Copyright

by

Luis Fonseca Tinoca

2004
The Dissertation Committee for Luis Fonseca Tinoca
certifies that this is the approved version of the following dissertation:

**From Professional Development for Science Teachers**

to **Student Learning in Science**

**Committee:**

______________________________________
James P. Barufaldi, Supervisor

______________________________________
Lowell J. Bethel

______________________________________
Edmund T. Emmer

______________________________________
Jack S. Turner

______________________________________
Julie A. Luft

______________________________________
Kamil A. Jbeily
From Professional Development for Science Teachers
to Student Learning in Science

by

Luis Fonseca Tinoca, Lic.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2004
DEDICATION

To my family.

Coming to Texas has been an incredible journey only possible thanks to the close
and constant support of my wonderful wife, and the distant but always present
encouragement of my parents.

To my children, Rafael and Carolina, my true muses. One look at them is all I
need to energize me to write and work

To the memory of my grandparents, António Silva Tinoca and Clarinda Vaz
Pereira da Fonseca, who passed away during our stay in Texas.
ACKNOWLEDGEMENTS

This work was only possible thanks to the continuous support of my wife Claudia. She has always been by my side since the moment we decided to embark in this Texas adventure. She cheered and encouraged me when I was stuck, and kept feeding and nourishing me when I was working on all cylinders. She gave me as much support as humanly possible. Thank you Claudia, you are truly my better half, and my achievements are a reflection of the person you help me be.

Professor James P. Barufaldi is the best academic role model that I could have wished for. He is highly knowledgeable in the field of science education, and a true Educator. He has been a lot more than a supervisor; he was always available to meet with me and has helped me overcome the many hurdles in my work. His guidance is focused and encouraging, and if today I feel prepared to take the next step in my academic career it is mainly thanks to him.

Professor Bethel is an excellent graduate advisor. He, also, helped steer my progress through graduate school in a foreign environment with great success. Moreover, he was a constant presence of support and encouragement as I navigated my way across this program from the first to the last day.

I must also thank Professors Turner, Emmer, Luft and Jbeily, for their feedback and contributions to my work. They all contributed with their expertise at different levels and were supportive of this study.
Finally, I must acknowledge the invaluable support of the Fundação para a Ciência e Tecnologia, part of the Portuguese Ministério da Ciência e do Ensino Superior, in the form of the scholarship FCT BD/1007/2000, for their continuous financial support throughout my doctoral program.
This study investigates the effects of professional development for science teachers on student learning. It is usually expected that professional development programs positively impact student learning, however this dimension is not commonly incorporated in the programs evaluation. It is simply assumed that students will be indirectly impacted through their participating teachers in the work with their students. Two main research questions are addressed: 1) Are professional development programs effective in enhancing student learning in science? 2) What are the characteristics of the most and least effective programs?

To answer these questions a meta-analysis of 37 professional development programs reporting their impact on student learning was performed. Program characteristics have been defined according to the categories defined by Loucks-Horsley
et al (1998), the National Science Education Standards (NRC, 1996), as well as new categories developed by us analyzing other variables such as the programs length.

A significant impact of professional development for science teachers on student learning has been found in the form of an overall correlation effect size of $r = 0.22$ ($p<0.001$). Moreover, a Fixed Effects Model was used to differentiate between the impacts of the different characteristics of professional development programs for science teachers. In particular, programs emphasizing work on curriculum development, replacement, or implementation, scientific inquiry, pedagogical content knowledge, lasting over 6 month and with a total duration of at least 100 hours have been identified as having a larger impact on student learning.

To enhance the findings vignettes have been developed based on the attained effect sizes describing possible professional development programs. Recommendations for present and future professional development programs are made based on what works best in order to maximize their impact on student learning.
TABLE OF CONTENTS

List of Tables                                                                 xii
List of Figures                                                               xv

Chapter 1: Introduction                                                      1
  Rationale                                                                  2
  Theoretical Framework                                                      5
  Problem                                                                   11
  Purpose                                                                   12
  Research questions                                                        13
  Assumptions                                                               13
  Importance of the Study                                                   14
  Definitions                                                               15
  Limitations                                                               17
  Organization                                                             18

Chapter 2: Review of the Literature                                          19
  Introduction                                                              19
  Theoretical framework                                                     21
  The professional Development Domain.                                      22
  The Teachers’ Domain                                                      29
  The Students’ Domain                                                      36
  Professional development worldwide                                        41
A model professional development program 45
Summary 49

Chapter 3: Methodology 50
Introduction 50
Rationale for the use of meta-analysis 50
Description of meta-analysis 51
Sampling Procedures 53
Data collection 55
Coding procedure 62
Effect size computation 62
Corrections 67
Treatment of the data/ Statistical methods 69
Vignettes 72
Pilot Study 74
Summary 76

Chapter 4: Data Analysis 77
Introduction 77
Description of the Sample 77
Vote counting 80
Fixed effects model effect sizes 80
Summary 95

Chapter 5: Discussion, Implications and Further Research 97
Introduction

Discussion

Findings portrayed in Vignettes

Implications

Further research

Conclusions

Appendix A

Coding Form

Appendix B

Example of coding for the effect size level

Appendix C

Sample Vignette

Appendix D

SPSS Macro for meta-analysis

Bibliography

Vita
List of Tables

Table 1.1: Relationship Between Selected Resource Variables and Student Achievement on NAEP Assessments 3

Table 2.1: Principles of effective professional development (adapted from Loucks-Horsley et al, 1998) 28

Table 2.2: The Concerns-Based Adoption Model (CBAM) 31

Table 2.3: Levels of Use 31

Table 2.4: A Paradigm for Professional Development in Learner Centered Schools 35

Table 2.5: Collaboration 36

Table 2.6: Core principles of the Texas Regional Collaboratives 46

Table 2.7: Comparison of Loucks-Horsley, et al. (1998) Principles to TRC 48

Table 3.1: Strategies for Professional Learning 57

Table 3.2: Changes in Success Rates (BESD) corresponding to various values of \( r^2 \) and \( r \) 66

Table 3.3: Summary of the sampled studies for the pilot study 75

Table 4.1: Study characteristics 78

Table 4.2: Studies and calculated effect size 79

Table 4.3: Overall Effect size (Zr) and descriptive statistics 81

Table 4.4: Overall Effect size (r) 82
Table 4.5: Effect size (Zr) and descriptive statistics for the group of studies with maximum confidence in the estimation of their effect size

Table 4.6: Effect size (Zr) and descriptive statistics for the group of studies with less than the maximum confidence in the estimation of their effect size

Table 4.7: Effect size (Zr) and descriptive statistics for the type of treatment variable

Table 4.8: Effect size (r) for the type of treatment variable

Table 4.9: Effect size (Zr) and descriptive statistics for the theoretical dimension variable

Table 4.10: Effect size (r) for the theoretical dimension variable

Table 4.11: Effect size (Zr) and descriptive statistics for the Additional context variable

Table 4.12: Effect size (r) for the additional context variable

Table 4.13: Effect size (Zr) and descriptive statistics for the Spread variable

Table 4.14: Effect size (r) for the spread variable

Table 4.15: Effect size (Zr) and descriptive statistics for the total duration variable

Table 4.16: Effect size (r) for the total duration variable
Table 4.17: The impact of key professional development characteristics on student learning

\[ \text{95} \]
List of Figures

Figure 1.1: A Model of the process of Teacher Change (Guskey, 1986) 5
Figure 1.2: Professional Development Design Process for Mathematics and Science Education Reform (Loucks-Horsley et al., 1998) 6
Figure 1.3: Professional Development Theoretical Framework 8
Figure 1.4: Focus of the Study 12
Figure 2.1: The Impact of the Developing Professional Developers model in place in the TRC 47
Figure 4.1: Distribution of observed correlation effect sizes $r$ 83
Teacher expertise – what teachers know and can do – affects all the core tasks of teaching. What teachers understand about content and students, for example, shapes how judiciously they select from texts and other materials and how effectively they present material in class. Teachers’ skill in assessing their students’ progress also depends on how deeply teachers know the content, and how well they understand and interpret student talk and written work. Nothing can fully compensate for the weakness of a teacher who lacks the knowledge and skill needed to help students master the curriculum.

(Darling-Hammond & Ball, 1998; p. 1, 2)

One can infer from the above passage that teachers’ knowledge and skills development are imperative for student learning to occur. The purpose of this study is to investigate the effects of professional development for science teachers on student learning. Two main research questions are addressed: 1) Are professional development programs effective in enhancing student learning in science? 2) What are the characteristics of the most and least effective programs?

In order to answer the research questions, a meta-analysis of research on science professional development programs has been conducted to determine the programs’
impact on student learning. Meta-analysis was selected as the statistical technique because this methodology has both the power to produce a synthesis of results from a large group of studies and, at the same time, distinguish between the impacts of the different variables of those studies. Moreover, this technique produces quantitative results.

Finally, the results of this meta-analysis will be used to identify trends in the characteristics of professional development programs. Based on these quantitative findings, vignettes have been developed that provide professional developers with examples of how a program with the characteristics identified are relevant to student learning.

