater risk of misunderstanding (Williams et al.; Ruberg et al., 1996). It
is important that the on-line facilitators provide guidance for “netiquette,” online
etiquette (Holt et al., 1998). These are issues facilitators need to be sensitive to when moderating a discussion.

There have been some agreements in the literature as to what should be included in a model distance-learning program. First of all, if there are any outside participants, non-student/teacher, all parties must benefit (Finarelli, 1998; Means, 1998). Without all parties benefiting there is likely to be little or inconsistent participation. In the case of GLOBE, to make sure there was valid data that could be used by scientists, data collecting protocols needed to be refined. The procedure used in this example was that experts developed a data collecting protocol and taught teachers at teacher-training sessions. Then teachers were able to teach the students (Becker et al. 1998).

The members of the learning community must not only realize, but also fulfill their role to the rest of the community. Members need to cooperate to develop good problem-solving skills. Members must question each other and request reasons for beliefs as well as clarifications for issues they do not understand. They need to build upon other’s ideas, deliberate together, and point out possible counter examples. When making judgments they must develop specific criteria for making decisions and examine these decisions against these criteria. Additionally, the facilitator must understand their role in distance learning. They are the person that sets the mood of the learning environment. The facilitator must be an organizational leader, clarifying the purpose, organizing and making available resources, and helping to summarize key points (Ruberg et al., 1996). They must also serve the role of a moderator,
staying alert to actions that induce strong emotions as well as controlling non-
contributors, distracters, monopolizers, and know-it-alls (Williams et al., 2001).

**Literature-Based Decisions**

**Social constructivism.** In this study, students were split into problem-
solving groups to solve the given tasks. This grouping was necessary if social
interactions were to modify their views of reality. In fact, while there are many
expert opinions of scientific truths, the one purpose of this study was to measure
the improvement in problem-solving skills. The end results in many of these
activities were not predetermined, “acceptable” scientific truths, but rather open-
ended problems that encourage discussion.

**Scaffolding instruction.** The majority of Linn’s (1995) scaffolding
knowledge integration framework was addressed through classroom interaction.
Students also had the option of using an online forum.

It this study, it was anticipated that the first and third aspects, identifying
new goals for learning and encouraging autonomous learning were dealt with
during team-teacher common planning time. This time was when classroom
teachers and experts from the field could create and modify the goals for student
learning so they could properly be linked to the outcomes of their everyday
observations. These goals needed to build on students' intuitions. Students
spend more time processing information in the outside world than in the science
classroom (Linn, 1995). Teachers must also emphasize knowledge integration
among students. If all of these experiences could be harnessed and shared, much more learning could occur. The facilitators must also help provide students with resources to continue developing their knowledge even after the study. Once students have the skills to continue to develop and learn on their own, they are on the road to becoming autonomous learners (Linn, 1995, 1996).

This study hoped to build the second and fourth criteria (making thinking visible and providing social supports) into the design of the problem-solving units and activities. As students had to explain their reasoning to other students in their group, their thinking would become visible on the forum. Teacher feedback to the students about their processes of problem solving was also important. According to Linn (1996) if students could recognize more than one reasonable explanation for phenomena, they may be more likely to expand their repertoire of problem-solving skills. As the teachers moderated problem-solving activities in this study, they along with a learner’s other group members, could provide the social support needed. A positive social structure involved was important, as negative social environment can discourage and frustrate students as much as a positive context can help them.

During the first semester of the Physical Science course, problem-solving activities were broken down into smaller parts that focused on as little as one problem-solving skill and its corresponding subskills. Fewer steps to focus upon should allow for thinking to become more visible. Instruction began with problem evaluation to help student autonomy. It was expected that teachers and fellow team members would provide social support.
Student maturation. Piaget (1964) stated that social transmission plays an important role in the assimilation of information; however, this factor in and of itself is insufficient if the learner is not in a state where he/she can understand the information. Due to the maturation of the students, most, if not all, of the students in the study, at the age of 14 and 15, were either in the concrete-operational or formal-operational stage of cognitive development. However, there is some concern that some of the students were not at these levels as the “average ages at which these stages appear … vary a great deal from one society to another” (Piaget, 1964, p. 178). While the ordering of these stages is constant, there is a risk that students in Integrated (remedial) Mathematics may not have been at these levels due to lack of growth academically, socially, and/or cognitively.

Problem solving outside the classroom. During the course of this study, students were encouraged to use an online forum to discuss issues and problem solve. Online support not only came from teachers but also students. In fact, students often gave more support when it came to higher-order thinking than teachers did.

Furthermore, during the assessment (problem-solving activities) not only were students encouraged to use the forum for their problem-solving activities, but anecdotal evidence from the online forum was used as evidence of problem solving. While using evidence for problem-solving skills and subskills for students using the forum may appear unfair to students not using the forum,
there are a few key things to remember. First, all students had access to the Internet, and more specifically the online forum, during study hall and lunch periods had they chosen to do so in order for use to not be limited to their homes. Second, use of the forum by some students and not others was not a drawback, but rather showed high initiative and thus the autonomy of those students, which as previously mentioned can be associated with high problem-solving skills.
Chapter 3: Materials and Methods

The methods of this study will be described first by detailing the pilot study, then the sample. Variables and instructional methods will then be introduced. Data collection and analysis will be detailed. Lastly, limitations and delimitations will be discussed.

Pilot Study

A preliminary study was conducted in 2008 to attempt to measure student problem-solving skills both with a Likert-type instrument and with authentic assessment. The sample of the pilot study (\(n=29\)) included freshman science students at a Midwestern suburban high school. Four different class sections were studied. Cohort 1p consisted of students taking Physical Science, Honors Geometry, and Honors Global Studies/Honors Language Arts. Cohort 2p consisted of students taking Physical Science, Algebra, and Honors Global Studies/ Honors Language Arts. Cohort 3p consisted of students taking Physical Science, Algebra, and Global Studies/ Language Arts. Cohort 4p consisted of
students taking Physical Science, Integrated (remedial) Mathematics, and Global Studies/ Language Arts. Figure 3.1 provides a summary of the cohorts.

<table>
<thead>
<tr>
<th>Cohort 1</th>
<th>Mathematics</th>
<th>Science</th>
<th>Social Studies</th>
<th>Language Arts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Honors Geometry</td>
<td>Physical Science I &amp; II</td>
<td>Honors Global Studies</td>
<td>Honors Language Arts</td>
</tr>
<tr>
<td>Cohort 2</td>
<td>Algebra</td>
<td>Physical Science I &amp; II</td>
<td>Honors Global Studies</td>
<td>Honors Language Arts</td>
</tr>
<tr>
<td>Cohort 3</td>
<td>Algebra</td>
<td>Physical Science I &amp; II</td>
<td>Global Studies</td>
<td>Language Arts</td>
</tr>
<tr>
<td>Cohort 4</td>
<td>Integrated (remedial) Algebra/Geometry</td>
<td>Physical Science I &amp; II</td>
<td>Global Studies</td>
<td>Language Arts</td>
</tr>
</tbody>
</table>

**Figure 3.1**. The core courses which students were enrolled in as part of Team B’s four cohorts. These cohort characteristics were used for both the pilot study and main study.

**Likert-type instrument.** The Likert-type instrument used was modeled on a study by Heppner and Peterson (1978). Four statements were created in each of three problem-solving skill areas: problem identification, problem planning, and problem evaluation, for a total of 12 Likert-type items. Through random assignment, six of the 12 statements were then worded in a negative manner (e.g., I am NOT good at looking at problems from different people’s points of view). Students had six choices for their response: Always (95%-100%), Usually (75%-95%), Frequently (50%-75%), Occasionally (25%-50%), Rarely (5%-25%), and Never (0%-5%). Responses were then recoded so that answers reflecting problem-solving skills were assigned the highest value (6) and
answers reflecting a lack of problem-solving skills were assigned the lowest value (1).

A principle components analysis was conducted to see if statements could be grouped into the three problem-solving skill areas based upon student responses to the questions. The principle components analysis suggested two or three groups, none of which matched the hypothesized problem-solving skill set. In fact, the best-fit pattern found was when two components were used. All of the positive statements were in one group and all but one of the negative statements were in the other group. It is possible that Heppner and Peterson’s instrument (1978), intended for university students, was too confusing, even after attempts to lower the reading level, for high school freshman.

Summated scores for the entire survey as well as individual problem-solving skill areas showed no significant difference among cohorts. Due to small sample size, a small number of questions, and/or student age/maturity level this test was deemed inappropriate for assessing student problem-solving skills.

**Authentic assessment.** An authentic assessment was also administered to assess problem-solving skills. Students were divided into groups of 3-4 and given a task to solve dealing with Newton’s Second Law of Motion. Students were instructed not to talk, but to write everything down, with each student having a different colored pencil.

The researcher assessed each of three problem-solving skills by looking for evidence of subskills. In Figure 3.2. the novice skills are in regular type, the
intermediate skills are underlined, and the expert skills are in bold type. There was only one rater for all samples and approximately 15% of the samples (4 out of 29) were re-evaluated by the same researcher to ensure intra-rater reliability. Each subskill for which evidence was identified received a “1,” those for which evidence was absent received a “0.” The percentage of subskills present for each skill was then calculated.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Subskills (novice, intermediate, expert)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Definition</td>
<td>Identify possible topics, identify topic choice, <strong>justify topic choice</strong>, identify variables, <strong>isolate variables</strong></td>
</tr>
<tr>
<td>Problem Planning</td>
<td>Brainstorm appropriate procedures, <strong>choose appropriate procedures</strong>, brainstorm appropriate materials, <strong>choose appropriate materials</strong>, brainstorm appropriate equations, <strong>choose appropriate equations</strong>, appropriate data collected, <strong>appropriate method of data presentation selected</strong>, Questions developed regarding end product, end product envisioned</td>
</tr>
<tr>
<td>Problem Evaluation</td>
<td>Find errors in data, find cause for errors in data, find solution for errors in data, evidence used to support results, validity of results questioned, validity of results supported or refuted, plan next step</td>
</tr>
</tbody>
</table>

*Figure 3.2. Problem-solving skills and subskills from preliminary study*

Using analysis of variance, a significant difference (at the $p=0.05$ level) was found in problem-solving skills between Cohorts 1p and 2p, Cohorts 1p and 3p, Cohorts 1p and 4p, Cohorts 2p and 4p, and Cohorts 3p and 4p. No significant difference was found between Cohorts 2p and 3p. Additionally, based upon research detailing the difference between expert and novice problem solvers, the
skills seen as expert skills typically received lower ratings across the board than those seen as intermediate and novice.

The next step, and the purpose of the current study, was to refine the assessment. In the pilot study, three problem-solving skills were settled upon; however, after some reflection and a look back at Polya, the researcher settled on a model with four problem-solving skills (problem definition, problem planning, plan execution, and problem evaluation) and several subskills for each of these skills as previously displayed in Figure 1.1. The refined assessment was then administered to determine if it could consistently detect differences in student groups, as well as detect changes in the problem-solving skills of students over time after engaging in classroom problem-solving units and problem-solving activities.

**Sample Population**

*School data.* The population studied was from a Midwestern suburban high school. Enrollment varies from year to year, but since the 2005-2006 school year (when the district’s third high school first became fully occupied) enrollment has typically ranged from 1,400-1,600 students. The number of students classified as economically disadvantaged is 20% district wide (Davis, Chang, Andrzejewski, & Poirier, 2010).
High School Transition

The students in the study were freshmen at a suburban secondary school. Therefore, almost all students (except for a few repeat freshmen who had not earned enough credits to become sophomores) were transitioning to a new building. Most students came from the district’s four middle schools; however, there were a small number transferring from outside the district from private schools or other public schools.

Social. Social issues are an area of concern in transitioning from middle school to a new building. Students have personal and interpersonal (be it with other students or teachers) struggles that arise in the transition to a new building (Barber & Olsen, 2004). Some students experience higher degrees of depression and lower self-esteem (Barber & Olsen) in a new setting. Additionally, increased social competition in new buildings (Eccles et al., 1993) can transcend personal/interpersonal boundaries.

Parents and teachers share concerns about social adjustment, especially regarding the issue of finding new friends (Akos & Galassi, 2004). These concerns seem justified given the greater amount of disciplinary issues (Lee, Smerdon, Alfeld-Liva, & Brown, 2000) and violence in these larger high school buildings (Darling-Hammond, Ancess, & Ort, 2002).