**Rationale**

If the correlation between achievement and learning is accepted, the question becomes, “What factors determine overall student achievement?” Several factors have been identified as contributing to student achievement, i.e., student characteristics such as socio-economic status, limited English proficiency and minority status; per-pupil spending; pupil teacher ratios; class sizes; and teacher quality. Of the afore mentioned factors, teacher quality is the most highly correlated with student learning (Darling-Hammond; 1999) (see Table 1.1).
Table 1.1: Relationship Between Selected Resource Variables and Student Achievement on NAEP Assessments (partial correlations, controlling for student poverty and resource variables) (adapted from Darling-Hammond, 1999; p. 30)

<table>
<thead>
<tr>
<th>Resource variables</th>
<th>Grade 4 Math</th>
<th>Grade 8 Math</th>
<th>Grade 4 Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of teachers well-qualified</td>
<td>.71**</td>
<td>.61**</td>
<td>.75**</td>
</tr>
<tr>
<td>(with full certification and a major in their field)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of teachers out-of-field</td>
<td>-.48*</td>
<td>-.44*</td>
<td>-.32</td>
</tr>
<tr>
<td>(with less than a minor in the field they teach) a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of all teachers fully certified b</td>
<td>.36</td>
<td>.20</td>
<td>.38</td>
</tr>
<tr>
<td>% of all teachers less than b fully certified</td>
<td>-.36</td>
<td>-.23</td>
<td>-.33</td>
</tr>
<tr>
<td>% of new entrants to teaching who are uncertified (excluding transfers)</td>
<td>-.51*</td>
<td>-.39</td>
<td>.43*</td>
</tr>
<tr>
<td>% of all newly hired teachers uncertified</td>
<td>-.40*</td>
<td>-.41*</td>
<td>-.53**</td>
</tr>
<tr>
<td>Per pupil spending</td>
<td>.32</td>
<td>.28</td>
<td>.19</td>
</tr>
<tr>
<td>Pupil:teacher ratio</td>
<td>.03</td>
<td>.22</td>
<td>.09</td>
</tr>
<tr>
<td>Average class size</td>
<td>-.03</td>
<td>.21</td>
<td>-.04</td>
</tr>
</tbody>
</table>

*p<.05, **p<.01

a - For this analysis of mathematics scores, percent of mathematics teachers out-of-field. For analysis of reading scores, percent of English/reading teachers out-of-field, also controlling for LEP status.

b - For this analysis, reading scores control for LEP status.

If, as Darling-Hammond’s work illustrates, teacher quality is one of the most determinant factors for student learning, then one must concentrate on how it can be
promoted and improved. Two ways are generally used to train teachers: pre-service education and professional development. Pre-service education is provided in colleges of science and education, during the student’s undergraduate career. Professional development is used after they have already started teaching and return to learn more and update their knowledge.

“As noted in the extensive body of evidence cited throughout this report, research is confirming that good teaching does matter” (NRC, 2001; p. 4). The Committee on Science and Mathematics Teacher Preparation (NRC, 2001) clearly emphasizes the crucial importance of teachers on student achievement and learning in both science and mathematics. With this in mind, the Committee provides a series of recommendations targeting both pre-service teacher education and teacher professional development, with the intent to improve teacher education due to its ultimate impact on student learning. One of the Committee recommendations states: “New Research that focuses broadly on synthesizing data across studies and linking it to school practice … would be especially helpful to the improvement of teacher education and professional development” (p. 121).

In an earlier work (1999) the National Research Council is also saying that “Research studies are needed to determine the efficacy of various types of professional development activities … extended over time and across broad teacher learning communities” (p. 240).

Without overlooking the importance of pre-service teacher education, no matter how effective it is, what happens when teachers are actually working in the field during their induction and subsequent years seems to be highly determinative of their continuous growth and development as a teacher (Luft, Roehrig, & Patterson, 2003). Therefore, any
study concerned with improving teacher quality, must pay special attention to the kind of professional development that is provided to them. Moreover, if the ultimate goal of improving teacher quality is to consequentially improve student learning, it is imperative to evaluate the impact of professional development programs on student learning.

**Theoretical Framework**

Guskey (1986, 1997) presented a model (Figure 1.1) of teacher change process that teachers go through when participating in professional development programs. This model has the professional development program as the initial triggering mechanism in the change process. However, it recognizes that the student learning outcomes, as observed by the teachers after they have transformed their classroom practices, are a determinant in promoting teachers’ change in beliefs and attitudes.

![Figure 1.1: A Model of the process of Teacher Change (Guskey, 1986)](image-url)

Loucks-Horsley, Hewson, Love, and Stiles (1998) presented a framework (Figure 1.2) for designing professional development programs for teachers of science and
mathematics. This framework also emphasizes the continuous and circular design permeating the implementation of professional development programs. This design is infused by the continuous reflection based on the outcomes of the program to reevaluate and further improve it.

Hein (1997) presented a view of the U.S. Educational System where student learning was the ultimate goal and the teachers were portrayed as the main level just before students through which that goal could be achieved. Nonetheless, according to Hein at that time “the evaluation of Teacher Enhancement efforts must focus on observable outcomes for teachers … (because) the research base that links teacher behavior and student learning is still inadequate” (p. 161).

Figure 1.2: Professional Development Design Process for Mathematics and Science Education Reform (Loucks-Horsley et al., 1998)
One may argue that professional development researchers must place student learning at the center of the educational system, and also that it must inform and influence the steps right above it, teacher professional development in particular, since the teachers are one of the greatest influences on student learning.

Based on the presented models and frameworks for the structure of professional development and its impact on teacher change, and on this researcher’s firm belief that student learning should be the focal point of the educational system, a theoretical framework for professional development was developed (Figure 1.3).

This framework reflects the researcher’s values about what should guide professional development programs. Three concentric domains characterize the framework. It includes the student learning domain, a teachers’ change/evolution domain, and a professional development domain.

Change is a complicated process, not an event (Loucks-Horsley & Roody, 1990). It is not reasonable to expect teachers to change overnight because of their participation in a professional development program. The impact on teachers’ beliefs and attitudes is much more probable to become a reality after they notice an improvement in their student learning outcomes than before. However, intensive and extensive follow up activities of the professional development program are other essential requirements to nurture teacher change.

The professional development domain with four main stages (Set goals, Plan, Do and Reflect) adapted from Loucks-Horsley et al. (1998) is incorporated within this framework. The initial stage is the Goal Setting Phase that should immediately start by
formulating its goals in terms of student learning outcomes, even though there may be other parallel goals such as building teacher leadership and improving teacher retention.

Figure 1.3: Professional Development Theoretical Framework
The second stage is the Plan Phase where the professional development practitioners develop an action plan. This plan is implemented during the Doing Phase when the professional development program actually interacts with the teachers.

The Reflection Phase is a vital part of the professional development program because it is based on the actual results of the implemented professional development program on student learning and teacher’s beliefs. Reviewing, during the Reflection Phase, accommodates necessary changes in the next iteration of every professional development program.

At the center of this framework is the ultimate goal of the professional development framework, the student learning domain. Leading to the goal of student learning is a change or improvement in teacher’s classroom practices, often essential to the improvement of student learning. Immediately after student learning, teacher’s beliefs are included, anchored in a theory of teacher change (Guskey, 1986; Horsley & Loucks-Horsley, 1998).

On the proposed framework, the students’ learning objective is the center and essentially the goal of the entire program. The teachers’ level is immediately above it, including teachers’ classroom practices, behaviors, and beliefs that have the greatest impact on student learning but are also influenced by feedback received from students’ experiences and outcomes. Finally, the professional development domain includes most of the teachers’ domain. It also includes the students’ learning at its core, and will not only influence the teachers and the students’ learning, but will also receive feedback from the students and teacher experiences to restructure the goal setting and planning phases.
The teachers’ domain goes beyond the professional development domain of the framework because professional development programs address only some of the issues that affect a teacher’s life. Situations such as administrative environment, mandated standardized tests, and teachers’ salaries are not within the realm of professional development programs but have important consequences in any teachers’ practice. In the same way, the students’ domain is extended beyond both the teacher and the professional development domains. Also in this case, some very important conditions such as number of students per class and family environment are not within the range of influence of either teachers or professional developers but do influence student learning. Moreover, the dotted lines at the edge of each domain in Figure 1.3 represent the permeability of the domains to external factors such as politics, economy, or mandated curriculum. These external factors, even though difficult to quantify, have an impact on the framework and must be acknowledged.

The next step in this research is the evaluation of professional development programs. The purpose of the evaluation is to identify the characteristics that have a greater impact on student learning. It is the opinion of this researcher that such research should be conducted at a cross-cultural level comparing professional development programs from different states, countries, and cultures. Assessing as many and as different professional development programs as possible may be a fruitful avenue of further investigations. Different cultures and countries have promoted professional development programs with different degrees of success. Learning from their strengths
and weaknesses is imperative. There is much to be gained by comparing a variety of professional development programs.

**Problem**

It is not clear how professional development programs are translated in teachers’ classrooms or how they impact students. This notion is reflected in the few studies that actually permeate the three domains presented in the framework (Figure 1.3).

This research will focus on exploring the part of the framework (Figure 1.3) that moves from what is done in professional development to the change in teachers’ classroom practices and finally to student learning as the ultimate goal (see Figure 1.4). This is the neurological center of the entire framework where the three domains intersect and how what is envisioned during the professional development programs creation impacts the teachers and through them their students.

To address this problem, it will be important to correlate different professional development programs with their impact on students. It is equally important to study the relationship between what occurs in the professional development programs and the impact on teachers and how they change their practices.
**Purpose**

The purpose of this study is to investigate the effects of professional development for science teachers on student learning. It is usually expected that professional development programs positively impact student learning, however, this dimension is not
commonly incorporated into the programs’ evaluation. It is simply assumed that students will be indirectly impacted through their participating teachers. This study hopes to inform present and future professional developers, of what works best in order to maximize their impact on student learning.

Research questions

The formulated research questions focus on the theoretical framework presented in this study. The research questions are:

1. Are professional development programs effective in enhancing student learning in science?

2. What are the characteristics of the most and least effective programs?

Assumptions

The major assumption of this study is the relationship between student achievement and student learning. For many years the issue of what matters most for student learning has been debated. The difficulty starts with the definition of what can be considered student learning, and how one can measure this outcome. Many tests such as the Third International Mathematics and Science Study (TIMMS), the National Assessment of Educational Progress (NAEP), the Scholastic Aptitude Test (SAT), the Graduate Record Examinations (GRE), the Texas Assessment of Academic Skills (TAAS), and the Texas Assessment Knowledge and Skills (TAKS) have developed strategies to assess student learning. This researcher assumes that student achievement as
determined by success on standardized tests is, correlated with the quality of student learning. Frequently these standardized texts are used to make “high-stakes” decisions such as passing, failing, graduating, or entering college.