The relationship students can have with their teacher is another social and academic issue. As students transition to higher grade levels, their relationship with their teachers tends to drop (Eccles et al., 1993), and students often feel
less support (Barber & Olsen, 2004). Students tend to feel there is less monitoring of their progress (Barber & Olsen). This can be a big issue since teacher connection is usually a good predictor of student function (Barber & Olsen).

**Academic.** Social adjustments in high school are often complicated by academic adjustments. Drop-out rates due to increased academic rigor, concerns about academic achievement, and increased homework load are three of the issues.

The two greatest worries students in a new building have are the increased homework load and increased academic rigor (Barber & Olsen, 2004). Academic rigor is a very real concern, as achievement and academic performance (including grades) tend to drop when transitioning to the high school building (Akos & Galassi, 2004, Barber & Olsen, Eccles et al., 1993). Furthermore, student drop out is another concern in the high school (Akos & Galassi), especially in large schools (Darling-Hammond et al., 2002).

As students transition to high school, their academic needs can be in conflict with the school's goal of creating autonomy. Students desire more organization and structure from their teachers (Barber & Olsen, 2004), while at the same time teachers would like to develop autonomous skills in their students (Akos & Galassi, 2004). One of these basic skills is for students to learn how to organize information on their own.
Smaller Learning Communities

The suburban high school in this study had about 1,600 students, thus defining it as a large high school and subjecting students to many of the concerns already discussed, such as transition to new buildings. In large high schools average students can “slip through the cracks” (Lee et al., 2000).

This high school implemented Smaller Learning Communities (SLCs) to help address these concerns. Entering freshman classes (on average about 400 students), would be split into four freshman teams of 100 students. Students on each freshman team would have the same four teachers for each of their four core classes: Language Arts, Math, Science, and Global Studies. Besides the physical scheduling aspect, a SLC also should have a common core of principles, performance based assessment, interdisciplinary instruction, and a common focus on skills and inquiry (Darling-Hammond et al., 2002).

In this smaller learning community, all freshmen were divided into one of four freshman teams. In a freshman team, most students had the same teachers for all of their core subject areas (history, language arts, mathematics, and science). All of the teachers had no less than one period in common to plan integrated activities and discuss student progress and cross-curricular instruction.

Goals. Some of the goals of the SLCs were to ease transition to the high school by using smaller groups, both a desire of parents in general (Akos & Galassi, 2004). Furthermore, research has shown that these smaller groups can
help transition (Barber & Olsen, 2004) and reduce student anonymity often associated with a large high school (Darling-Hammond et al., 2002).

Besides helping students transition socially, SLCs are able to facilitate teaching of skills to students (Darling-Hammond et al., 2002). Because students share the same core teachers, these teachers can provide a unified and overlapping approach to implicitly and explicitly teach many skills (e.g., organizational, problem-solving, and study skills).

Another academic benefit in SLCs is the ability of teachers to plan together and provide rigorous and interdisciplinary instruction (Darling-Hammond et al., 2002). For the freshman team in this study, teachers met daily to discuss student progress and course content so that any opportunity to relate one course to another could be emphasized and students could learn the “big picture.”

The high school in this study desired to create these smaller learning communities in the freshman year. The goal of this change would be to help social adjustment and increase academic achievement much like that in smaller schools (Darling-Hammond et al., 2002).

**Concerns.** One major drawback in this school’s implementation of SLCs was the tracking of students. Students on the freshman team in this study were placed in one of four cohorts, primarily based on math level (advanced, regular, and remedial) and secondarily, by Language Arts (honors and regular). Although convenient for this study, this tracking may have been to the detriment of the
students as a majority of the students that would be expected to struggle the most with problem-solving skills were in the same cohort.

Some research has shown that tracking is ineffective and leads to an achievement gap between white/non-white students and high SES/low SES students (Burris & Welner, 2005). There is evidence to suggest these issues were present in the group being studied. Furthermore, although it was hoped that 100% of all students would be included in the study, there was some risk, based on past precedent, that economically disadvantaged students would be under-represented in the study due to lack of parental consent (Davis et al., 2010).

**Team Data.** Freshman classes of 350-450 students were split into four teams. Three teams (Teams A, B, & C) typically contained 90-120 students, while the fourth team (Team D) typically contained from 70-100 students. Each team had a common lunch and teachers shared a common planning period. Teams did not share the exact same demographics (see Figure 3.3). Because all students in the study had the same four teachers, in the same building, in the same suburban Ohio school district, the potential confounding variable of differences in instruction between classes should have been limited.

Teams in the secondary school in this study were not uniform in the courses offered. Some teams had more honors courses than others (See Figure 3.3).
<table>
<thead>
<tr>
<th></th>
<th>Team A</th>
<th>Team B</th>
<th>Team C</th>
<th>Team D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Math</strong></td>
<td>1 Honors Geometry, 2 Algebra, 1 Integrated-Algebra/Geometry (remedial)</td>
<td>1 Honors Geometry, 2 Algebra, 1 Integrated-Algebra/Geometry (remedial)</td>
<td>1 Honors Geometry, 2 Algebra, 1 Integrated-Algebra/Geometry (remedial)</td>
<td>1 Honors Geometry, 1 Algebra, 1 Integrated-Algebra/Geometry (remedial)</td>
</tr>
<tr>
<td><strong>Science</strong></td>
<td>2 Honors Biology, 2 Physical Science I &amp; II</td>
<td>4 Physical Science I &amp; II</td>
<td>1 Honors Biology, 3 Physical Science I &amp; II</td>
<td>3 Physical Science I &amp; II</td>
</tr>
<tr>
<td><strong>Global Studies</strong></td>
<td>1 Honors Global Studies, 3 Global Studies</td>
<td>2 Honors Global Studies, 2 Global Studies</td>
<td>1 Honors Global Studies, 3 Global Studies</td>
<td>1 Honors Global Studies, 2 Global Studies</td>
</tr>
<tr>
<td><strong>Language Arts</strong></td>
<td>1 Honors Language Arts, 3 Language Arts</td>
<td>2 Honors Language Arts, 2 Language Arts</td>
<td>1 Honors Language Arts, 3 Language Arts</td>
<td>1 Honors Language Arts, 2 Language Arts</td>
</tr>
<tr>
<td><strong>Notes</strong></td>
<td>Contains a special education inclusion unit</td>
<td>Honors Global Studies is blocked and team taught with Honors Language Arts (as are the Global Studies and the Language Arts Classes)</td>
<td>Contains a special education inclusion unit</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3.3. Distribution of class sections among four teams at a Midwestern suburban high school*

Within the freshman team (Team B) studied in this investigation, students were divided into four freshman cohorts of approximately 25-30 students each (See Figure 3.1). With few exceptions, each student within a cohort had the same four core courses, making the cohort rather homogeneous. No other freshman team had clear cohorts (classes were more heterogeneous) and it was because of these cohorts that Team B was selected for this study.

Lastly, although individual student data was collected, to stay within the framework of social constructivism, students typically worked in groups of 3-4
when solving problems. A summary of the organizational hierarchy at this suburban high school can be found in Figure 3.4.

<table>
<thead>
<tr>
<th>Organizational Division and Subdivisions</th>
<th>Typical Size per Division or Subdivision</th>
</tr>
</thead>
<tbody>
<tr>
<td>High School</td>
<td>1,500-1,700</td>
</tr>
<tr>
<td>Freshman class</td>
<td>300-450/ class</td>
</tr>
<tr>
<td>Freshman Teams</td>
<td>80-120/ team</td>
</tr>
<tr>
<td>Cohorts</td>
<td>18-30/ cohort</td>
</tr>
<tr>
<td>Problem-solving groups</td>
<td>3-4/ group</td>
</tr>
</tbody>
</table>

*Figure 3.4. Organizational hierarchy*

**Class selection.** In this school district tracking of students was not an exact science as parents had the ultimate say as to which courses their child would take (B. Ackerman, personal communication, January 5, 2009). There were, however, some tests used along the way to assist teachers, guidance counselors, and administrators in making recommendations for student placement.

The first major tests, taken in fifth grade, were used to place students in either sixth-grade mathematics or pre-algebra. Students were administered the Orleans-Hanna, a mathematics process assessment, and the Terra Nova, the qualitative reasoning portion of the Terra Nova that was used for mathematics placement (J. Metzger, personal communications, January 22, 2009).

Throughout the middle school grades, various administrations of the Terra Nova, Ohio Achievement Test (OAT), as well as Yearly Progress Pro Digital
Learning, which assesses OAT strands weekly, were given to the students. It is these tests, along with student classroom achievement that became the basis for staff recommendations for placement (J. Metzger, personal communications, January 22, 2009; P. Pierpoint. personal communications, January 22, 2009, R. Shrilla, personal communications, January 21, 2009); Once again, however, parents had a large voice in the final placement. Lastly, upon entering their freshman year, all students who were enrolled in either Algebra or Integrated Algebra/Geometry were given an entrance exam to ensure they were placed in the most suitable math class (L. Ferguson, personal communications, August 28, 2009).

Other than the 25 students per school who were pulled out for the Able and Talented (A&T) Program, there was no tracking at three of the four middle schools for science, history, and language arts. One feeder middle school was an exception, and had three different level courses: intervention, regular, and advanced. (M. Lutz, personal communications, January 5, 2009; J. Metzger, personal communications, January 22, 2009). The advanced course dealt with more critical-thinking and problem-solving skills. As students transitioned from middle school to high school, they had the opportunity to take Honors Biology, Honors Global Studies, and Honors Language Arts in their freshman year.

Although there may be a logical argument that parents who push their child into honors courses may have had a more vested interested in their child’s education, children with strong parent influence are more likely to do better in school, and students who do better in school will have learned better use of
problem-solving skills. This argument contains many assumptions, which may or may not be valid.

However, placement into mathematics does have a stronger emphasis on reasoning, critical thinking, and other skills related to problem solving measured by the aforementioned tests. Additionally, students in higher-level mathematics courses can be expected to have expert problem-solving skills.

Because students were not placed in cohorts based on science ability, all students were enrolled in the same science course, it is therefore the researcher’s opinion that mathematics placement (as opposed to the global studies/language arts placement) would be most likely to show significant differences among novice and expert student problem-solving skills. Furthermore, it would be of value to document, which students were in the middle school with science tracking to see if there was a significant difference between the problem-solving skills of this subgroup as compared to the students in the other middle schools.

**Participation**

The freshman team participating in the study began with 99 students and ended with 96 students as three students left the team during the course of the study. Data for these three students were not included in the study because their data was incomplete. Not only were these students not in attendance for all four problem-solving assessments, but attendance was so poor, that all three
students were absent for at least part of the problem-solving assessments even during their time of enrollment at the school.

Permission was obtained via consent forms for 77 out of 96 students. Four forms were returned requesting not to participate, and 15 forms were not returned even after multiple reminders. The final tally of students participating in the study were 18 of 19 students in Cohort 1, 26 of 29 students in Cohort 2, 16 of 21 students in Cohort 3, and 17 of 27 students in Cohort 4.

**Cross-teamed students.** Eight of the 77 students included in the study were “cross-teamed” students. These students did not have all four core teachers for their courses. Three of the eight students matched the cohort classification scheme quite well. Of the three students, two matched the same four courses as Cohort 4 (Integrated Algebra/Geometry, Physical Science, Global Studies and Language Arts) and were thus placed in that cohort for data analysis purposes. One student was mathematically advanced and taking Algebra 2 as a freshman. This student was placed in Cohort 1 as this student matched Cohort 1 best (advanced math, Physical Science).

The remaining five students did not match any of the cohorts perfectly. Since the pilot study showed that math courses taken had the highest correlation of all courses with student problem-solving skills, students were first separated into cohorts by math course. The data for the two students enrolled in an honors math course were included in Cohort 1 and the data for the two students enrolled in an Integrated (remedial) Algebra/Geometry course had their data included in
Cohort 4; both these placements were done irrespective of whether or not they were talking Honors Global Studies and/or Honors Language Arts. The one remaining Algebra student was home schooled in Global Studies and Language Arts. Because this student was involved in the problem-solving instruction in both their math and science courses, where the majority of problem-solving instruction related to the PSSA tasks was given, this student was included in the study. Additionally, since this student was not required to think at the higher level required by the Honors Language Arts and Honors Global Studies courses, data from this student was included with Cohort 3.