It is also assumed that teachers change their practices as a consequence of their participation in professional development programs. Teachers enrolled in professional development programs are expected to reform their practices as a consequence of their participation in a more significant way than those who do not have the opportunity to participate.

Finally, it is assumed that student learning is impacted by teacher change. It is common to accept the determinant role of teachers in student leaning. It is assumed that teachers who participate in the professional development programs have different practices from “non-participating” teachers and that this impacts their respective students differently.

**Importance of the Study**

The main goal of this study is to focus on improving student learning in science, by investigating the impact of professional development programs for science teachers on student leaning to provide a new focus for these programs. Professional development programs are often general and loosely focused instead of aiming to improve student learning by reforming and enhancing classroom teaching.

Furthermore, this study addresses an identified “lacuna” (“hole” in Latin) in research. In the past, few studies evaluated the impact of professional development on
student learning, and none attempted to synthesize their results in order create an overall understanding of their impact. Ultimately, it is the goal of this study to inform present and future professional development programs concerning these practices that work best to maximize student learning.

**Definitions**

*Student learning*

For the purpose of this study student learning is defined as a function of student achievement. There are several other indicators and facets of student learning. Ideally, other indicators such as student attitudes towards science, problem solving skills and ability to employ critical thinking in science would also be included. However, most studies dealing with professional development do not attempt to evaluate their impact on such characteristics. There are a few very complete studies that assess several of the mentioned desired outcomes of professional development (e.g., Yager & Weld, 1999). Unfortunately, these are exceptions to the general trend. Most studies do not evaluate the impact of professional development on students at all. Therefore, this study limits the evaluation of student learning only as a function of student achievement, the aspect most commonly evaluated by studies that attempt to assess the impact of professional development on students.
Professional development

Professional development has experienced a metamorphosis in the last 30 years. From in-service, to staff training, to staff development, to workshop, to professional development, the changes have been sometimes in name only, and other times so profound as to make the programs unrecognizable to their precedents. Even today, the offerings of professional development programs are tremendously varied. From program characteristics to total duration one can find almost every type of imaginable program. In order to investigate the overall effectiveness of professional development and to differentiate between the impact of different variables in these programs we have decided to define it as broadly as possible. The Darling-Hammond and McLaughlin (1995) definition has been adopted, considering professional development for teachers as any program aimed at “deepening teachers’ understanding of the processes of teaching and learning and of the students they teach” (p.597)

Program characteristics

In order to answer the second research question it is necessary to differentiate between the different types of characteristics that are common and characterize most professional development programs for science teachers. Therefore four different types of variables characteristic of professional development programs were defined; first, variables dealing with the type of treatment being used to deliver the professional development process (e.g., workshop or study group); second, variables dealing with the content that the programs focus on (e.g., physical science or biology); third, additional
content variables embedded in the program (e.g., pedagogical content knowledge or assessment); and finally, variables discriminating the program duration, both in terms of total number of hours and spread throughout the year.

Limitations

The major limitation of this study is the number of professional development programs included. It is not common practice to include a student learning component in the evaluations of professional development programs. Therefore, many studies about professional development could not be included in this investigation.

Another limitation is the relationship between student learning and student achievement. There is often a lot more involved in student learning than what can be evaluated by their achievement in any type of test or activity. However, it was decided to use achievement as an indicator of student learning.

Finally, generalizability of the conclusions attained with this study is a limitation. Every professional development program is molded by the geographical and social environment of the society that it is targeting. Therefore, even though general recommendations for a professional development program are presented, the local setting must be taken into account to finalize any specific program design. Moreover, despite the main focus on student learning it is recognized that student learning may not be the main goal of some professional development programs. Other alternative focuses, such as teacher retention or administrative proficiency, are equally legitimate, depending upon
the local context situation. That must always be the determinant factor behind the type of program developed.

**Organization**

Chapter 1 provides a general introduction to this study. It introduces the rationale, purpose, and the research questions. Chapter 2 presents a review of the literature dealing with professional development. Chapter 3 discusses the methodology used, and its implementation. Chapter 4 presents the data analyses. Finally, Chapter 5 discusses the implications of the results, suggests further areas of research, and presents the findings in the form of vignettes developed from the data analyses.
Chapter 2

Review of the Literature

Introduction

The major rationale for this study is one supported by Darling-Hammond (1999), presented in chapter one (see Table 1.1). Teacher quality is one of the most determinant factors for student learning. This view is also shared by the National Research Council (2001) in their report “Educating Teachers of Science, Mathematics, and Technology: New Practices for the New Millennium” when they say “As noted in the extensive body of evidence cited throughout this report, research is confirming that good teaching does matter” (p. 4). Loucks-Horsley and Matsumoto (1999) noted that “teacher expertise is one of the most important factors in student learning” (p. 259). This position alone could be sufficient to convince every one who cares about education to support professional development for teachers. Again, this position stresses that the main goal of professional development should be the improvement of student learning.

Without overlooking the importance of pre-service teacher education, no matter how good it is, what happens when teachers are actually working in the field during their induction and subsequent years seems to be highly determinative of their continuous growth and development as a teacher. Within this context, any study concerned with improving teacher quality, must pay special attention to the kind of professional development provided to teachers.
For this researcher this is as crucial as the arguments between those in favor of the professionalization of the teaching career and those in favor of its deregulation (no need for certification based on educational background). This crucial battle has received front page credits in major journals, and has even influenced the U.S. Secretary of Education Annual Report on Teacher Quality when he sides with those pro deregulation citing the 2001 report from the Abell Foundation (Walsh, 2001). However, as Darling-Hammond (2002) points out in her rebuttal, there are outrageous flaws in that report. It is biased and its methods are not clear. Despite that, Walsh (2002) answers in the same tone criticizing Darling-Hammond methods and conclusions. The necessity for more research is evident and imperative. Especially, research directed at the linkage between Research-Policy-Practice: the theme of the 2004 National Association of Research in Science Teaching (NARST, 2004). One can even argue that this is even more crucial in the field of teacher education. The consequence of failing to prove the relationship between teacher education in general, and professional development in particular, with increased student learning is cornerstone to the argument of those in favor of professionalization. Walsh’s (2001, 2002) recommendations are grounded in data analysis that shows no relationship between teacher certification and student learning. Moreover, we must be proficient in the methodologies, such as meta-analysis, that have been used by those in favor of deregulation, and are often not mastered by those against it.
Theoretical Framework

Guskey (1986, 1997) presented one of the few models (see Figure 1.1) of teacher change during their participation in professional development. This model has the professional development program as the triggering mechanism in the change process. However, it recognizes that the student learning outcomes, as observed by the teachers after they have transformed their classroom practices, promote change in teachers’ beliefs and attitudes.

Loucks-Horsley et al. (1998) presented a framework (see Figure 1.2) for designing professional development programs for teachers of science and mathematics. This framework also emphasizes the continuous and circular design permeating the implementation of professional development programs. This design is infused by the continuous reflection on the outcomes of the program to reevaluate and constantly improve it.

Based on the presented models and frameworks for the structure of professional development and its impact on teacher change, and the firm belief that student learning should be the focal point of the educational system, a theoretical framework for professional development was developed (see Figure 1.3).

This framework reflects this researcher’s values about what should guide professional development. Three concentric domains characterize the framework that includes the student learning domain, a teachers’ change/evolution domain, and a professional development domain. Exploration of the literature describing each of these domains follows.
The Professional Development Domain

Encompassing most of the teacher and student domain is the professional development domain with four main stages adapted from Loucks-Horsley et al. (1998). The initial stage is the Goal Setting Phase that starts with the formulation of goals in terms of student learning outcomes, even though there may also be other parallel goals such as building teacher leadership and improving teacher retention. The second stage is the Plan Phase, where the professional development practitioners develop an action plan. This plan is implemented during the Doing Phase when the professional development program actually interacts with the teachers. The Reflection Phase is a vital part of the professional development program because it is based on the actual results of the implemented professional development program on student learning and teacher’s beliefs. Reviewing is always imposed in order to accommodate necessary changes in the next iteration of the professional development program.

The Evolution of professional development

Professional development has evolved during the last 40 years. Traditionally, professional development programs have had as their objective what used to be called the in-service education of K-12 teachers. This notion has contributed to the quasi-professional status of teaching (Darling-Hammond, 1993). This status fails to promote the importance of a profession that has a tremendous impact on students. Teaching and teachers must be recognized and valued for the tremendous role they play in the education of our children. Moreover, it is essential that the status of the teaching career
undergoes profound changes, in order to establish itself as a much more prestigious profession. This change in status of the teaching profession is essential to attract motivated individuals to the profession and to increase their retention in the profession. Professional development has the potential to improve the status of the profession. Additionally, it should help to make obvious the need to treat teachers as established professionals that deserve to be paid as any other highly skilled professional such as lawyers or diplomats – after all, a good teacher sometimes has to assume roles of both of the above!

Frequently the education teachers receive during their pre-service education is not enough to prepare them for their life-time career. In this context, many labels have been given to what Darling-Hammond and McLaughlin (1995) call “deepening teacher’s understanding of the processes of teaching and learning and of the students they teach” (p.597). From in-service, to staff training, to staff development, to workshop, to professional development it has been a long and well fought battle. Professional development can occur at many different levels and formats (informal, nonformal, and formal) (Schwartz & Brian, 1998).

Professional development is the result of the “sociological/Darwinian” evolution of what used to be called “in-service teacher education”. Confronted with the shortcomings of traditional in-service teacher education programs, the concept evolved into a long term, career long process called professional development. The following quote reflect this new thinking about professional development.
Professional development is a lifelong collaborative learning process that nourishes the growth of educators both as individuals and as team members to improve their skills and abilities. … the focus of professional development must be to improve student learning.

(Speck, 2001, p.4)

Literature in this area is abundant. Many suggest what they consider “best practices” for professional development (Darling-Hammond & Ball, 1998; Loucks-Horsley et al, 1998; O’Brien, 1992; Sparks & Loucks-Horsley, 1989). Others point out the ineffectiveness of short, one time programs (Gall, Borg & Gall, 1996). However, there is minimal research investigating how these characteristics are related to better instruction and increased student learning. Most professional development programs are implemented as a response to a perceived educational problem or deficiency (Hyde, 1992). The professionalization movement mentioned above is an additional attempt to transform teacher training from a reactive process, focused on remediation, to a proactive one focused on education.

**Shortcomings of traditional professional development programs**

Many teachers identify traditional professional development with short, after-school initiatives commonly held at the school library or cafeteria. These are commonly one-time initiatives in which the teachers’ participation is often mandated and organized by the school administration. The topics are often too broad for any given teacher and
disconnected from application in their own classrooms. The individual teacher has little or no participation in the decision-making process of what should be explored, leaving the teacher disconnected from the learning experience (O’Brien, 1992). In his review of “effective” guidelines for in-service workshops, O’Brien (1992) calls our attention to the lack of input that the teachers usually have in the experiences planned for them.

Moreover, the strategies, such as lecture, used are often examples of poor and ineffective methodology. Workshops where the teachers are lectured to are common and the lack of participation from them follows as a natural consequence (Radford, 1999). The outdated notion that learning comes in small factual pieces is passed along through boring, impractical means. Radford (1999) reports about the impact of the project LIFE professional development program, a three-week summer workshop, on over 90 teachers and 2100 students. Even though his study does not control for initial differences between the treatment and control groups, triangulation from several data sources makes his findings more robust. He reports statistically significant gains on teachers and students process skills and attitudes towards science as a result of the LIFE program.

Even though one recognizes the important role of teachers in the success of the educational enterprise, there is clearly a gap between the reform goals and the practices in place to translate them into reality. (Lynch, 1997). Teachers are not given the chance to experience the new reform approaches themselves. They are only lectured to about what to do. It is a classical case of “do what I say not what I do”. Lynch’s (1997) study explores how 25 science teachers interact with reform-based curricula such as the National Science Education Standards (NRC, 1996) and Project 2061 (AAAS, 1989). She
points out that “each teacher must construct for himself or herself a notion of what the particular reform means… since before a teacher can be empowered to implement a reform effort, it must first be understood well and appear to be implementable” (p.14). What then are the shortcomings of traditional in-service programs? According to Barufaldi (1997, adapted) they include:

- Are fragmented, short-term, and lack follow up activities
- Lack message and relevance to what actually happens in the teachers’ classrooms
- Are of insufficient intensity and duration to make a positive impact on teacher’s performance and students’ achievement
- Lack incentives and do not respond to teachers’ needs and concerns
- Lack built-in release time for teachers to plan new strategies and to interact with other teachers and professional development experts
- Are driven by mandated requirements rather than by student learning’s
- Are too theoretical and lack practical classroom applications and developmentally appropriate instruction
- Are not aligned with recent developments in curriculum, assessment, and methodology nor with technological advances
- Lack science content rigor and are not thought by credible professionals in science content and methodology
• Do not make effective use of business and industry expertise and resources
• Lack collaboration between colleges of natural science and colleges of education
• Are not coordinated with pre-service teacher preparation programs

These shortcomings are found in a wide range of different settings all over the world, even though with slight changes. The lack of time and incentives to attend professional development activities, for example, changes a lot in different countries (Hayes, 1997 & Hopkins, 1986). All these shortcomings could be easily identified if the teachers participating in traditional in-service teacher education programs were invited to evaluate them.

Keys for successful professional development (from the literature)

However, all is not lost in the “kingdom of teaching”. Several initiatives are promoting a profound shift in teacher education towards more reform oriented practices. Despite the fact that professional development programs differ greatly in their context and specificities, there are several key characteristics that have been identified as crucial to improve their success. Among them are elements unique to adult learning, attention to the change process in which teachers are engaged, duration of the program, opportunities for modeling exemplary practices, and a collaborative structure.

Loucks-Horsley et al (1998) present seven principles for effective professional development experiences (see Table 2.1)
Table 2.1: Principles of effective professional development (adapted from Loucks-Horsley et al, 1998)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Well defined image of effective classroom learning and teaching</td>
</tr>
<tr>
<td>2</td>
<td>Provide opportunities for teachers to build their knowledge and skills</td>
</tr>
<tr>
<td>3</td>
<td>Use or model with teachers the strategies they will use with their students</td>
</tr>
<tr>
<td>4</td>
<td>Building a learning community</td>
</tr>
<tr>
<td>5</td>
<td>Support teachers to serve in leadership roles</td>
</tr>
<tr>
<td>6</td>
<td>Provide links to other parts of the education system</td>
</tr>
<tr>
<td>7</td>
<td>Continuously assessing themselves and making improvements</td>
</tr>
</tbody>
</table>

Providing science teachers with the pedagogical content knowledge (Shulman, 1986), that is appropriate to their area and represents the most recent reform based initiatives, is also one of the values that should guide professional development programs. Providing teachers opportunities to experience reform-based curriculum will enhance the degree of reform-based activity implementation in their respective classrooms.

The duration of the programs has also been referenced as contributing to its success (Lawrenz, 1984; Lynch, 1997). The longer the duration of the program the greater the chance that the teachers are engaged in learning and change. Time is necessary for teachers to reflect upon what they are learning, and to process and apply it in their own classrooms. Time is also necessary for them to share their experiences with their colleagues. Lawrenz’s (1984) study separated teachers into two groups representing two different lengths of professional development sessions: one group with 140 participants in a shorter, one-credit program, and another group with 296 teachers in a longer three-credit program. The researcher implemented a very well designed
methodology, with three different instruments, reporting the reliability of each one of them. A pre-test, post-test design was implemented with all three instruments. A statistically significant difference was found between the impact of the two different duration programs on all levels tested. Attitudes towards science, the value of curricular change, the willingness to participate, and the belief in teaching specific science concepts all increased for the longer duration group and decreased for the shorter duration group.

One other characteristic often identified as important for professional development programs is the necessity of tailoring the program to the teachers needs (Bethel, 1982; Fletcher, Bethel & Barufaldi, 2000; Loucks-Horsley et al., 1998). The importance of taking into account the needs of the teachers is also present in the work of Bethel (1982). His study, involving 254 randomly chosen elementary teachers, was tailored to meet their specific needs and reported increased test scores assessing the teachers’ knowledge of science and, increased amount of time teaching science per week from 8 to 100 minutes.

**The Teachers’ Domain**

On the proposed framework, the student’s learning objective is the center and essentially the goal of the entire program. The teacher’s level is immediately above it, including teachers’ classroom practices, behaviors, and beliefs that have the greatest impact on student learning but are also influenced by feedback received from the students’ experiences and outcomes.
Change is a complicated process. It is not reasonable to expect teachers to change overnight because of their participation in a professional development program. The impact on teachers’ beliefs and attitudes is much more probable to become a reality after they notice an improvement in their students’ learning outcomes than before. However, that is often not enough; intensive and extensive follow up activities of the professional development program are another essential requirement to nurture teacher change.

The teachers’ domain follows the model presented by Guskey (1986). Teacher’s beliefs are included anchored in the theory of teacher change (Guskey, 1986; Horsley, & Loucks-Horsley, 1998). Teachers may begin changing their practice as a consequence of their participation in a professional development program. However, the deeper and harder to achieve change in beliefs and attitudes is much more likely to occur only later after they see and experience the impact of the new practices that they have implemented on the learning of their students.

One essential point in this line of reasoning is that change or evolution of teachers’ practices and beliefs is not a simple process. Moreover, fostering change is never easy and requires huge investments of time and hard work. Considering the Concerns-Based Adoption Model (CBAM) (Horsley, & Loucks-Horsley, 1998; Loucks-Horsley, & Roody, 1990) described in Tables 2.2 and 2.3, one realizes that moving teachers to higher stages of concern and levels of use is indeed very challenging.
Table 2.2: The Concerns-Based Adoption Model (CBAM) (adapted from Loucks-Horsley, 1990; Horsley and Loucks-Horsley, 1998)

<table>
<thead>
<tr>
<th>Stages of Concern</th>
<th>Expressions of concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Refocusing</td>
<td>I have some ideas about something that would work even better.</td>
</tr>
<tr>
<td>5 Collaboration</td>
<td>I am concerned about relating what I am doing with what others are doing?</td>
</tr>
<tr>
<td>4 Consequence</td>
<td>How is my use affecting kids?</td>
</tr>
<tr>
<td>3 Management</td>
<td>I seem to be spending all my time in getting materials ready.</td>
</tr>
<tr>
<td>2 Personal</td>
<td>How will using it affect me?</td>
</tr>
<tr>
<td>1 Informational</td>
<td>I would like to know more about.</td>
</tr>
<tr>
<td>0 Awareness</td>
<td>I am not concerned about it.</td>
</tr>
</tbody>
</table>

Table 2.3: Levels of Use (Horsley & Loucks-Horsley, 1998)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Renewal</td>
<td>Seeks more effective alternatives to the established use of innovation,</td>
</tr>
<tr>
<td>5 Integration</td>
<td>Makes deliberate efforts to coordinate with others in using the innovation.</td>
</tr>
<tr>
<td>4B Refinement</td>
<td>Assesses impact and makes changes to increase it.</td>
</tr>
<tr>
<td>4A Routine</td>
<td>Has established a pattern of use and is making few, if any, changes.</td>
</tr>
<tr>
<td>3 Mechanical</td>
<td>Is poorly coordinated, making changes to better organize use of the innovation.</td>
</tr>
<tr>
<td>2 Preparation</td>
<td>Prepares to use the innovation</td>
</tr>
<tr>
<td>1 Orientation</td>
<td>Seeks information about the innovation</td>
</tr>
<tr>
<td>0 Nonuse</td>
<td>Takes no action with respect to the innovation</td>
</tr>
</tbody>
</table>

This researcher, reflecting on his thinking about the goals and models of professional development recognized that only three years ago, he was only at the Stage 1
of the levels of use. He thought professional development was a very important aspect of the educational system and sought all the information that he could obtain about more innovative professional development projects. Even now, he believes that he still hasn’t achieved the highest level of use; he view himself more at level 5, since he tends to coordinate his efforts with others about innovations. Teacher change is often an essential step to promote student learning, but it is definitely a lengthy and complicated process, making the case for teacher professional development and extensive follow up activities.