Data Collection in the Current Study

The structure of this study was a Time Series Design as depicted in Figure 3.5. An initial baseline reading (O₁) of the dependent variables was taken. The dependent variables in this study were the level of problem-solving skills and subskills. These variables were also measured again at three different intervals (O₂, O₃, and O₄).

\[ O₁ \ X₁ \ O₂ \ X₂ \ O₃ \ X₃ \ O₄ \]

*Figure 3.5. Time series design*

The treatments (Xᵢ) included problem-solving units and problem-solving activities used in the core subject areas. These units and activities were part of
the normal curriculum. As much as possible, the skills and subskills addressed were documented (Figure 3.6-3.8).

**Measurement of Variables**

**Independent variables.**

*Cohort.* There were several independent variables that may have affected selected student problem-solving subskills. The first variable was student cohort. Based on course selection, students may have been enrolled in one of four cohorts (see Figure 3.1). While a causal relationship would be impossible to establish, a correlation between course selection and presence of selected problem-solving subskills was expected.

Students were tracked within the SLC team based upon cohort (course selection) in mathematics, language arts, and social studies but not science. All students were currently enrolled in Physical Science II and had completed Physical Science I first semester. No Honors Biology students were included in this study, science this course was not offered on this freshman team. One science class contained students taking 3 honors courses (mathematics, language arts, and social studies); one science class contained students taking two honors courses (language arts and social studies); one science class contained students taking no honors courses; and one science class contained students enrolled in no honors courses and a remedial mathematics course.
**Problem-solving activities and units.** In order to improve student problem-solving skills, problem-solving activities were introduced in each course. Problem-solving activities were scaffolded activities that typically lasted one to several days and were designed to increase the level of one or more of the subject’s problem-solving skills. It was anticipated that these problem-solving activities should contribute to a statistically significant increase in one or more of the problem-solving skills/subskills for a population.

SLC teachers also cross-referenced any class and integrated activities with the researcher’s problem-solving skills/subskills summary to determine which activities in their classes would likely increase student problem-solving skills. These individual lessons could thus be classified as problem-solving activities.

Improvement in student problem-solving skills/subskills may also be a function of problem-solving units. Problem-solving units, primarily in Physical Science II, Geometry, and Algebra, were parts of the course of study, lasting up to several weeks, that were expected to increase the level of one or more of the subject’s problem-solving skills/subskills. These units were not specifically designed to improve problem-solving skills/subskills, but often inadvertently required the use of one or more subskills. There were often scaffolded instructions that guided students through these activities.

Teachers involved in the SLC cross-referenced their course of study with the researcher’s problem-solving skills/subskills summary to determine which
units would likely increase student problem-solving skills. These large segments in the course of study could thus be classified as problem-solving units.

**Instructional procedures and instructional content.** The freshmen teachers in the team chosen for the study valued problem-solving and critical-thinking skills. Two of the areas they stressed in their instruction were the evaluation and justification of statements made based upon readings, especially in the team taught Global Studies/Language Arts class, and data, especially in science and mathematics. These skills also happened to be two areas that tend to separate novice problem solvers from expert problem solvers.

However, in an attempt to attain a higher precision of understanding of student problem-solving skills, the freshmen team teachers determined the specific subskills that tended to be addressed in their instruction. The team-taught Global Studies/Language Arts teachers were asked to identify which subskills they addressed during the semester (see Figure 3.6).
Figure 3.6. Problem-solving activities and the corresponding skills and subskills addressed in the team-taught Global Studies/Language Arts class. Activities occurred between the observations as designated in the figure. Subskills in bold were included in this study. Abbreviations for subskills can be found in Figure 1.1

In the case of Mathematics and Physical Science, the problem-solving subskills emphasized varied depending on the unit. There were three different Mathematics courses that students were taking: Honors Geometry, Algebra or Integrated Algebra/Geometry. Each of these courses needed to be addressed separately. The Mathematics and Science teachers described which subskills were addressed (see Figures 3.7-3.8).
<table>
<thead>
<tr>
<th>Observation 1: Paper Airplanes, 01-Feb-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honors Geometry</td>
</tr>
<tr>
<td>11-FEB-2010: Scientific notation (B6 &amp; C2)</td>
</tr>
<tr>
<td>19-FEB-2010: Operations with exponents, Why do properties work? (D4 &amp; D6)</td>
</tr>
<tr>
<td>24-FEB-2010: Graphing exponential functions (C2)</td>
</tr>
<tr>
<td>24-FEB-2010: Modeling exponential decay, and compound interest (D4, D5 &amp; D6)</td>
</tr>
<tr>
<td>Algebra</td>
</tr>
<tr>
<td>19-FEB-2010: Multiplying polynomials (B5 &amp; B6)</td>
</tr>
<tr>
<td>Integrate Algebra/Geometry</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation 2: Mousetrap Cars, 08-Mar-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS Unit - MAR-2010: Proofs</td>
</tr>
<tr>
<td>12-MAR-2010: Theorems and proofs involving circles (B6, D4, D5, &amp; D6)</td>
</tr>
<tr>
<td>22-MAR-2010: Applying Pythagorean Theorem – surveying &amp; measuring indirectly (B6, D4, D5, &amp; D6)</td>
</tr>
<tr>
<td>01-APR-2010: Learning tools and applying them to new situations (B6, B7, B8, D4, D5, &amp; D6)</td>
</tr>
<tr>
<td>26-APR-2010: Finding areas of figures (B6, D4, D5, &amp; D6)</td>
</tr>
<tr>
<td>PS Unit – MAR/APR-2010: Solving equations</td>
</tr>
<tr>
<td>09-MAR-2010: Different methods for multiplying polynomials (B6)</td>
</tr>
<tr>
<td>24-MAR-2010: Choosing factoring methods (B5 &amp; B6)</td>
</tr>
<tr>
<td>29-MAR-2010: Graphing quadratics (C2)</td>
</tr>
<tr>
<td>19-APR-2010: Choosing the best method to solve quadratic equations (B6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation 3: Catapult Projects, 03-May-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-MAY-2010: Surface areas and volumes of 3-D objects (B6, D4, D5, &amp; D6)</td>
</tr>
<tr>
<td>13-MAY-2010: Coordinate Geometry, knowing which formula to use. (B7, B8, &amp; D4)</td>
</tr>
<tr>
<td>07-MAY-2010: Checking for extraneous solutions (D5 &amp; D6)</td>
</tr>
<tr>
<td>19-MAY-2010: Using Pythagorean Theorem to find distance and midpoint (D4, D5, &amp; D6)</td>
</tr>
<tr>
<td>20-MAY-2010: Similar Figures (B5, B6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observation 4: Rube-Goldbergs, 08-Jun-2010</th>
</tr>
</thead>
</table>

Figure 3.7. Problem-solving units and activities and the corresponding skills and subskills addressed in Mathematics courses. Activities occurred between the observations as designated in the figure. Subskills in bold were included in this study. Abbreviations for subskills can be found in Figure 1.1
### Physical Science II

#### Observation 1: Paper Airplanes, 01-Feb-2010
- 05-FEB-2010: Race Lab (C2, D1, D2, D3, D4, D5, D6, & D7)
- 12-FEB-2010: Speed and velocity (B1, B2, B5, B6, C2, D4, D5, & D6)
- 19-FEB-2010: Opinion of the week (A1, A2, A3, D4, D6)
- 23-FEB-2010: Dropper lab (C2, D1, D2, D3, D4, D5, D6, & D7)
- 23-FEB-2010: Acceleration (B1, B2, B5, B6, C2, D4, D5, & D6)
- 24-FEB-2010: Motion concept maps (A1, A2, D4)

#### Observation 2: Mousetrap Cars, 08-Mar-2010
- PS Unit - MAR/APR-2010: Newtons Laws
- 19-MAR-2010: Opinion of the week (A1, A2, A3, D4, D6)
- 23-MAR-2010: N1LM, force diagrams (A1, A2, B7, B8, & C2)
- 31-MAR-2010: N2LM Lab (A1, A2, A3, B1, B2, B3, B4, B5, B6, B7, B8, C1, C2, D1, D2, D3, D4, D5, D6, & D7)
- 01-APR-2010: N2LM (B1, B2, B5, B6, C2, D4, D5, & D6)
- 14-APR-2010: N3LM, momentum (A1, A2, B1, B2, B5, B6, B7, B8, C2, D4, D5, & D6)
- 21-APR-2010: Projectile Lab (A1, A2, B1, B2, B5, B6, B7, B8, C1, C2, D1, D2, D3, D4, D5, D6, & D7)

#### Observation 3: Catapult Projects, 03-May-2010
- 06-MAY-2010: Mechanical energy (B1, B2, B5, B6, C2, D4, D5, & D6)
- 11-MAY-2010: Pendulum Lab (A1, A2, A3, B1, B2, B3, B4, B5, B6, B7, B8, C1, C2, D1, D2, D3, D4, D5, D6, & D7)
- 12-MAY-2010: Energy conservation (A1, A2, A3, D4, D5, & D6)
- 27-MAY-2010: Dice Lab (C1, C2, D1, D2, D3, D4, D5, D6, & D7)
- 27-MAY-2010: Nuclear energy (B5, B6, & D4)

#### Observation 4: Rube-Goldbergs, 08-Jun-2010

---

*Figure 3.8. Problem-solving units and activities and the corresponding skills and subskills addressed in Physical Science II. Activities occurred between the observations as designated in the figure. Subskills in bold were included in this study. Abbreviations for subskills can be found in Figure 1.1*

---

A summary of the number of lessons for each problem-solving skill and subskills in each course, Global Studies/Language Arts (GS/LA), Physical Science II (Sci), Geometry (Geo), Algebra (Alg) and Integrated Algebra/Geometry (Int), can be found in Table 3.1. The occurrence of the treatments are designated by the following labels: X1, occurring between O1 and O2; X2, occurring between O2 and O3; and X3, occurring between O3 and O4. The totals are calculated by adding the occurrences in Global Studies/Language Arts,
Science, and the average for the three math courses. For example the total for treatment period X1 for “evidence used to support results” (D4) was found by adding 3+6+(0+2+0)/3=9.7.

Table 3.1. Problem-solving activities for each subskill and course. Problem-solving subskill abbreviations, A1 through D7 can be found in Figure 1.1. Also displayed are the totals for the three treatment time periods: X1, X2, and X3. Abbreviations for subskills can be found in Figure 1.1

Dependent variables.

Variables. The primary dependent variables for which data was collected was evidence of student problem-solving subskills. The proficiency of these skills were assessed during problem-solving assessments. The researcher looked for evidence of each of these subskills and assigned a “2” if there was
direct/conclusive evidence of a subskill present, a “1” if there was indirect or inconclusive evidence of a subskill present, and a “0” if there was no evidence of a subskill present. From the data collected, the total percentage of subskill points possible for each of the main problem-solving skills could then be calculated for each student.

The focus of this study was on 12 of the subskill areas that have been characterized as expert problem-solving skills, and one subskill, “appropriate method of data presentation selected,” that has been characterized as an intermediate problem-solving skill, to ensure that all four skill areas are represented. Each subskill was analyzed individually as well as collectively with each subskill receiving an equal weight (e.g., since there were 13 subskills in total, each subskill contributed to $1/13^{th}$ of the total).

**Problem-solving assessment.** A problem-solving assessment was conducted in the beginning of the study, as a baseline check of student problem-solving skills, as well as twice during the study to observe changes. At the end of the study, a problem-solving assessment was administered to provide a final comparison, for a total of four administrations.

**Levels.** There are several different levels the researcher chose to examine regarding each the problem-solving assessments. The most comprehensive level is referred to as the “problem-solving ability.” Problem-
solving ability is the mean for all thirteen problem-solving subskills so that one mean score is reached for each student and for each of four assessments.

The next, more detailed view, is referred to as the “problem-solving skills.” At this level the four problem-solving skills including: problem definition, problem planning, plan execution, and problem evaluation were examined. Two of the four measures, problem definition and plan execution, only had one subskill included, which were justify topic choice and appropriate method of data presentation selected respectively. Problem planning had four subskill scores including: isolate variables, choose appropriate procedures, questions developed regarding end product, and end product envisioned. These were averaged to obtain the problem planning skills scores. Problem evaluation had seven subskill scores including: find errors in data, find cause for errors in data, find solution for errors in data, evidence used to support results, validity of results questioned, validity of results supported or refuted, and plan next step. These were averaged to obtain the problem evaluation scores.