It must be remembered that the “teachers are learners and the principles of learning and transfer for student learners apply to teachers” (NRC, 1999; pg 231). If we want teachers to develop their knowledge to better teach children we must remember that they will need valuable learning experiences themselves as learners in order to be able to successfully improve their knowledge. Some of the important themes related to how teachers learn include the following (Loucks-Horsley & Matsumoto, 1999, adapted):

- To gain meaning and deep understanding, learners must build coherent structures of information organized around core concepts or big ideas of a discipline, rather than collect facts and principles through memorization. Thus, teachers need a sound foundation in the major ideas of the disciplines they teach and a deep understanding of how students come to learn those disciplines.
- Studies of expert performance illustrate what successful learning looks like. Experts use problem solving techniques unique to their disciplines to access relevant pieces of their store of information. Thus teachers need to be skilled in
how to make decisions about what students know, what they need to know, and how they can be helped to gain that knowledge – and the knowledge to help their students to do so.

• Learners need to understand major concepts and generalized principles, plus when and how to apply what they have learned. Thus teachers need to know what knowledge to apply in what learning and teaching situations.

• The opportunities and the tools for self-assessment and the disposition to act on information they gather enhance teachers’ learning.

• Learning is influenced by participation in a community, by its norms, its constraints and resources, and its limits and possibilities. Thus teacher learning is enhanced by interactions that encourage them to articulate their views, challenge those of others, and come to better understandings as a community.

All these points are supported by research on learning to teach to new standards summarized by Darling-Hammond and Ball (1998, adapted):

• Teachers’ prior beliefs and experiences affect what they learn.

• Learning to teach to the new standards takes time and is not easy. Most teachers, even if their beliefs are consonant with the new reforms, must develop new ways of teaching and assessing their work.

• Content knowledge is key to learning how to teach subject matter so that students understand it.
• Knowledge of children, their ideas, and their ways of thinking is crucial to teaching for understanding.

• Opportunities for analysis and reflection are central to learning to teach.

The importance of the success of this change process that teachers undergo is essential to the achievement of our main goal – improve student learning. Following is an adapted list of identified facts on the research on change (Loucks-Horsley & Roody, 1990):

• It takes 3 to 5 years to implement meaningful changes in schools.

• Teachers and administrators need concrete, specific models of innovations rather than philosophical arguments.

• Effective models can be replicated (adapted) across districts and schools.

• Change programs need to provide a wide range of support and help to teachers and administrators, including information, training, moral support, problem solving, materials and equipment, trouble shooting, and protection from competing demands.

• Those advocating change need to move away from mandates and rhetoric as their primary interventions. The mandates work well to create the expectation for change and get the attention of those concerned, but they must be followed by attention to the human and organizational needs for input and implementation decisions and for a wide variety of assistance and support.
The process of change described by the researcher is well illustrated in Table 2.4 adapted from the work of Loucks-Horsley (1995):

Table 2.4: A Paradigm for Professional Development in Learner Centered Schools (adapted from Loucks-Horsley, 1995)

<table>
<thead>
<tr>
<th>From Too Much</th>
<th>To More</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on teachers needs</td>
<td>Focus on students learning outcomes</td>
</tr>
<tr>
<td>Focus on individual development</td>
<td>Focus on individual and system development</td>
</tr>
<tr>
<td>Transmission of knowledge, skills, and strategies</td>
<td>Inquiry into teaching and learning</td>
</tr>
<tr>
<td>“Pull-out” training</td>
<td>Job-embedded learning</td>
</tr>
<tr>
<td>Generic teaching skills</td>
<td>Content and content-specific teaching skills</td>
</tr>
<tr>
<td>Fragmented, piecemeal, one-shot experiences</td>
<td>Driven by clear, coherent, long-term strategic plan</td>
</tr>
<tr>
<td>District direction and decision making</td>
<td>School direction and decision making</td>
</tr>
<tr>
<td>Professional development as some people’s jobs</td>
<td>Professional development as everyone’s job</td>
</tr>
<tr>
<td>Professional development for teachers</td>
<td>Professional development for everyone</td>
</tr>
<tr>
<td>Professional development as a “frill”</td>
<td>Professional development as essential</td>
</tr>
</tbody>
</table>

In order to facilitate the implementation of all these changes, several aspects should be considered. Sparks (1989) has identified five necessary components:

- Schools possessing norms that support collegiality and experimentation
- District and building administrators who work with staff to clarify goals and expectations, and actively commit to and support teachers’ efforts to change their practice
- Efforts that are strongly focused on changes in curricular, instructional, and classroom management practices with improved student learning as the goal.
Adequate, appropriate staff development experiences with follow-up assistance that continues long enough for new behaviors to be incorporated into ongoing practice.

Collaboration is also regarded as a very powerful tool to foster teacher change through professional development (Barufaldi & Reinhartz, 2001). In their description of the dynamics of collaborations in a state-wide professional development program for science teachers across Texas, Barufaldi and Reinhartz (2001) explore the impact and the nature of collaboration in 20 sites across the state. Their findings are summarized in Table 2.5.

Table 2.5: Collaboration (Barufaldi and Reinhartz, 2001 pp. 102,103)

| 1.  | Collaboration, working together, rather than alone, can produce beneficial results |
| 2.  | Collaboration is a learned process |
| 3.  | Incentives and rewards are imperative to sustain meaningful collaboration |
| 4.  | The degree or intensity of collaboration within a system is a multi-tiered process |
| 5.  | The interconnectivity of the system must be well articulated |
| 6.  | Support and encouragement are necessary to create linkages among the partners, and to help them realize that they can all learn from others |

**The Students’ Domain**

At the center of the framework developed for this study is the student learning domain, the ultimate goal of the professional development framework. Leading to the goal of student learning is a change or improvement in teachers’ classroom practices, essential to the improvement of student learning.
Inquiry and student achievement have received the least systematic attention in the professional development literature (Guskey, 1997; Guskey & Sparks, 1996; Saab, Steel, & Shive, 1997; Teitel, 1996; Valli, Cooper, & Franks, 1997). There is clearly a need to identify and conceptualize professional development programs’ interactions with student learning. Traditionally, professional development programs have had as their objective what used to be called the in-service education of K-12 teachers. However, learning, not teaching is the issue and ultimate objective. Teaching is one of our methods of attempting to promote learning within our children.

It is timely to conceptually link professional development and teacher preparation with the impact on students’ learning. We have to promote “research as a weapon” (Cochran-Smith, 2002) that makes this connection and creates the framework that will support and validate professional development as one of the ways of promoting student learning.

The importance of linking professional development to student learning is also found in the National Staff Development Council (2001) document “Standards for Staff Development”. The National Staff Development Council advocates that “staff development that improves the learning of all students” (p. 1) is essential, as is the usage of “disaggregated student data to determine adult learning priorities, monitor progress, and help sustain continuous improvement” (p.1) the rational behind a data-driven process for professional development. The fact that teacher education matters the most to student learning (Darling-Hammond, 1999; Guskey, 1998; Killion, 1998) has been defended by several authors in the last decade, others take issue with this notion (Walsh, 2001, 2002).
Restructuring our schools toward a more learner-centered environment has resulted in a new conception of teaching and consequently of teacher education (Darling-Hammond, 1992). This is not a new idea, we can read in the work of Bethel (1982) “The real measure of a staff development program, of course, is its effect on students” (p. 416). But the link of this learner-centered community with professional development has not yet been established.

Professional development is central to educational reform (Loucks-Horsley, & Matsumoto, 1999; National Education Goals Panel, 1998) because the teachers are responsible for our children’s education. Other factors, such as class size, school environment, and socio-economic status, have also been presented as having a significant impact on student learning. However, the teacher is the most important factor in providing a rich learning atmosphere for students (Darling-Hammond, 1999).

In this context, the ultimate goal of professional development is improving student achievement (Guskey, 1997, 1998; Guskey & Sparks, 1996; Killion, 1998; Speck, 2001). Loucks-Horsley (1999) believes that “teacher expertise is one of the most important factors in student learning” and professional development is probably the best way of improving teacher expertise.

There is limited research in this area. Moreover, most “research on or evaluation of professional development does not assess student learning” (Loucks-Horsley, & Matsumoto, 1999, p. 258). Killion (1998) reviewed 450 projects of the Middle Grades Initiative of the National Staff Development Council and found that more than 90% did not include any measure of professional development impact on student performance. It
is imperative that researchers consider the impact of professional development programs on student learning as stated in the National Science Education Program Standards (NRC, 1996):

In an effective science program, a set of clear goals and expectations for students must be used to guide the design, implementation, and assessment of all elements of the science program.

The most important resource is professional teachers.

(Standards A and D, National Research Council, 1996)

Some professional development programs have already addressed this crucial issue (Adelman, 1998; Cohen-Regev & Strobel, 2001; Fletcher & Barufaldi, 2002; Goertz, Massel & Corcoran, 1998; Killion, 1998; Lewis, 1998; Marek & Methven, 1991; Raghavan, Rubin & Norman, 1992; Parke & Coble, 1997; Zucker & Marder, 1998). Priorities are changing. Evaluation that was described as costly, time-consuming, and that diverted attention from the more important planning, implementation and follow-up activities is now being regarded as vital to assess the impact of the professional development program on students. Moreover, the current climate of accountability requires students to achieve higher standards, teachers to be held accountable for the students’ results, and professional developers to show that they really matter (Bush, 2002; Guskey, 1998).
Guskey (1998) suggested five different levels of professional development evaluation. The levels are hierarchically arranged, from simple to more complex, with success at a lower level being necessary to achieve success at the levels that follow. The suggested levels are:

1- Participants’ reactions
2- Participants’ learning
3- Organizational support and change
4- Participant’s use of new knowledge and skills
5- Student learning outcomes

Naturally, the bottom line in education is addressed in Level 5: The impact on students (Guskey, 1998). The factors responsible for the impact on student learning must be identified.