The most detailed view is referred to as the “problem-solving subskills.” This level examines the 12 higher-level and 1 intermediate-level problem-solving subskills: justify topic choice, isolate variables, choose appropriate procedures, questions developed regarding end product, end product envisioned, appropriate method of data presentation selected, find errors in data, find cause for errors in data, find solution for errors in data, evidence used to support results, validity of results questioned, validity of results supported or refuted, and plan next step.
The first two levels were the major foci of this study. However, to get good resolution, the problem-solving subskills level was also used. This level was not expected to yield significant results and was included to assist in the understanding and interpretation of the problem-solving ability and problem-solving skills levels results.

Assessment type. There were two types of assessments to measure problem-solving skills: classical and supplemental. Both types included evaluation of student work in the science classroom to assess the presence of problem-solving subskills.

Classical assessment (see Appendix A) included typical teacher observations of problem-solving subskills during the problem-solving process. The science teacher had a rubric and an oral-question script to complete as they observed students conducting a problem-solving task. The science teacher documented observations of problem-solving subskills both in the classroom and via the available online forum. Additionally, a question script was designed which allowed the teacher to interact with the students to probe for problem-solving skills that may not have been directly observed. Lastly, any other student writings that would be used for normal classroom grading (i.e. lab write-ups and online discussions) were included. This represented assessments that could be conducted by a classroom teacher under ordinary circumstances while trying to assign a grade to student performance. The amount of information collected was
limited due to time, but the ability to conduct such an assessment was considered to be the most feasible.

Supplemental Assessments (see Appendix B) were given at the conclusion of the problem-solving activity. These assessments contained a group of questions designed to elicit examples of student problem-solving skills displayed or performed during the preceding activity. These assessments took additional classroom time, but were designed to catch any evidence of problem-solving that a teacher of up to 30-students in a classroom might miss while circulating around the classroom.

The problem-solving assessments ($O_1$, $O_2$, $O_3$, & $O_4$) in this study were chosen based upon a variety of factors. First, they were all novel problems or at least novel circumstances for each student. A few students had mentioned that they had made paper airplanes in science class or Rube-Goldbergs in gifted studies; however, none of them had made it with the parameters given to them.

A flow chart displaying the relationship between the independent and dependent variables can be found in Figure 3.9. A specific example of these variables as related to the physical science course can be found in Figure 3.10.
Figure 3.9. Independent and dependent variables in the main study

Figure 3.10. An example of independent and dependent variables in the physical science course
Another reason for choosing these assessments was to emphasize the intrinsic goal, to complete the problem successfully, as well as the satisfaction of learning problem-solving skills. Vansteenkiste, Lens, and Deci (2006) showed that an emphasis on intrinsic goals alone has greater benefit than in combination with extrinsic goals or than extrinsic goals by themselves. The thought was that the drive to succeed at the task (intrinsic motivation) would be the emphasis rather than the extrinsic motivation of getting a good grade.

Observation 1 (Pre-Test): Paper Airplane Activity. Students had to design a paper airplane that could hit a stationary teacher in the front of the room from a set distance away. Students were only to use computer paper and paper clips. Typical data collected by students would include independent variables like design specifications (e.g., number of folds, folding orientation, and paperclip placement) as well as dependent variables like distance and accuracy.

Observation 2: Mousetrap Cars. Students were given various materials (e.g. wood shims, CDs, and dowel rods) to construct a car powered by the spring in a mousetrap. At the end of the construction and testing, students competed for the “best” mousetrap car. “Best” was defined by criteria collected from the students. Typical data collected by students would include independent variables like number of wheels and form of traction as well as dependent variables like distance, speed, and acceleration.

Observation 3: Catapults. Students were given various materials (e.g., wood block, craft sticks, and mousetraps) to construct and test a catapult to be used for a small battle. The ammunition, play-dough, was provided to hit mini-
army men. Typical data collected by students included independent variables like launch angle and type of spoon as well as dependent variables like distance and straightness.

Observation 4 (Post-Test): Rube-Goldberg. Students used left over materials from past problem-solving assessments, and were free to use any of their own materials to construct a Rube-Goldberg device that contained at least 10 energy transfers and six different types of energy. Typical data collected by students included independent variables like design specifications and adjustments as well as dependent variables like reliability.

**Data Analysis**

Analysis of Variance (ANOVA) with repeated measures was conducted to see if there were any significant changes in problem-solving skill assessment scores related to time of assessment or cohort membership. Total problem-solving ability, the four problem-solving skill areas, and 13 problem-solving subskills were examined.

Additionally, the researcher examined changes in problem-solving skills/subskills in relation to the skills/subskills addressed in the problem-solving units and problem-solving activities that transpired during the semester to see if there were any perceivable patterns.
**Validity.** These Problem-Solving Assessments were designed to be direct tests of student problem-solving skills. The skills evaluated were grouped into four distinct subsets as articulated in Polya (1957). These skills were identified as either being present/showing conclusive evidence (2 points), inconclusive (1 point), or no evidence (0 points).

As stated by Adams (1964), the purpose of the experiment determines what types of validity must be ensured. Because this instrument was being used as an absolute measure and it was not being compared to another source of data, concurrent validity was not particularly applicable.

Predictive validity was assessed by comparing the skills addressed during classroom instruction, with changes in problem-solving skills/subskills. The researcher kept a journal documenting skills/subskills addressed in each course as they occurred. This information was summarized and is displayed in Figures 3.6-3.8.

Content validity was integrated into the instrument by conducting a literature review of problem-solving skills. A list of the “universe” of problem-solving skills and sub-skills was documented (see Figures 2.1) to create a rather comprehensive skill set.

Construct validity was evaluated after data from the study was collected. If the instrument was valid, it should have been able to help distinguish among various cohorts and episodes of student instruction.

For example, one-quarter of the freshmen in the study were enrolled in Honors Geometry, while another half were in Algebra. If the test is valid, than the
group of students in Honors Geometry would be expected to demonstrate more
evidence of problem-solving skills, especially those related to the Honors
Geometry syllabus. Even though Honors Language Arts and Honors Global
Studies completed the same activities as the regular course, the depth of
engagement would likely show slightly higher levels of the problem-solving skills
that were required in those courses (see Figures 3.6-3.8).

Also, over the course of the study, all students should show more
evidence of problem-solving skills in later episodes than in earlier episodes. This
improvement should be especially true for those skills purported to be developed
during various problem-solving activities during the course of the school year
(see Figure 3.6-3.8). These problem-solving activities were documented so that
construct validity could be assessed.

Various science and mathematics teachers at this school evaluated the
assessments for face validity. Teachers were given the proposed set of
skills/subskills to determine if the problem-solving skills being evaluated were not
only comprehensive, but also age appropriate. All five teachers, two science,
one mathematics, one language arts, and one history, found the list to be quite
comprehensive and suggested only minor wording changes that were included in
the final list of skills/subskills (Figure 1.1.).

Lastly, reliability would support the validity of the instrument. In the case
of this study, there was only one person marking the assessments, the
researcher. Therefore, intra-rater reliability was assessed by having the
researcher mark a complete set of assessments twice and compute a Pearson's
correlation coefficient. A coefficient close enough to 1.0, would support the hypothesis that marking was consistent throughout the study.

**Limitations.** The students could not be, in the purest form of the word, randomly assigned to class sections. Students were pseudo-randomly assigned to class sections by the school’s scheduling computer, with some confounding variables.

One major set of variables was the level of mathematics, global studies, and/or language arts classes. Even though this was part of the basis of this study, due to the fact that the team selected was divided into cohorts, certain combinations were impossible. For instance, if a student enrolled in Honors Geometry, they were also in Honors Global Studies and Honors Language Arts, where for other teams they could have been enrolled in a regular course. Likewise, no students taking Honors Biology were included in the study because the team included in this study did not have Honors Biology as a course option. These limitations were simply byproducts of the logistics of scheduling 400 students with a wide selection of courses.

Another limitation affecting the student schedule was the elective courses students selected. Some electives only met at one or two times during the day, thus limiting when a student might enroll in their science course and thereby limiting their team assignment. For example, if a student was signed up for orchestra, which was only offered during the sixth period of the day, they were unable to take science during the sixth period of the day or be on a team that had
core classes sixth period. Therefore, some students were not able to be selected for this freshman team which had core courses periods 3, 4, 6, and 7.

Another limitation due to scheduling was variation in class-size. Two classes were relatively large, Cohort 2 and Cohort 4, and two were quite small, Cohort 1 and Cohort 3. As a result, group sizes varied; some classes had more groups of 3 and some classes had more groups of 4, based upon class size.

However, if this instrument is to be used by everyday classroom teachers, it is important that it withstand non-ideal conditions to come up with significant results, as most classroom environments are not ideal.

**Delimitations.** Due to the start time of the study (January), the teachers had already begun to teach some problem-solving skills as part of their classes. It would be unethical to hold off instruction until the research began. However, the researcher felt confident that there were so many problem-solving activities still to occur that growth could still be measured.

Another variable that could potentially affect student problem-solving skills and subskills was time as related to student cognitive development and motivation. The researcher assumed time was not significant from the standpoint of maturity. Very few students were expected to have a significant increase in maturity or conceptual development stage during the 4-month course of the study. Due to the short time period of this study, the researcher treated time as a delimitation.
Lastly, the researcher chose to operate under the assumption that the tasks associated with the problem-solving assessments were similar in difficulty level. Tasks had to vary to ensure each problem was novel and not simply an exercise or algorithm for the students to perform. Also, tasks needed to be related to class content so students had the knowledge bases to refer to when solving problems. Given these two constraints, and the attempt to compare student skills and subskills from one assessment to the next, the researcher sought four problem-solving tasks similar in difficulty level.
Chapter 4: Results

Included in the results is data regarding intra-rater reliability, as well as comparisons of problem-solving ability and problem-solving skills across assessments and cohorts.

Intra-Rater Reliability

To determine intra-rater reliability, the same assessor marked the final assessment twice, with the two markings being 1 month apart. The final assessment was chosen for several reasons. First, the 1 month apart ensured that there was no memory of the first marking; there needed to be significant time between markings. Second, because students needed timely feedback, and some papers were written too lightly to make photocopying an option, the final assessment was chosen as the assessor could hold onto those papers after the end of the school year. A Pearson’s Correlation between the intial marking and the marking 1 month later (see Appendix C) was computed to determine intra-rater reliability. The correlation for combined subskills was 0.85 and for total Problem-Solving Ability was 0.96.
However, there were several subskill areas that, although they displayed a strong intra-rater correlation, had a correlation that was lower than desirable. Four subskills (isolate variables, find solutions for errors in data, validity of results questioned, and validity of results supported or refuted) had correlation coefficients between 0.68 and 0.80. In most individual cases it was quite easy to determine whether a subskill was present (2) or absent (0). However, the difficulty arose, especially for these four subskills, in determining if there was enough evidence to warrant a rating of “inconclusive” (1). Because evidence for these four subskills may have been a bit more subtle, the rubric needs to have clear operational definitions of these four subskills to ensure more reliable marking. Recommendations for these subskills can be found in the conclusion.