One danger of which we must be aware is excessive focus on testing and a lack of ability to distinguish between testing and assessing. The consequences of excessive testing with high accountability for students, teachers, schools, and professional development practitioners are treacherous because they may lead the effort away from the promotion of student learning towards students achievement in standardized tests and lead into practicing routines that will improve the students rate of success in that kind of test and lead further away from enriching inquiry embedded learning activities. We cannot have a system that teaches and values creativity and then measures for conformity; they simply are not compatible.
Professional Development Worldwide

Professional development initiatives are a concern across the world. Teacher expertise is recognized in most countries as the most preponderant way to improve student learning. Because of this, studies from several countries are included in this investigation. In our initial review of literature we were able to locate articles about professional development in science from countries such as the United Kingdom, Israel, Portugal, Spain, Australia, Canada, Japan, Korea, Taiwan, The Netherlands, Germany, South Africa, and the United States. The improvement of professional development is unquestionably a worldwide effort. While findings of a study performed in Portugal may not be directly applicable to Korea, one will certainly profit from this sharing of research. Even programs from different countries will share characteristics common to science education reform. Disregarding research done worldwide would be like trying to drive with only one eye open. We need all the help we can get, especially in such a relevant and high-stakes arena as education. Moreover as Scott, Stone and Dinham (2001) point out “teachers everywhere enter the profession to serve children” (p.13). What is more, they all struggle to improve their practice and teaching through professional development. However, due to the established eligibility criteria (see Chapter 3) not all of these were included in our final sample. Only studies from the United States, Canada, Portugal and Korea ended up being used.

Naturally, professional development across such different countries has different emphasis. This is due not only to the variety of curricula being implemented, but also to the range of pre-service education and certification programs that teachers must go
through. Local conditions are different and lead to distinct emphasis for professional development programs. For example, out-of-field teaching and shortage of qualified teachers with majors in the subject that they are teaching, are two very pertinent problems in the USA (Darling-Hammond, 1998) that are practically nonexistent in Portugal and Germany (Valente, 2004). Some of the characteristics of professional development programs in some of the mentioned countries will be addressed.

However, one must be careful not to generalize from the findings reported in professional development programs in these countries. They usually represent only one particular program in place in that country. Frequently that program is not experienced by the majority of teachers. These programs often reflect reform based initiatives and usually find their way into a published format, though, their impact is limited.

In the some of the European Union countries such as, Portugal, Spain, United Kingdom and Germany; different institutions cooperate to ensure distribution of competences (Usarralde, 2001; Valente, 2003). However intense collaboration was not really apparent. Professional development in these countries is usually associated with a teaching career. The teachers’ participation is mandatory, but they can enroll in a wide variety of programs though often not relevant to their own area of expertise (Ponte, 1993) In Portugal, for example, these programs can be as short as 10 hours or as long as 120.

In Portugal, professional development is regulated by law. Teacher certification is awarded only by universities whose degrees have been recognized by the government. Every teacher must complete a one year internship to become fully certified. Once a teacher has completed the internship he formally enters the teaching career. However,
progression in the career requires a certain number of professional development hours per year, (usually about 20) depending on placement of teachers on the career ladder. This placement will allow them advance in the career ladder and achieve higher salaries. However, the variety of programs available for the teachers is enormous. It is possible, for example, to have a physics teacher enrolled in a professional development program dealing with traditional folklore dancing! In the past 5 years, however, there has been a significant move toward master degree programs as the choice for professional development for science teachers. This is mainly due to the fact that the degree program provides the teachers with an automatic two years advance on their career ladder.

Shimahara (1998) provides an overview of the professional development model in Japan entitled “Teaching as a Craft” (p.451). In this article he discusses several types of in-service education available for Japanese teachers. He mentions “in-house inservice education, which is offered by virtually every school” (p. 455); out of school programs organized by subject areas associations called “bukai” (p. 456); teacher initiated networks; and a one year mandatory program enrolling all beginning public school teachers.

In Korea, similarly to Japan, the school system is centralized and coordinated through the Ministry of Education (Lee, 2001). As Lee (2001) reports, it is a system where the teachers generally have more authority than American teachers. The teachers participating in the professional development program described by Lee (2001) were selected through a screening process coordinated by the Ministry of Education. This
procedure is described as a very competitive process due to the large number of applicants and their high qualifications.

In Canada there has been a call for substantial curriculum change in teacher education (Bencze & Hodson, 1999). However, this is probably the international context closer to the US reality. According to Ross, Rolheiser and Hogaboam-Gray (1998) work two types of professional development activities occur in Canada. Their descriptions of the implementation of action research and skills training professional development programs are very similar to the ones described by U.S. authors.

In the U.S. the major curriculum restructure of the late 1960’s with the Biological Sciences Curriculum Study, the Physical Science Study Committee, and the Chemical Education Material Study, led to a variety of professional development programs in science. The Nation at Risk (1983) report, during the nineteen-eighties called for a substantial financial investment in teacher education. These programs have been evolving through successive waves of change throughout the eighties and nineties, but the variety is still great. The U.S. Department of Education and the National Research Council have attempted to direct a more consistent professional development strategy during the nineteen-nineties with documents such as the National Science Education Standards (1996). However, most of their recommendations in this area are based on theoretical beliefs as opposed to quantifiable data.

The rewards system also varies greatly. In Massachusetts, for example, teachers are rewarded for their participation in professional development programs with the climbing of their career ladder. In Texas there is no career ladder. A professional
development program like The Texas Regional Collaboratives for Excellence in Science (TRC) recruits its teachers as volunteers. Some of them do gain credit towards Master Degree program through local partnerships with universities, but most of them participate only for the rewards of continuously learning.

A Model Professional Development Program

The TRC headquartered in the Center for Science and Mathematics Education at The University of Texas at Austin, provides professional development for science teachers throughout Texas. The professional development is provided locally by instructional teams composed by four providers including:

- Science specialists
- Science education professors
- Science professors
- Master teachers

Before working with the local teachers, each team member has the opportunity to receive professional development by the TRC. This program was designed around six core principles called “systemic threads”. These principles are presented on Table 2.6.

The impact of this model for developing professional development teams (Loucks-Horsley et al, 1998) is represented in Figure 2.1. As one can note in this figure, the training provided by four instructional teams, can easily impact more than 40,000 students in a period of only five years. The TRC currently has 20 sites in Texas. During the 2001 - 2002 school year, the TRC provided over 700 science teachers, from all grade
levels and diverse socioeconomic schools, with more than 100 contact hours of professional development (TRC, 2004). This program has been in operation for more than ten years. The TRC has been recognized in a variety of ways including induction into the Texas Science Hall of Fame in 2000. The TRC has been commended as an exemplary professional development program that promotes science in Texas schools.

Table 2.6: Core principles of the Texas Regional Collaboratives (Jbeily & Barufaldi, 1998)

- Science Literacy: Curricular and pedagogical decisions should reflect the goal of scientific literacy for all students
- Technology Integration: The use of technology is an integral part of achieving science literacy.
- Standards-Based Instruction: Reform must be systemic in nature and decisions standards-based.
- Equity: Science is for ALL students
- Constructivist Instructional Techniques: Science education is an active process.
- Authentic Assessment Strategies: Assessment should be authentic and aligned with the goal of achieving science literacy

The TRC is funded by a variety of agencies. This collaborative partnership was created by local agencies and several business partners at the State and National levels. “Major partners include: the National Science Foundation, Texas Education Agency, CASIO Inc., Delta Education, ExxonMobil Education Foundation, Frey Scientific, Holt, Rinehart and Winston, Shell Oil Company Foundation, Southwestern Bell Corporation, and Toyota USA Foundation” (Meyer et al, 2004, p.3).

The tenets of the TRC were aligned with Susan Loucks-Horsley work to ensure a quality professional development program. The TRC addressed each of her recommendations as shown in Table 2.7.
Moreover, this program is exemplary because of the amount of research conducted during the last ten years. This research focused on the evaluation of its effectiveness and impact by studying the increase in teacher content and pedagogical knowledge, and student achievement in science.

During the last ten years, evaluations of this program have revealed significant gains in teacher pedagogical and content knowledge (Fletcher, Bethel, & Barufaldi, 2002; TRC, 2004). Moreover, Meyer and Barufaldi (2003) also reported an increase in teachers’ confidence and understanding of the teaching of science. Meyer, Lee, Fletcher, Tinoca, and Barufaldi (2004) conducted a study of 65 teachers participating in the TRC.
She used questionnaires and interviews to determine what attracted the teachers to this program. She reported that the teachers are seeking ways to improve their practice and appreciate the sustained support of a program like the TRC. In addition, these teachers expressed appreciation for the opportunity to network and collaborate with other teachers across the state.

Table 2.7: Comparison of Loucks-Horsley, et al. (1998) Principles to TRC (Meyer et al, 2004)

<table>
<thead>
<tr>
<th>Loucks-Horsley, et. al. (1998)</th>
<th>Texas Regional Collaboratives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Well defined image</td>
<td>1. Six shared systemic threads</td>
</tr>
<tr>
<td>2. Teachers building knowledge and skills</td>
<td>2. Professional Development Academies</td>
</tr>
<tr>
<td>3. Provides models teachers use with students</td>
<td>3. Teachers trained as they would teach their students</td>
</tr>
<tr>
<td>5. Teachers in leadership roles</td>
<td>5. Science Teacher Mentors</td>
</tr>
<tr>
<td>6. Links to other parts of the system</td>
<td>6. Access to up-to-date state and national information as well as experts in science and science education</td>
</tr>
<tr>
<td>7. Continually assessing</td>
<td>7. Pre &amp; Post assessment each year plus Formative Assessments</td>
</tr>
</tbody>
</table>

Lee, Fletcher, Tinoca, Barufaldi and Meyer (2004) also conducted research on the TRC, focusing their study on the impact of mentoring on teachers. They used a quasi-experimental design to evaluate the importance of mentoring on teacher practice, and found a statistically significant impact on both mentors and mentees as a result of their collaboration when compared to a control group. This mentoring support is crucial
especially during the induction years and later, as a rejuvenation factor for more experienced professionals (Ingersoll, 2003; Luft, Roehrig, & Patterson, 2003).

Fletcher, Bethel and Barufaldi, (2000) provides an introduction to the types of instruments used to assess the program. This work is ongoing (Fletcher & Barufaldi, 2002; Fletcher, Bethel & Barufaldi, 2001; Fletcher, Tinoca, Lee, Barufaldi, & Meyer, 2004). The researchers have implemented a variety of techniques, including the Science Classroom Profile (2004), used to collect student data to evaluate the impact of the TRC.

Summary

Chapter 2 presented a wide range of literature dealing with each of the three domains described in the theoretical framework. However, this literature search focused on only each of the domains separately. Moreover, the few studies that permeate more than one domain are, usually of a theoretical nature. The need for further empirical research is evident. In particular to promote the connection between Research-Policy-Practice, the theme of this year’s annual meeting of the National Association for Research in Science Teaching, more empirical studies are needed. The TRC, through evaluation and research, is promoting the interconnectivity among research, policy and practice.
Chapter 3
Methodology

Introduction

In this chapter the methodology used for this study and the steps behind its implementation are presented. The reasons why a meta-analysis was chosen to address the posed research questions are explained. The reader is presented with the basic concepts of meta-analysis and the fixed models, and vote counting cases in particular. The collection and analysis methods are presented, including an explanation of the corrections used. Finally vignettes are used to illustrate the results of this study.

Rationale for the use of meta-analysis

The purpose of this study is to investigate the effects of professional development on science teacher practice and student learning. It is usually expected that professional development programs have a positive impact on student learning; however this dimension is not commonly incorporated in program evaluation. The following are the research questions: Are professional development programs effective in enhancing student learning in science?; and, What are the characteristics of the most and least effective programs? These questions call for an investigation studying the impact of professional development programs for science teachers on student learning in science. This inquiry requires not only the evaluation of the overall impact of professional
development programs on student learning. But also, to compare and differentiate different program characteristics and their impact on students. The methodology was chosen because of its ability to address both formulated questions.

Meta-analysis was selected as the method of choice due to its ability to synthesize results from a large array of studies, differentiating between the impacts of different variables. Moreover, this type of analysis has the capacity to produce quantitative results about the overall impact of professional development on student learning, and to distinguish between program characteristics as desired. This enables the exploration of the overall impact of professional development on student learning by synthesizing the available research. In addition coding the different characteristics of the programs enables the investigation of what makes them more or less effective.

**Description of meta-analysis**

Meta-analysis is a methodology very similar to survey research with the difference that the target population of the surveys is a sample of studies instead of people (Lipsey & Wilson, 2001). This methodology relies on the estimation of effect sizes from different studies to combine them and provides powerful integrated results. Meta-analysis is not a recent methodology. As Olkin (1990) showed, examples of quantitative literature synthesis are found as early as 1904 in the work of Pearson. The term “meta-analysis” however originates from Glass’s (1976) work to refer to “the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings” (p. 3, Glass, 1976 cit in Cooper & Hedges, 1994). He
was one of the first to implement the use of quantitative statistical analysis in the social sciences (Cooper & Hedges, 1994). This method became popular during the 1970’s when it was used to tackle the emerging debate about the effectiveness of psychotherapy by Smith and Glass (1977) since it has the power to synthesize results across a larger number of studies (375 studies were used in this effort). Glass (1976) and Rosenthal (1979) were two of the pioneers to promote its use in social research including education. Since then, “literally thousands of meta-analyses have been conducted and great improvements have been made in meta-analysis methodology” (p. 1, Lipsey & Wilson, 2001).

The key concept behind every meta-analysis is the notion of effect size. Given any set of quantitative data dealing with a certain topic it is expected (with occasional exceptions) that different measurement procedures and instruments have been used. In this study, for the sample of professional development programs for science teachers collected, there are a variety of procedures that have been used to measure impact on student learning. Some studies may use local state standardized tests, others may use adaptations of international tests such as the Third International Mathematics and Science Study (TIMSS), while others still may develop their own measurement tool. Results from such different studies are comparable due to the concept of standardization related to the use of the effect size statistic. Traditionally the effect size statistic has been defined by dividing the mean difference between the control and experimental groups by the pooled within-group standard deviation (Glass, 1976).
Usually, effect sizes range from -3.0 to +3.0. An effect size of 0.9, for example, represents that the subjects in the experimental group scored 0.9 standard deviations above the control group.

Specific procedures used in meta-analysis research include:

1. Literature search and study retrieval
2. Coding
3. Calculation of effect sizes and correction factors
4. Data analysis

There are major benefits in utilizing a meta-analytic approach (Rosenthal, 1991), including completeness and explicitness of the process. Even though we are aware of the “elusive literature” (Cooper & Hedges, 1994) problem, when limited research studies are available, the use of meta-analytic procedures produces an exhaustive review of a research domain, with appropriate weighing of different studies and explicit quantitative relationships. Moreover, it has been shown empirically, that meta-analytic procedures increase power at the same time that they decrease type II errors (Rosenthal, 1991).

**Sampling Procedures**

*Literature search and Study retrieval*

The first step in any meta-analysis enterprise is to gather a group of studies as the sample. This sample will then be surveyed or questioned with a coding form, very similarly to what would happen to one individual answering a survey or a questionnaire.
In order to compile the sample of studies to be used an exhaustive literature search will be performed. This search will be conducted using the following methods:

Review articles – key articles in science education and professional development (e.g.; Kennedy 2001; Shymansky, 1983), as well as prior meta-analysis in overlapping areas (e.g.; Hurley, 2001; Lederman, 1999), are used to identify other retrievable studies.

References in studies – some studies are identified when referenced in some of the studies initially found (e.g.; Shymansky, 1983).

Computerized bibliographic databases – Educational Resources Information Center (ERIC); UMI ProQuest Digital dissertations; EBSCOhost® Electronic Journals Service (EJS), FirstSearch (aka: Wilson Select Plus), WorlCat and SSCI (Social Sciences Citation Index) are also important sources of studies. Some of the search terms to be used include: “professional development for science teachers,” “in-service education for science teachers,” “student learning” and “student achievement.”

*Eligibility criteria*

An initial literature search on professional development for science teachers and student learning in science led the investigator to over 4000 references since 1970. Even though the research literature on professional development is vast, the number of studies employing true or quasi-experimental designs permitting the calculation of an effect size
is limited. Only 182 studies of a quantitative nature were identified. From this, only 37 meet the eligibility criteria. There are three basic criteria for inclusion in this study.

1. The study has to employ a true or quasi-experimental design, of a quantitative nature permitting the calculation of an effect size.

2. The study has to have been conducted since 1970. To locate the investigation in a time frame after the major curriculum revolutions (BSCS, PSSC) of the late 1960’s.

3. The study has to address the consequences of the implementation of a professional development program for science teachers on their students’ learning in science.

These criteria were created to limit the studies included in the sample to those with the characteristics that permitted performance of the meta-analysis in order to answer the posed research questions. Articles from periodicals and the ERIC data base were obtained through the University library or, when available, on-line articles were downloaded from the respective websites. Articles not available from the library or online were obtained through inter-library loans.

**Data collection**

Once the study sample has been assembled the coding form to “question” it was created. The creation of the coding form is one of the most important steps in any meta-analysis because decisions about what to code will ultimately determine the research
questions that can be addressed and answered. In this case, two major objectives were
determined by the two original research questions. To address the question: “Are
professional development programs effective in enhancing student learning in science?”,
the effect size reported by each study was identified in order to synthesize and report the
overall medium effect size. However, to address the question: “What are the
characteristics of the most and least effective programs?”, the different variables
characterizing professional development programs in science education were identified
and coded in order to differentiate between their impact on student learning.

In order to identify the variables four major theoretical frameworks were used.
For the descriptive components, as well as for the effect size computation of the studies
the variables suggested by some meta-analysis specialists such as Cooper and Hedges
(1994), and Lipsey and Wilson (2001) were included. These are clearly defined and
unambiguous variables. They included simple descriptive variables such as grade level,
socio economic status, and ability level (e.g.; GRADLEV; SES; ABILEV, see appendix
A variables 1-7, 14-19, and 21-23). Special relevance to the variables Type of criterion
and Congruence of treatments and outcome measurements (see variables 21 and 22).
These variables were used a decision making criteria aligned with the imposed eligibility
criteria. Only studies with outcomes based on achievement and congruent treatments and
measurements were included in the final sample. The method of measurement variable
(see variable 23 appendix A) distinguishes between studies that developed specific tests
to be used and those that used already developed tests (or adaptations of these) such as
the NAEP, TIMSS or local state tests. At a different level, other variables dealt with the
identification and computation of the observed effect size such as sample sizes, group means, and standard deviations (e.g.; NTREAT, TREATMEAN and TREATSD, see appendix A variables 24-36).

The categories defined by Loucks-Horsley et al. (1998) were used to organize, categorize and define the variable TYPETREAT, distinguishing between the different activities of professional development present in this study. The work of these researchers has been influential in professional development in science education, and explores a variety of different professional development activities. The defined types of professional development are presented in Table 3.1.

This extensive list of types of activities that can be used in professional development programs were interpreted as non-exclusive, allowing any particular program to be categorized in more than one of the defined categories. For example a program in which the teachers worked collaboratively with “real” scientists to develop new curricula would be categorized in both the “Partnerships with scientists” category as well as the “Curriculum development” category.