**Problem-Solving Ability Comparison**

Comparisons using multivariate analysis of variance with repeated measures were made to determine if there were any significant problem-solving assessment differences by cohort. Significance was found in the tests of between-subjects cohort effects ($F(1,73)=5.92, p=.000$, see table 4.1) and the within-subject assessment contrasts ($F(1,3)=154.84, p=.000$, see table 4.2). Pairwise comparisons were also made between cohorts to determine if there were any significant overall problem-solving assessment differences detected.
Table 4.1. Multivariate analysis of variance by cohort by assessment with repeated measures-between-subject effects

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>231.87</td>
<td>1</td>
<td>231.87</td>
<td>1194.13</td>
<td>.000</td>
<td>0.94</td>
</tr>
<tr>
<td>Cohort</td>
<td>17.76</td>
<td>3</td>
<td>5.92</td>
<td>30.49</td>
<td>.000</td>
<td>0.56</td>
</tr>
<tr>
<td>Error</td>
<td>14.18</td>
<td>73</td>
<td>0.194</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2. Multivariate analysis of variance by cohort by assessment with repeated measures-within-subject contrasts

<table>
<thead>
<tr>
<th>Source</th>
<th>Assessment</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment</td>
<td>Linear</td>
<td>11.24</td>
<td>1</td>
<td>11.24</td>
<td>154.84</td>
<td>.000</td>
<td>0.68</td>
</tr>
<tr>
<td>Assessment by cohort</td>
<td>Quadratic</td>
<td>.63</td>
<td>3</td>
<td>.21</td>
<td>2.68</td>
<td>.053</td>
<td>.10</td>
</tr>
</tbody>
</table>

Repeated measures. Means across all the cohort groups increased over all four assessments (see Table 4.3 and Figure 4.1). Pairwise comparisons of the overall problem-solving ability mean scores across the assessments resulted in significant differences (at the 0.05 level) for problem-solving ability scores between all assessments except Assessment 1 ($M_1=0.68$) and Assessment 2 ($M_1=0.75$).
<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Cohort 4</th>
<th>Overall Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.79 (+/-0.22)</td>
<td>0.81 (+/-0.31)</td>
<td>0.70 (+/-0.24)</td>
<td>0.34 (+/-0.21)</td>
<td>0.68 (+/-0.31)</td>
</tr>
<tr>
<td>Paper Airplanes</td>
<td>1.06 (+/-0.34)</td>
<td>0.83 (+/-0.30)</td>
<td>0.67 (+/-0.31)</td>
<td>0.38 (+/-0.35)</td>
<td>0.75 (+/-0.40)</td>
</tr>
<tr>
<td>Mousetrap Cars</td>
<td>1.34 (+/-0.46)</td>
<td>1.18 (+/-0.32)</td>
<td>0.89 (+/-0.45)</td>
<td>0.47 (+/-0.22)</td>
<td>1.00 (+/-0.48)</td>
</tr>
<tr>
<td>Catapults</td>
<td>1.41 (+/-0.25)</td>
<td>1.35 (+/-0.28)</td>
<td>1.09 (+/-0.43)</td>
<td>0.81 (+/-0.44)</td>
<td>1.19 (+/-0.42)</td>
</tr>
<tr>
<td>Rube-Goldberg</td>
<td>1.15 (+/-0.05)</td>
<td>1.05 (+/-0.04)</td>
<td>0.84 (+/-0.06)</td>
<td>0.50 (+/-0.05)</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.15 (+/-0.05)</td>
</tr>
</tbody>
</table>

Table 4.3. Means (M) and standard deviations (SD) for problem-solving ability by assessment by cohort and overall across assessments and cohorts.

Figure 4.1. Change in mean problem-solving ability scores by assessment by cohort and cohorts combined. See Table 4.3 (df=9)
Although the increase in problem-solving ability between Assessment 1 ($M=0.68, SD=0.31$) and Assessment 2 ($M=0.75, SD=0.40$) was not significant, there was a significant improvement ($p=0.000$) in problem-solving ability detected with this instrument between Assessment 2 ($M=0.75, SD=0.40$) and Assessment 3 ($M=1.00, SD=0.48$), as well as detected ($p=0.000$) between Assessment 3 ($M=1.00, SD=0.48$) and Assessment 4 ($M=1.19, SD=0.42$) as shown in Figure 4.2. These results are reported in Appendix D.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Mean Assessment Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1$</td>
<td>0.68</td>
</tr>
<tr>
<td>$O_2$</td>
<td>0.75</td>
</tr>
<tr>
<td>$O_3$</td>
<td>1.00</td>
</tr>
<tr>
<td>$O_4$</td>
<td>1.19</td>
</tr>
</tbody>
</table>

*Figure 4.2. Display of overall mean assessment differences. The mean values underscored by a common line are not significantly different while all other mean differences are significant.*

The growth shown is rather consistent with classroom instruction. As seen in Table 3.1, the greatest number of problem-solving activities occurred between Assessment 2 and Assessment 3 and that’s where the greatest amount of growth was seen. Likewise, it is reasonable to expect the growth between Assessments 3 and 4. The cumulative effect of all problem-solving activities throughout the semester certainly could explain student ability to “get it.”

**Cohorts.** The overall problem-solving ability scores across the four assessments matched predictions for all four Cohorts. Cohort 1 had the highest mean ($M=1.15, SD=0.05$), followed by Cohort 2 ($M=1.05, SD=0.04$), Cohort 3 ($M=0.84, SD=0.06$), and Cohort 4 ($M=0.50, SD=0.03$) with the lowest mean (see
Figure 4.3 and Figure 4.4). Pairwise comparisons of the overall problem-solving ability scores across the four assessments (see Table 4.4 and Figure 4.3) show significant differences at the 0.05 level between every Cohort except “Cohort 1” and “Cohort 2.”

<table>
<thead>
<tr>
<th>Cohort</th>
<th>C4</th>
<th>C3</th>
<th>C2</th>
<th>C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Assessment Score</td>
<td>0.50</td>
<td>0.84</td>
<td>1.05</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Figure 4.3. Display of overall mean assessment differences by cohort. The mean values underscored by a common line are not significantly different while all other mean differences are significant.

Figure 4.4. Comparing means and standard deviations for the overall problem-solving ability scores across the four assessments by cohort.
<table>
<thead>
<tr>
<th>(I) CHT</th>
<th>(J) CHT</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>.10</td>
<td>.07</td>
<td>.506</td>
<td>-.09</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>.31</td>
<td>.08</td>
<td>.001</td>
<td>.10</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>.65</td>
<td>.07</td>
<td>.000</td>
<td>.44</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.21</td>
<td>.07</td>
<td>.037</td>
<td>.01</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>.54</td>
<td>.07</td>
<td>.000</td>
<td>.35</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>.34</td>
<td>.08</td>
<td>.001</td>
<td>.12</td>
</tr>
</tbody>
</table>

Based on observed means. The error term is Mean Square (Error) = .049.

Table 4.4. Multiple comparisons for problem-solving ability mean scores by cohort. Significant results ($p=0.05$) are marked in bold.

Cohort 1 and Cohort 2 both included students taking no less than two honors courses. Although Cohort 1 had students taking Honors Geometry and Cohort 2 had students taking Algebra, no significant difference was found. When looking back at the emphasis on problem-solving skills, there was greater emphasis in the Algebra class than the Honors Geometry class during the second semester. Although Cohort 1 had a greater mean score than Cohort 2, the greater emphasis on problem-solving skills in Algebra could help explain the closing of the gap and lack of a statistically significant difference ($p=.506$) between overall problem-solving ability mean scores for Cohort 1 ($M=1.15$, $SD=0.05$) and Cohort 2 ($M=1.05$, $SD=0.04$).

Both Cohort 1 ($M=1.15$, $SD=0.05$) and Cohort 2 ($M=1.05$, $SD=0.04$) had significantly higher overall problem-solving ability scores than students without honors classes. Students in Cohort 4 ($M=0.50$, $SD=0.05$) who were enrolled in remedial math, showed significantly lower scores than all three of the other
cohorts.

Finally, by examining Figure 4.1 one can look at the timing of cohort improvement. Although the assessment by cohort interaction was not significant ($p=0.053$), the timing of growth, as seen by the slopes of the graph, appears consistent with cohort level. Between Assessment 1 and Assessment 2, Cohort 1 appears to have greatest growth; Between Assessment 2 and Assessment 3, Cohort 2 appears to have greatest growth, although Cohort 1 and Cohort 3 are not far behind; Lastly, between Assessment 3 and Assessment 4, Cohort 4 appears to have greatest growth this time with Cohort 2 and Cohort 3 slightly behind.

**Problem-Solving Skills Comparison**

Comparisons using multivariate analysis of variance with repeated measures were made to determine if there were any significant differences in four problem-solving skills (problem definition, problem planning, plan execution, and problem evaluation) detected by the problem-solving assessments. As shown in Table 4.5, there was a significant multivariate assessment by cohort interaction effect ($p=0.000$). There was also significant assessment by cohort between-subjects contrast effect (see Table 4.6) for all four problem-solving skills: problem definition ($p=0.000$), problem planning ($p=0.000$), plan execution ($p=0.000$), and problem evaluation ($p=0.000$) and a significant assessment by cohort within-subjects contrast effect (see Table 4.7) for three of the four
problem-solving skills: problem definition ($p=0.002$), plan execution ($p=0.009$), and problem evaluation ($p=0.004$).

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>$F$</th>
<th>Hypothesis $df$</th>
<th>Error $df$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.034</td>
<td>490.87</td>
<td>4.00</td>
<td>70.00</td>
<td>.000</td>
</tr>
<tr>
<td>Cohort</td>
<td>.25</td>
<td>10.66</td>
<td>12.00</td>
<td>185.49</td>
<td>.000</td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment</td>
<td>.24</td>
<td>16.81</td>
<td>12.00</td>
<td>62.00</td>
<td>.000</td>
</tr>
<tr>
<td>Assessment by Cohort</td>
<td>.32</td>
<td>2.44</td>
<td>36.00</td>
<td>183.91</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Table 4.5.* Multivariate analysis of variance of problem-solving skills: problem definition, problem planning, plan execution, and problem evaluation assessments with repeated measures by cohort

<table>
<thead>
<tr>
<th>Source</th>
<th>Measure</th>
<th>Type III Sum of Squares</th>
<th>$df$</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>ProbDefinition</td>
<td>425.54</td>
<td>1</td>
<td>425.54</td>
<td>554.85</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>ProbPlanning</td>
<td>405.99</td>
<td>1</td>
<td>405.99</td>
<td>1248.29</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>PlanExecution</td>
<td>413.75</td>
<td>1</td>
<td>413.75</td>
<td>1024.18</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>ProbEvaluation</td>
<td>119.09</td>
<td>1</td>
<td>119.09</td>
<td>443.39</td>
<td>.000</td>
</tr>
<tr>
<td>Cohort</td>
<td>ProbDefinition</td>
<td>31.60</td>
<td>3</td>
<td>10.54</td>
<td>13.74</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>ProbPlanning</td>
<td>16.28</td>
<td>3</td>
<td>5.43</td>
<td>16.69</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>PlanExecution</td>
<td>47.60</td>
<td>3</td>
<td>15.87</td>
<td>39.28</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>ProbEvaluation</td>
<td>15.19</td>
<td>3</td>
<td>5.06</td>
<td>18.85</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Table 4.6.* Univariate analyses of variance of problem definition, problem planning, plan execution, and problem evaluation problem-solving subskill by cohort between-subject effects
Table 4.7. Univariate analyses of variance of problem definition, problem planning, plan execution, and problem evaluation problem-solving subskill by cohort within-subject effects

Problem-definition skills.

Repeated measures. Due to the significant assessment by cohort interaction effects, the focus for interpretation will be on the within-subject results. Pairwise comparisons of assessment scores resulted in significant differences within-subjects for problem-definition skills for an assessment by cohort interaction ($p=0.002$) see Table 4.7. For the most part all cohorts have the same general pattern across assessments with a few exceptions.

According to Table 3.1, the greatest emphasis on the problem-definition skill studied, justify topic choice (A3), took place between the first ($O_1$) and second assessment ($O_2$), during the $X_1$ treatment period, including 5 activities, as compared to 3 activities for the other two treatment periods ($X_2$ and $X_3$). Additionally, because this subskill was being introduced in the first treatment interval ($X_1$), more classroom time was spent teaching to scaffold this subskill.
Therefore, the steeper slopes representing larger growth for Cohort 1, Cohort 2, and Cohort 3, as compared to Cohort 4, may be due to the relative ease with which these three cohorts were able to apply problem-definition skills as compared to the lower-level cohort, Cohort 4. As we continue to look at change from the second \((O_2)\) and third assessment \((O_3)\), during the \(X_3\) treatment period, Cohort 1 is the only cohort with a positive slope. This difference may be due to differences in problem-assessment tasks, which require abilities that only the highest-level cohort, Cohort 1, possessed.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Cohort 4</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Airplanes</td>
<td>(0.83)</td>
<td>(1.15)</td>
<td>(0.88)</td>
<td>(0.59)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>(M) ((+/0.51))</td>
<td>((+/0.61))</td>
<td>((+/0.89))</td>
<td>((+/0.71))</td>
<td>((+/0.70))</td>
<td></td>
</tr>
<tr>
<td>Mousetrap Cars</td>
<td>(1.50)</td>
<td>(1.58)</td>
<td>(1.56)</td>
<td>(0.82)</td>
<td>(1.39)</td>
</tr>
<tr>
<td>(M) ((+/0.62))</td>
<td>((+/0.64))</td>
<td>((+/0.73))</td>
<td>((+/0.88))</td>
<td>((+/0.76))</td>
<td></td>
</tr>
<tr>
<td>Catapults</td>
<td>(1.61)</td>
<td>(1.38)</td>
<td>(1.25)</td>
<td>(0.53)</td>
<td>(1.22)</td>
</tr>
<tr>
<td>(M) ((+/0.70))</td>
<td>((+/0.70))</td>
<td>((+/0.78))</td>
<td>((+/0.72))</td>
<td>((+/0.81))</td>
<td></td>
</tr>
<tr>
<td>Rube-Goldberg</td>
<td>(1.94)</td>
<td>(1.62)</td>
<td>(1.25)</td>
<td>(0.65)</td>
<td>(1.40)</td>
</tr>
<tr>
<td>(M) ((+/0.24))</td>
<td>((+/0.57))</td>
<td>((+/0.86))</td>
<td>((+/0.86))</td>
<td>((+/0.80))</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>(1.47)</td>
<td>(1.43)</td>
<td>(1.23)</td>
<td>(0.65)</td>
<td></td>
</tr>
<tr>
<td>(M) ((+/0.10))</td>
<td>((+/0.09))</td>
<td>((+/0.11))</td>
<td>((+/0.11))</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(Table\ 4.8. \text{Mean}\ (M)\) \text{and standard deviation (SD)} for problem-definition skills by assessment by cohort and overall across assessments and cohorts
Figure 4.5. Change in mean problem-definition scores by assessment by cohort and cohorts combined (df=9)

**Cohorts.** Overall problem-definition scores matched predictions for all four Cohorts. Cohort 1 had the highest overall mean ($M=1.47$, $SD=0.10$), followed by Cohort 2 ($M=1.43$, $SD=0.09$), Cohort 3 ($M=1.23$, $SD=0.11$), and Cohort 4 ($M=0.65$, $SD=0.11$) with the lowest mean. Because of assessment by cohort interactions, t-tests, with corrections for multiple comparisons, were conducted to compare cohorts over the various assessments.
<table>
<thead>
<tr>
<th>Assessment #</th>
<th>Cohort</th>
<th>Cohort</th>
<th>Cohort</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>0.10</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
<td>0.47</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td></td>
<td>0.16</td>
</tr>
</tbody>
</table>

Table 4.9. P-values for T-tests, with corrections for multiple comparisons, comparing problem-definition mean scores by cohort by assessment. Bolded values are significant at the \( p=0.05 \) level

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 4</th>
<th>Cohort 1</th>
<th>Cohort 3</th>
<th>Cohort 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( O_1 )</td>
<td>0.59</td>
<td>0.83</td>
<td>0.88</td>
<td>1.15</td>
</tr>
<tr>
<td>( O_2 )</td>
<td>0.82</td>
<td>1.50</td>
<td>1.56</td>
<td>1.58</td>
</tr>
<tr>
<td>( O_3 )</td>
<td>0.53</td>
<td>1.25</td>
<td>1.38</td>
<td>1.61</td>
</tr>
<tr>
<td>( O_4 )</td>
<td>0.65</td>
<td>1.25</td>
<td>1.62</td>
<td>1.94</td>
</tr>
</tbody>
</table>

Figure 4.6. Display of mean problem-definition skill score differences by assessment by cohort. The mean values underscored by a common line are not significantly different while all other mean differences are significant. Means obtained from Table 4.8 and significance data from Table 4.9
As the data show (see Table 4.9 and Figure 4.6), there are quite a few significant differences between cohorts, and these incidences tend to increase over the course of the study, possibly due to a greater growth in the higher-level cohort problem-definition skills. This lack of growth in Cohort 4 would also be consistent with a lack of transition from Piaget’s (1964) concrete operational to formal operational stages, as students with the lowest cognitive development would have difficulty applying physical science concepts from class to specific outcomes in the problem-solving assessments.

**Problem-planning skills.**

*Repeated measures.* The overall means for problem-planning skills across the cohorts increased for each of the four assessments (see Table 4.10 and Figure 4.7), with a significant main effect ($p=0.000$). However, pairwise comparisons of the four assessments only showed significant differences (at the 0.05 level) in problem-planning skills between the first assessment ($O_1$) and assessments three ($O_3$) and four ($O_4$). This difference may suggest significant improvement between the beginning and end of the trial period in problem-planning skills (consisting of subskills B2, B6, B7, and B8). The greatest number of subskills addressed by problem-planning activities (22.3) occurred between Assessment 2 and Assessment 3 (see Table 3.1). This emphasis may be responsible for the significant difference between scores for Assessment 1 ($M=0.97$) and Assessment 3 ($M=1.26$), which appeared to be most similar in task and difficulty.
<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Cohort 4</th>
<th>Overall Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Paper Airplanes</td>
<td>$M$ 1.26 (+/-0.35)</td>
<td>1.01 (+/-0.32)</td>
<td>1.00 (+/-0.47)</td>
<td>0.56 (+/-0.39)</td>
<td>0.97 (+/-0.44)</td>
</tr>
<tr>
<td>2 Mousetrap Cars</td>
<td>$M$ 1.40 (+/-0.46)</td>
<td>1.27 (+/-0.59)</td>
<td>1.06 (+/-0.48)</td>
<td>0.62 (+/-0.61)</td>
<td>1.11 (+/-0.61)</td>
</tr>
<tr>
<td>3 Catapults</td>
<td>$M$ 1.54 (+/-0.44)</td>
<td>1.29 (+/-0.43)</td>
<td>1.31 (+/-0.60)</td>
<td>0.87 (+/-0.49)</td>
<td>1.26 (+/-0.53)</td>
</tr>
<tr>
<td>4 Rube-Goldberg</td>
<td>$M$ 1.56 (+/-0.28)</td>
<td>1.47 (+/-0.46)</td>
<td>1.39 (+/-0.58)</td>
<td>1.09 (+/-0.48)</td>
<td>1.39 (+/-0.48)</td>
</tr>
<tr>
<td>Overall</td>
<td>$M$ 1.44 (+/-0.07)</td>
<td>1.26 (+/-0.06)</td>
<td>1.19 (+/-0.07)</td>
<td>0.78 (+/-0.07)</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.10. Mean ($M$) and standard deviation ($SD$) for problem-planning skills by assessment by cohort and overall across assessments and cohorts.*

*Figure 4.7. Change in mean problem-planning scores by assessment by cohort and cohorts combined ($df=9$)*
**Cohorts.** Problem-planning skill scores matched predictions for all four Cohorts (see Figure 4.8). Cohort 1 had the highest mean ($M=1.44$, $SD=0.07$), followed by Cohort 2 ($M=1.26$, $SD=0.06$), Cohort 3 ($M=1.19$, $SD=0.07$) and Cohort 4 ($M=0.78$, $SD=0.07$) with the lowest mean. Pairwise comparisons of problem-planning skill scores showed significant differences at the 0.05 level between Cohort 4 and all other cohorts with Cohort 4 being significantly lower. The relatively higher scores for Cohorts 1, 2, and 3, but low scores for Cohort 4 once again highlight the lower skill level of Cohort 4. In spite of the lower skill level of Cohort 4, significant improvement in problem-planning skill scores for Cohort 4 was shown from Assessment 1 to Assessment 4 ($p=0.00$) as was also the case for Cohort 1 ($p=0.01$), Cohort 2 ($p=0.00$), and Cohort 3 ($p=0.04$).

![Figure 4.8](image_url). Comparing mean scores and standard deviations across the four assessments for problem-planning skill scores by cohort
Plan-execution skills.

Repeated measures. Due to the significance assessment by cohort interaction effects, the focus for interpretation will be on the within-subject results. Pairwise comparisons of assessment scores resulted in significant differences within subjects for problem-execution skills for an assessment by cohort interaction ($p=0.009$) see Table 4.7.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Cohort 4</th>
<th>Overall Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Paper Airplanes</td>
<td>M 1.39</td>
<td>M 1.46</td>
<td>M 0.87</td>
<td>M 0.53</td>
<td>M 1.12</td>
</tr>
<tr>
<td></td>
<td>(+/-0.50)</td>
<td>(+/-0.65)</td>
<td>(+/-0.34)</td>
<td>(+/-0.51)</td>
<td>(+/-0.65)</td>
</tr>
<tr>
<td>2 Mousetrap Cars</td>
<td>M 1.61</td>
<td>M 1.35</td>
<td>M 0.44</td>
<td>M 0.59</td>
<td>M 1.05</td>
</tr>
<tr>
<td></td>
<td>(+/-0.50)</td>
<td>(+/-0.75)</td>
<td>(+/-0.51)</td>
<td>(+/-0.71)</td>
<td>(+/-0.79)</td>
</tr>
<tr>
<td>3 Catapults</td>
<td>M 1.94</td>
<td>M 1.92</td>
<td>M 0.81</td>
<td>M 1.06</td>
<td>M 1.51</td>
</tr>
<tr>
<td></td>
<td>(+/-0.24)</td>
<td>(+/-0.27)</td>
<td>(+/-0.66)</td>
<td>(+/-0.90)</td>
<td>(+/-0.74)</td>
</tr>
<tr>
<td>4 Rube-Goldberg</td>
<td>M 1.72</td>
<td>M 1.23</td>
<td>M 1.13</td>
<td>M 0.82</td>
<td>M 1.23</td>
</tr>
<tr>
<td></td>
<td>(+/-0.46)</td>
<td>(+/-0.65)</td>
<td>(+/-0.72)</td>
<td>(+/-0.88)</td>
<td>(+/-0.74)</td>
</tr>
<tr>
<td>Overall</td>
<td>M 1.67</td>
<td>M 1.49</td>
<td>M 0.81</td>
<td>M 0.75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(+/-0.08)</td>
<td>(+/-0.06)</td>
<td>(+/-0.08)</td>
<td>(+/-0.08)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.11. Mean ($M$) and standard deviation ($SD$) for plan-execution skills by assessment by cohort and overall across assessments and cohorts.

Plan-execution scores consisted of only one expert level skill, appropriate method of data presentation selected (C2). Mean scores for plan-execution skills did not match predictions consistently from Assessment 1 to Assessment 4 (see Table 4.11 and Figure 4.9). One possible reason for this lack of relationship is the difference in tasks set forth for the four problem-solving assessments. The tasks for the first and third problem-solving assessment were very similar in the data that was to be collected, distance and accuracy. The data that was collected for Assessment 2, distance, and Assessment 4, percent success, were
different not only from the Assessment 1 and 3 tasks, but from each other as well.

Figure 4.9. Change in mean plan-execution scores by assessment by cohort and cohorts combined (df=9)

By examining Figure 4.9, there appear to be a couple of instances where cohorts do not match the general pattern. Between the first (O₁) and second assessments (O₂), Cohort 1 has a positive slope while the other cohorts have negative (Cohort 2 and Cohort 3) or nearly flat (Cohort 4) slopes. Once again, this could be due to a more difficult task, which required abilities that only Cohort 1 possessed. Also, between the third (O₃) and fourth assessments (O₄), Cohort 4 is the only cohort that has a positive slope. While at first glance this result does not make sense, upon further review, there may be a plausible explanation. This
task (O4) required students to collected data regarding how often specific steps in the Rube-Goldberg worked as planned. While this is rather simple data to collect, it is not inconceivable that the higher-level cohorts (Cohort 1, Cohort 2 and, Cohort 3), felt the need to over think the data collection process, thus collecting more and inappropriate data.

**Cohorts.** Once again means for all four cohorts matched predictions. The relatively higher scores for Cohorts 1 (\(M=1.67, \ SD=0.08\)), Cohort 2 (\(M=1.49, \ SD=0.06\)), but lower scores for Cohort 3 (\(M=0.81, \ SD=0.08\)), and Cohort 4 (\(M=0.75, \ SD=0.08\)) may highlight the higher skill level of the students enrolled in Honors Language Arts/ Honors Global Studies. In fact, for Assessments 1, 2, and 3 the Cohorts 1 and 2 were significantly higher than Cohorts 3 and 4 (see Table 4.12 and Figure 4.10).
Table 4.12. P-values for T-tests, with corrections for multiple comparisons, comparing plan-execution mean scores by cohort by assessment. Bolded values are significant at the \( p=0.05 \) level.

<table>
<thead>
<tr>
<th>Assessment #</th>
<th>Cohort 1</th>
<th>Cohort 3</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Figure 4.10. Display of mean plan-execution skill score differences by assessment by cohort. The mean values underscored by a common line are not significantly different while all other mean differences are significant. Means obtained from Table 4.11 and significance data from Table 4.12.
Problem-evaluation skills.

Repeated measures. Due to the significant problem-evaluation skills assessment by cohort interaction effects, the focus for interpretation will be on the within-subject results. Pairwise comparisons of assessment scores resulted in significant differences within subjects for problem-evaluation skills for the assessment by cohort interaction (p=0.004) see Table 4.7.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 1 M (+/-SD)</th>
<th>Cohort 2 M (+/-SD)</th>
<th>Cohort 3 M (+/-SD)</th>
<th>Cohort 4 M (+/-SD)</th>
<th>Overall M (+/-SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Airplanes 1</td>
<td>0.44 (+/-0.27)</td>
<td>0.56 (+/-0.45)</td>
<td>0.47 (+/-0.29)</td>
<td>0.16 (+/-0.19)</td>
<td>0.42 (+/-0.36)</td>
</tr>
<tr>
<td>Mousetrap Cars 2</td>
<td>0.71 (+/-0.48)</td>
<td>0.41 (+/-0.37)</td>
<td>0.35 (+/-0.37)</td>
<td>0.16 (+/-0.22)</td>
<td>0.41 (+/-0.41)</td>
</tr>
<tr>
<td>Catapults 3</td>
<td>1.10 (+/-0.61)</td>
<td>0.99 (+/-0.41)</td>
<td>0.61 (+/-0.50)</td>
<td>0.15 (+/-0.19)</td>
<td>0.75 (+/-0.57)</td>
</tr>
<tr>
<td>Rube-Goldberg 4</td>
<td>1.20 (+/-0.39)</td>
<td>1.26 (+/-0.34)</td>
<td>0.89 (+/-0.46)</td>
<td>0.66 (+/-0.55)</td>
<td>1.04 (+/-0.49)</td>
</tr>
<tr>
<td>Overall</td>
<td>0.86 (+/-0.06)</td>
<td>0.80 (+/-0.05)</td>
<td>0.58 (+/-0.07)</td>
<td>0.28 (+/-0.06)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.13. Mean (M) and standard deviation (SD) for problem-evaluation skills by assessment by cohort and overall across assessments and cohorts

In examining Table 4.13 and Figure 4.11, there are a few instances when individual cohorts do not match the general pattern. Between Assessment 1 and Assessment 2, Cohort 1 stands out, as it is the only cohort with a positive slope. Once again, this may be due to a more difficult task requiring abilities that only Cohort 1 possessed. Between Assessment 2 and Assessment 3, Cohort 4 is the only cohort that does not have a positive slope. This may be due to slightly more difficult task requiring abilities that only Cohort 4 lacks.
Cohorts. Overall problem-solving evaluation skill scores matched predictions for all four Cohorts. Cohort 1 had the highest mean ($M=0.86$, $SD=0.06$), followed by Cohort 2 ($M=0.80$, $SD=0.05$), Cohort 3 ($M=0.58$, $SD=0.05$), and Cohort 4 ($M=0.28$, $SD=0.06$) with the lowest mean. Because of assessment by cohort interactions, t-tests, with corrections for multiple comparisons, were conducted to compare cohorts over the various assessments (see Table 4.14 and Figure 4.12).
<table>
<thead>
<tr>
<th>Assessment #</th>
<th>Cohort</th>
<th>Cohort</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.12</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 4.14. P-values for T-tests, with corrections for multiple comparisons, comparing problem-evaluation mean scores by cohort by assessment. Bolded values are significant at the $p=0.05$ level.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Cohort 4</th>
<th>Cohort 1</th>
<th>Cohort 3</th>
<th>Cohort 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1$</td>
<td>0.16</td>
<td>0.44</td>
<td>0.47</td>
<td>0.56</td>
</tr>
<tr>
<td>$O_2$</td>
<td>0.16</td>
<td>0.35</td>
<td>0.41</td>
<td>0.71</td>
</tr>
<tr>
<td>$O_3$</td>
<td>0.15</td>
<td>0.61</td>
<td>0.99</td>
<td>1.10</td>
</tr>
<tr>
<td>$O_4$</td>
<td>0.66</td>
<td>1.89</td>
<td>1.20</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Figure 4.12. Display of mean problem-evaluation skill score differences by assessment by cohort. The mean values underscored by a common line are not significantly different while all other mean differences are significant. Means obtained from Table 4.13 and significance data from Table 4.14.
Summary of problem-solving skills. Both problem-planning skill scores and problem-evaluation skill scores showed an increase in mean scores over the course of the four assessments. Problem-definition skills and plan-execution skills fluctuated, but the final assessments were higher than the initial assessments. Results of a t-test, with corrections for multiple comparisons showed significant differences between Assessment 1 and Assessment 4 for problem-definition scores ($p=0.000$), problem-planning scores ($p=0.000$), and problem-evaluation scores ($p=0.000$).

Significant assessment by cohort interactions were found for problem-definition scores ($p=0.002$), plan-execution scores ($p=0.009$), and problem-evaluation scores ($p=0.004$). Furthermore, when looking at graphs to examine assessment by cohort interactions, differences from the general pattern often matched the cohort level; Higher-level cohorts, especially Cohort 1, would often show a positive slope when others did not, while lower-level cohorts, especially Cohort 4, would often show a negative or flat slope when the others did not.

Although not always significant, mean scores for all four problem-solving skill areas matched predictions with Cohort 1 having the highest means, followed by Cohort 2, Cohort 3, and lastly Cohort 4. Additionally, for the skills that showed assessment by cohort interactions problem-definition, plan-execution scores, and problem-evaluation, often showed significant differences between cohorts on many of the assessments. All of these significant results marched Cohort predictions.
Lastly, problem-evaluation had the most expert-level skills and the lowest overall mean scores of all the problem-solving skills ($M=0.66$, $SD=0.53$). Furthermore, a t-test, with corrections for multiple comparisons, showed that problem-evaluation skills scores were significantly lower than scores for problem-definition skills ($p=0.000$), problem-planning skills ($p=0.000$), and plan-execution skills ($p=0.000$).
Chapter 5: Conclusion

Summary of Findings

The Problem-Solving Skills Assessment (PSSA) showed much strength in documenting improvement in overall problem-solving ability and the four individual problem-solving skills: problem definition, problem planning, plan execution, and problem evaluation. The PSSA was also able to differentiate different levels of students based on problem-solving skills. The PSSA also would be quite useable by a classroom teacher without much disruption to the classroom environment. However, consistent results were not always found from one assessment to the next.

Problem-solving ability. When using the instrument as a summary of problem-solving ability and problem-solving skills, the instrument did a good job of showing differential improvement over time for the different cohorts of students. The means of the four assessments and the four cohorts each appeared to match the prediction 100% of the time, although there were only significant differences 83% of the time. This is quite a high percentage
considering the homogeneity of the instruction for the four cohorts. All of the students were at the same suburban high school and all had the same four core teachers.

**Problem-solving skills.** Looking at the four main problem-solving skills, the instrument still held up rather well with means that appeared to match predictions for assessment (79%) and cohorts (100%) most of the time. Additionally, for three of the problem-solving skills: problem definition, plan execution, and problem evaluation, there was a significant assessment by cohort interaction effect. Furthermore, for these three problem-solving skills, differences between cohorts were often significant for each of the four assessments. The instrument demonstrates the ability to measure problem-solving skills well enough to use in the classroom setting.

Two problem-solving skill areas, problem definition and plan execution, both showed inconsistencies over the four assessments as mean scores did not continually improve. One reasons for this inconsistency could be due to differences in the tasks used for the PSSA. Another possible reason may have been that problem definition and plan execution were the two skills that only had one subskill each, justify topic choice and appropriate method of data presentation selected, respectively. Only having one subskill for each of these areas decreased the precision of this assessment. If teachers were to focus on these skill areas it is recommended that they include some of the novice and intermediate problem-solving skills, such as brainstorm possible topics, and
identify topic choice for problem definition, and appropriate data collected for plan execution.

**Evaluating cohorts vs. assessments.** The instrument was much better at differentiating the four cohorts than showing the improvement over the four assessments. In hindsight this result should have been expected. When comparing cohorts, the students in each cohort did not change. Every comparison was made on an equal playing field. However, when comparing assessments, there were two factors to be considered. First, student problem-solving skills and how these changed over time. Another variable that was not initially taken into account was the difficulty level of the problem-solving assessments. Not only did the problem-solving assessments vary in difficulty overall, but especially in some of the skill areas. For instance, the final problem-solving assessment, the Rube Goldberg, was much more difficult in the plan-execution area as the data that needed to be collected was not obvious. Rather than being an easily measurable and tangible quantity like distance and speed, the best data to collect would have been the percent of time the device worked as planned, which was quite a bit different from most of the data collected during the course of the year.

To get a fairer assessment of growth, it is recommended to use problems that are of the same level of difficulty. Perhaps it would be beneficial to create a classification scheme for problems so that like problems could be compared to one another.
**Variation in problem-solving assessments.** There was more variation in results from problem-solving assessment to problem-solving assessment than expected. That is, in some cases, problem-solving assessment showed a decrease in problem-solving subskills (most notably isolate variables, appropriate method of data collection, and plan next step) from one assessment to the next. While it would not be totally unheard of for some regression to occur, some of this was most likely due to the differences in difficulty of tasks assigned to the students, which will be addressed further in the recommendations.

**Possible Errors**

There are several reasons why the analysis may not have detected statistical significance and may have lacked agreement with predictions. Errors in reliability and precision need to be addressed and dealt with to improve the overall validity of the instrument.

**Lack of reliability.** Because the instrument did not change from task to task, reliability in the context of this experiment was determined by the consistency or repeatability with which the instrument was implemented and skill/subskill scores were marked. First, it is worth noting that when the final assessment was marked and re-marked, all results had a strong positive correlation (the lowest being 0.68). Therefore, any critique here is a bit nit picky to improve data even further. However, the simplest way to increase intra-rater
reliability (and inter-rater reliability should somebody else choose to implement this instrument) is to improve upon the rubric.

The current criteria used for each subskill was simply “no evidence,” “inconclusive evidence,” and “conclusive evidence.” While these three criteria can apply to all the subskills, they can also be quite ambiguous (See Figure 5.1). A rubric could be made for each subskill articulating much more clearly what represents “inconclusive evidence” and “conclusive evidence.” In some cases, there may even be some criteria that must be met to even demonstrate “inconclusive evidence.”

Figure 5.1 provides three examples that use three of the subskills with the lowest Pearson’s Correlation Coefficient in this study. Including more specific and subskill-unique statements would likely improve this rubric.

<table>
<thead>
<tr>
<th>Subskill</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2 - Isolate variables</td>
<td>No evidence</td>
<td>Indirect/inconclusive evidence</td>
<td>Direct/conclusive evidence</td>
</tr>
<tr>
<td>D3 - Find solution for errors in data</td>
<td>No evidence</td>
<td>Indirect/inconclusive evidence</td>
<td>Direct/conclusive evidence</td>
</tr>
<tr>
<td>D6 - Validity of results is supported or refuted</td>
<td>No evidence</td>
<td>Indirect/inconclusive evidence</td>
<td>Direct/conclusive evidence</td>
</tr>
<tr>
<td>D5 - Validity of results is questioned</td>
<td>No evidence</td>
<td>Indirect/inconclusive evidence</td>
<td>Direct/conclusive evidence</td>
</tr>
</tbody>
</table>

Figure 5.1. Subskill scoring rubric with criteria from current study
**Isolate variables.** In order to demonstrate that variables have been isolated, a student must first clearly identify the variables being measured. This statement would not receive any points, but is a requirement to receive points.

“Inconclusive evidence” could be demonstrated by first explaining how the variable will be measured. This shows the ability to at least focus on the variable at hand. In order to demonstrate “conclusive evidence” there must be some description of how other variables are limited.

**Find solution for errors in data.** There are two major points in finding solutions for errors in data. First, the student must be able to show that the solution they have suggested is in fact, related to the error in question. Second, they must support the fact that the solution would likely fix the error. If only one of them is present there is “inconclusive” evidence for this skill. If both are present “conclusive evidence” would be demonstrated.

**Validity of results is supported or refuted.** In order to support or refute validity of results, a student must first state whether or not the data is accurate. This statement would not receive any points, but is a requirement to receive points.

There are two major parts to supporting or refuting validity. First, the student must be able to demonstrate that the data is/is not precise. Second, the student must be able to demonstrate that the data is/is not reliable. If only one of
these is present there is “inconclusive” evidence for this skill. If both are present “conclusive evidence” would be demonstrated.

**Validity of results is questioned.** This topic was quite unreliable and really should not be included in future research as it is nearly impossible to assess under normal classroom circumstances. In order to assess this subskill ample time must be dedicated to questioning a student without asking them directly to question the validity, which this instrument did. Once the teacher asks the student to question the validity, then the student is no longer doing it on their own and it is impossible to tell if the student would have done it on their own. By creating a more reliable template not only should data improve, but teachers should be able to scaffold activities and units better.

Below is an example of a rubric with the above recommendations implemented. Please note that “validity of results is questioned” is not included as per recommendations.
<table>
<thead>
<tr>
<th>Subskill</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2 - Isolate variables</td>
<td>Variables being measured not identified</td>
<td>Student explains how variable will be measured</td>
<td>Student described how other variables are limited</td>
</tr>
<tr>
<td>D3 - Find solution for errors in data</td>
<td>No evidence</td>
<td>Student suggestion is related to error</td>
<td>Student explains how solution would likely fix error</td>
</tr>
<tr>
<td>D6 - Validity of results is supported or refuted</td>
<td>Student does not state whether or not data is accurate</td>
<td>Student explains how data is or is not precise -OR-</td>
<td>Student explains how data is or is not precise AND how data is or is not reliable</td>
</tr>
</tbody>
</table>

Figure 5.2. Subskill scoring rubric with recommendations implemented.

**Lack of precision.** In the context of this study, precision refers to the exactness of mean scores for overall problem-solving skill and each problem-solving subskill. For each of the subskills there are only three possible scores (0, 1, or 2), resulting in mean scores for an individual of 0.0, 0.5, or 1.0. Even though this lack of precision is quite large, it is worth mentioning that subskills were not the focus of this study and that speed and accessibility was a higher priority.

For problem-solving skills, two of the skills, problem identification and plan execution, still lacked precision because there was only one subskill for each. The other two skills, problem planning and problem evaluation, could have any integer score in the range of 0-8 and 0-14 respectively, resulting in mean scores for an individual occurring at intervals of 0.125 and 0.071 respectively.
Increasing precision can help reduce standard deviation, and thus lead to greater statistical significance.

Finally, when examining overall problem-solving abilities, scores could be an integer from 0-26, resulting in mean scores for an individual occurring at intervals of 0.038. Therefore, the larger, problem-solving ability level gives much more precision than any individual subskill.

**Lack of supplemental information.** Although an online discussion group was set-up to help monitor student problem solving outside of the classroom, and serve as a way to gather supplemental information on student problem-solving skills, student use was minimal. There were only a few occasions when students used the forum, and no problem-solving skills were evident, only clarification of instructions, and requests for contact information. Students chose other ways to communicate with each other outside the classroom. More student-friendly modes of technological communication that dominated their use included texting, e-mails, and facebook, all of which are much more private and are nearly impossible to evaluate. Furthermore, several groups met outside of class, not in the classroom, but at another student’s house.

**Student motivation.** Vansteenkiste et al. (2006) describe various motivations that students may have. Their two main categories are “Goal Content” and “Goal Motive.” The activities were selected for problem-solving assessments to enhance both goal content and goal motive. The goal was to
have students want to see their creation succeed, i.e. goal content, and see the value in learning how to solve problems, i.e. goal motive.

However, some students thought the tasks were too hard and thus lost intrinsic motivation. The extrinsic motivation of grades, and a bonus for the “winners” was not motivational for them. This issue was especially apparent in the last problem-solving assessment. At the end of the semester, some of the students had already failed the course, even before the final assessment. These students would not have the extrinsic motivation of grades or controlled motivation of bringing home bad grades to parents, as these are consequences that would occur regardless of their performance on the final assessment.

Because the study did not examine motivation directly, some difficulty arises in determining if students valued the acquisition of problem-solving skills, even given the context of real-life problems for them. The two main motivations appeared to be the fear of taking a bad grade to their parent(s) (controlled motive), or the value in good grades as a stepping-stone to college (autonomous motive). In future research, especially longitudinal studies, tracking motivation along with problem-solving skills would be a good area for further study.

**Question level.** When the study was started it was thought all freshman would be at Piaget’s highest level of development, formal operations stage. One major difference between students in the formal operations stage and the concrete operational stage is the ability to use abstract thought. In the case of problem solving this ability would include relating experimental data collected to
abstract theories from their own knowledge base, a skill of expert problem solvers (Lester & Kehle, 2003). For this reason, it was thought that focusing on expert problem-solving skills would be suitable. Based on the results from Cohort 4, who were generally significantly below the other Cohorts, it may have been wise to include some questions, even if not marked as part of the assessment, that were at the novice and intermediate level to help scaffold learning and guide the students to the more in-depth answers. This scaffolding in questions may have helped with motivation by encouraging students who might otherwise give up.

Overall Validity

As previously mentioned content validity and face validity were dealt with in the preliminary phases of this study. Predictive validity and construct validity were evaluated after the study.

Results regarding the predictive validity of the instrument are inconclusive. One of the difficulties found when assessing predictive validity is determining the time at which students would be able to demonstrate skills they have gained. In some instances the time of skill score improvement matched predictions; however, this was not always the case and may suggest that skill acquisition is a bit more complicated than initially thought.

The PSSA assessment showed rather good construct validity. Scores often showed significant differences between cohorts and significant differences
between assessments. Some fine-tuning of the instrument could improve the construct validity.

**Recommendations**

1. With the biggest weakness in precision relating to problem definition and plan execution, it is recommended that in future research other novice and intermediate subskills be included in evaluating these skills, such as brainstorm possible topics, identify topic choice, and appropriate data collected.

2. To aid in the instruction of problem-solving skills, creating a unique rubric for all documented subskills is recommended. By documenting novice intermediate subskills, this additional data would be especially helpful for lower level students that are unable to demonstrate any expert problem-solving skills.

3. Additionally, with a more complete rubric, a teacher could focus on just one problem-solving skill, and its multiple subskills, to help scaffold instruction at the beginning of a course and then use the complete instrument only in the middle and end of a course.

4. The PSSA aided in the differentiation of the cohorts in this study rather well. However, these cohorts were largely determined by courses other than science. Therefore it is recommended that this study be conducted again in a situation in which students are tracked by science ability.
5. With today’s online grading software, documenting these various problem-solving skills through an online grading program could be a useful way to pass on information to a student’s teachers in later years to help monitor progress and create more useful instruction.

6. Develop a classification for problems so that results from like assessments could be compared with each other and not with unlike assessments. For instance, some problems could be classified as “shallow” or “deep” based on the ease of identifying variables. Also, some problems could be classified as “obvious” or “hidden” based on the ease with which data can be collected.

7. Develop a better method to document online student work outside of the classroom. This can be done in one of two ways. The first way is to encourage students to use a teacher moderated forum better. The second way would be to encourage students to include the teacher in communication (e.g. CC the teacher on all e-mails).

8. Lastly, once the instrument is further refined, an interesting study would be to see if improving problem-evaluation skills improves the other three problem-solving skills. Problem-evaluation skills require reflection back to earlier steps in the process, which would seem to reinforce all other skills.

**Final Thoughts**

The findings of this study are consistent with past research in many aspects. First of all, based on the low scores for problem-evaluation skills,
Linn's (1995) statements regarding the need for explicit teaching of self-evaluation skills is supported. Furthermore, other aspects that differentiate novices from experts are supported. Smith and Good (1984), note that experts often base the best solution on the ability to justify answers. The ability to differentiate cohorts based on expert skills in this study, which included justify topic choice, evidence used to support results, and validity of results supported or refuted, supports Smith and Good's statements, as well as beliefs by Lester and Kehle (2003) that experts exhibit more self-regulatory behaviors. It would be hoped that the Problem-Solving Skill Assessment could be used in the future to help articulate differences between experts and novices even better, and help evaluate various methods for scaffolding problem-solving instruction.
References


Appendix A: Classical Assessment: Problem-Solving Skills Assessment

Oral Script
Problem Definition
(Justify Topic Choice)
• Why did you choose your topic?
• What are some of the topics you did not choose? Why didn't you choose these topics?

Problem Planning
(End product envisioned)
• What do you expect to find out from your experiment?
• What do you expect your data to look like?
Appendix B: Supplemental Assessment: Problem-Solving Skills

Assessment Written Script
Problem Definition
(Justify Topic Choice)
• Why did you choose your topic?
• What are some of the topics you did not choose? Why didn't you choose these topics?

Problem Planning
(Isolate Variables)
(Choose Appropriate Procedures)
• What were your variables?
• What procedures did you use to test these variables?
• What confounding variables were you able to avoid? How did you avoid them?

Problem Evaluation
(Evidence used to support results, validity of results questioned. validity of results supported or refuted)
• What do you KNOW you found out?
• What do you THINK you found out?
• Can you pinpoint how your data supports this claim?
• Is your data "good?" (accurate, reliable and precise)
• How do you know it is good?

(Find errors in data, find causes for errors, find solution for errors in data)
• Can you identify any errors in your in your data? (Elaborate)
• What do you think caused these errors?
• Can these errors be corrected if somebody else did the experiment? How?

(Plan next step)
• What questions does this study leave you with?
• What would you like to find out next?
• How do you think the information you learned is useful to anybody else?
Appendix C: Pearson Correlation Coefficients to Determine Intra-rater Reliability
All Problem-Solving Subskills - 0.8468**
Mean Problem-Solving Skills - Pearson Correlation 0.960**

Skills
Problem Definition - 0.826**
Problem Planning - 0.897**
Plan Execution - 0.918**
Problem Evaluation - 0.924**

Subskills
Justify Topic Choice - Pearson Correlation 0.826**
Isolate Variables - 0.713**
Choose Appropriate Procedures - 0.826**
Questions Developed Regarding End Product - 0.866**
End Product Envisioned - 0.882**
Appropriate Method of Data Presentation Selected - 0.918**
Find Errors in Data - 0.815**
Find Cause For Errors in Data - 0.883**
Find Solution for Errors in Data - 0.766**
Evidence Used to Support Results - 0.921**
Validity of Results Questioned - 0.682**
Validity of Results Supported or Refuted - 0.799**
Plan Next Step - 0.801**
Appendix D: Repeated Measures for Problem-Solving Ability
### Pairwise Comparisons

**Measure:** PAVG

<table>
<thead>
<tr>
<th>(I) Assessment</th>
<th>(J) Assessment</th>
<th>Mean Difference (I-J)</th>
<th>Std. Error</th>
<th>Sig. a</th>
<th>95% Confidence Interval for Difference*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>-.073</td>
<td>.044</td>
<td>.591</td>
<td>Lower Bound: -.192, Upper Bound: .045</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-.309 *</td>
<td>.048</td>
<td>.000</td>
<td>Lower Bound: -.438, Upper Bound: -.180</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.073</td>
<td>.044</td>
<td>.591</td>
<td>Lower Bound: -.045, Upper Bound: .192</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-.236 *</td>
<td>.049</td>
<td>.000</td>
<td>Lower Bound: -.369, Upper Bound: -.102</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.309 *</td>
<td>.048</td>
<td>.000</td>
<td>Lower Bound: .180, Upper Bound: .438</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-.428 *</td>
<td>.046</td>
<td>.000</td>
<td>Lower Bound: -.554, Upper Bound: -.302</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>.501 *</td>
<td>.042</td>
<td>.000</td>
<td>Lower Bound: .388, Upper Bound: .615</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>.428 *</td>
<td>.046</td>
<td>.000</td>
<td>Lower Bound: .302, Upper Bound: .554</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>.192 *</td>
<td>.046</td>
<td>.000</td>
<td>Lower Bound: .069, Upper Bound: .316</td>
</tr>
</tbody>
</table>

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

* The mean difference is significant at the