Table 3.1: Strategies for Professional Learning (adapted from Loucks-Horsley et al., 1998)

<table>
<thead>
<tr>
<th>Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immersion in inquiry into science and mathematics</td>
</tr>
<tr>
<td>Engaging in the kinds of learning teachers are expected to practice with their students—that is, inquiry based science investigations or meaningful mathematics problem solving</td>
</tr>
<tr>
<td>Immersion in the world of scientists and mathematicians</td>
</tr>
<tr>
<td>Participating in an intensive experience in the day-to-day work of a scientist or mathematician, often in a laboratory, industry, or museum, with full engagement in research activities</td>
</tr>
<tr>
<td>Curriculum implementation</td>
</tr>
<tr>
<td>Learning, using, and refining use of a particular set of instructional materials in the classroom</td>
</tr>
<tr>
<td>Curriculum replacement units</td>
</tr>
<tr>
<td>Implementing a unit of instruction that addresses one topic or concept and incorporates effective teaching and learning strategies to accomplish learning goals</td>
</tr>
</tbody>
</table>
Table 3.1: Strategies for Professional Learning (continued)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Curriculum development and adaptation</strong></td>
<td>Creating new instructional materials and strategies or tailoring existing ones to meet the learning needs of students</td>
</tr>
<tr>
<td><strong>Workshops, institutes, courses and seminars</strong></td>
<td>Using structured opportunities outside of the classroom to focus intensely on topics of interest, including science or mathematics content, and learn from others with more expertise</td>
</tr>
<tr>
<td><strong>Action research</strong></td>
<td>Examining teachers’ own teaching and their students’ learning by engaging in a research project in their classroom.</td>
</tr>
<tr>
<td><strong>Case discussions</strong></td>
<td>Examining written narratives or videotapes of classroom teaching and learning and discussing what is happening, the problems, issues, and outcomes that ensue</td>
</tr>
<tr>
<td><strong>Study groups</strong></td>
<td>Engaging in regular, structured, and collaborative interactions regarding topics identified by the group, with opportunities to examine new information, reflect on their practice, or assess and analyze outcome data.</td>
</tr>
<tr>
<td><strong>Examining student work and student thinking and scoring assessments</strong></td>
<td>Carefully examining students’ work and products to understand their thinking and learning strategies and identify their learning needs and appropriate teaching strategies and materials</td>
</tr>
<tr>
<td><strong>Coaching and Mentoring</strong></td>
<td>Working one-on-one with an equally or more experienced teacher to improve teaching and learning through a variety of activities, including classroom observation and feedback, problem solving and troubleshooting, and co-planning</td>
</tr>
<tr>
<td><strong>Partnerships with scientists and mathematicians in business, industry, and universities</strong></td>
<td>Working collaboratively with practicing scientists and mathematicians with the focus on improving teacher content knowledge, instructional materials, access to facilities, and acquiring new information</td>
</tr>
<tr>
<td><strong>Professional networks</strong></td>
<td>Linking in person or through electronic means with other teachers or groups to explore and discuss topics of interest, set and pursue common goals, share information and strategies, and identify and address common problems</td>
</tr>
<tr>
<td><strong>Developing professional developers</strong></td>
<td>Building the skills and knowledge needed to create learning experiences for other educators, including design of appropriate professional development strategies; presenting, demonstrating, and supporting teacher learning and change; and understanding in-depth the content and pedagogy required for effective teaching and learning of students and other educators</td>
</tr>
<tr>
<td><strong>Technology for professional learning</strong></td>
<td>Using various kinds of technology to learn content and pedagogy, including computers, telecommunications, videoconferencing, and CD-ROM and videodisc technology</td>
</tr>
</tbody>
</table>

This is also the way that Loucks-Horsley et al. (1998) introduced them. In their work they suggest that most professional development programs are a combination of two or more different types of activities.
For the science content of the professional development programs the variable THEODIM was created using the eight facets proposed in the National Science Education Standards (NRC, 1996). These categories include:

1. Science as inquiry
2. Physical science
3. Life science
4. Earth and space
5. Science and technology
6. Science and society
7. History and Nature of Science
8. Unifying concepts

These categories were also interpreted as non-exclusive, as most programs focus on more than one of them. This is a broad definition of the characteristics of school science; contrary to more restrictive ones that focus on a particular content such as physics, chemistry, or biology (NRC, 1996).

Additional context variables were created to give a fuller description of the professional development program (see variable 10 appendix A), grounded on characteristics suggested by isolated authors and personal experiences, suggesting possible implications for the success of professional development. These variables include:
Specific content – the program focuses only on one particular content (e.g. electricity) as opposed to a more broad focus

Pedagogy – a strong emphasis in general pedagogical practices not specific to any discipline

PCK – focus is on pedagogical content knowledge (Shulman, 1986). Pedagogical content knowledge is the ability to transform subject matter to a diverse group of students using multiple strategies and methods of instruction and assessment while understanding the contextual, cultural, and social limitations within the learning environment

Technology – technology is used as part of the professional development program even if it is not its main focus

Testing, assessment – focus on developing new assessments and preparing the students to take new mandatory tests

Standards – focus on promoting reformed based practices aligned with the National Science Education Standards

Additional variables include duration of the treatment, housing of the program, and funding per student (e.g.; TOTALDUR; HOUSE; FUND see appendix A variables 11-13 and 20)

Duration of the professional development program has been cited by several authors (e.g.; Desimone, Garet, Birman, Porter, & Kwang, 2003; Lawrenz, 1984) as important for the effectiveness of the treatment. However, no study has been found that
clearly investigated this relationship or suggested the number of hours that are sufficient. As a result, this variable was created and analyzed at the following levels:

**TOTALDUR** – the total number of hours, defined as up to eight hours (corresponding to approximately one day); 9 to 16 hours (two days); 17 to 50 hours (one week); 51 to 100 hours (two or three weeks); and over 100 hours.

**SPREAD** – to reflect the total time frame, in which the program is implemented, defined as up to one week; up to one month; up to six months; and over six months.

Another variable of interest is the housing of the program. Some authors (Shawl, 1984; Wendell, 1977) suggest that housing the program in a university or a school district may have a different impact. Consequently, the variable **HOUSE** was created with possible values of university, school district, or school to reflect the organization behind the program implementation.

An additional variable selected was the presence of an initial needs assessment activity to align the program with the needs of the teachers (Eleser & Chauvin, 1998; Fletcher, Bethel & Barufaldi, 2000). Therefore, the variable **NEEDASSES** was created to distinguish between programs with or without that characteristic.

Finally, to investigate the relationship between the amounts of funding (Lieberman, & Grolnick, 2000) involved in the program and its effectiveness, the variable **FUND** was created. This variable is defined as the amount of funding per teacher.
Coding procedure

The author of this study is the principal coder for all of the selected studies. However, because one of the main flaws of meta-analysis can be coding bias, additional coders will be used to improve reliability. Therefore, two additional coders knowledgeable in the field of science education (i.e., graduate students in science education) will be used. These coders are volunteers and will code three studies each. This will provide a total overlap between coders of 17% of the total sample.

Whenever we were unable to code one of the variables present in the coding form due to lack of information in the study, an attempt will be made to contact the primary author of that study. This will allow completion of most of the coding forms. However, sometimes there may still some estimation involved. This is normally due to the fact of absent data (such as the standard deviation for the experimental group). To reflect the degree of estimation involved, the variable confidence rating will be coded. This variable has five possible values to reflect the degree of estimation involved: 1 highly estimated; 2 moderate estimation; 3 some estimation; 4 slight estimation; and 5 no estimation.

Effect size computation

The traditional Cohen (1977) effect size $d$ is defined by the formula (Equation 1):

$$
d = \frac{\bar{x}_{exp} - \bar{x}_{con}}{\sigma_w}
$$

(1)

where: $d$ – is Cohen’s effect size
\( \bar{x}_{\text{exp}} \) – Average score for the experimental group

\( \bar{x}_{\text{con}} \) – Average score for the control group

\( \sigma_w \) – Estimated pooled within groups standard deviation

This formula is in the same family as Hedges’s \( g \) where only the control group is used to calculate the standard deviation or Glass’s \( \Delta \) where the standard deviation from the true population is used and often makes it impossible to use in real-life meta-analysis (Rosenthal, 1994). An analogous result is the correlation effect size \( r \) (Equation 2). The correlation effect size \( r \), represents the correlation coefficient for the relationship between two variables, \( x \) and \( y \), and is directly the effect size estimate:

\[
 r_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y}
\]  

(2)

where: \( r_{xy} \) – correlation effect size

\( \sigma_{xy}^2 \) – Variance

\( \sigma_x \) – Standard deviation for group \( x \)

\( \sigma_y \) – Standard deviation for group \( y \)

The correlation effect size \( r \) can also be calculated directly from the \( t \) or \( F \) statistics (see Equation 3 for an example):

\[
r = \sqrt{\frac{t^2}{t^2 + df}}
\]

(3)

where: \( df = n_1 + n_2 - 2 \)

\( n_1 \) – number of elements in group 1
These two types of estimates of the effect size, \( d \) and \( r \), can be related by the formula in Equation 4:

\[
r = \frac{d}{\sqrt{d^2 + \frac{1}{pq}}} \tag{4}
\]

Where \( p \) and \( q \) are the proportion of the population in each of the two groups being compared.

Even though both of these effect size estimates can be used as the base for the meta-analysis calculations, Rosenthal (1993) suggested four reasons to prefer \( r \) over \( d \). First, when using \( r \) there is no need to make adjustments in moving from t tests for independent observations to those for correlated observations. With \( d \) that is not the case. Second, investigators often do not report sample sizes making the calculation of \( d \) difficult unless we assume equal sample sizes for both the experimental and control groups. Third, \( r \) can be more simply interpreted due to its relationship with the Binomial Effect Size Display (BESD). Finally, \( r \) offers greater flexibility than \( d \). It can always be used whenever \( d \) can while the reverse is not true. For these reasons \( r \) was chosen over \( d \) in this investigation but \( d \) equivalents can be obtained to the achieved \( r \)’s to facilitate the understanding of the readers more familiar with the more traditional estimate of effect size from Cohen’s work.

By calculating an effect size statistic for each study in the meta-analysis a comparison of their results in the same scale is possible. Furthermore, by combining the
effect size statistic from different studies an overall mean effect size can be calculated, and differentiation between the effect sizes of groups of studies with different characteristics is possible.

Cohen (1977, 1988) in his work on statistical power suggested that standardized mean differences effect sizes should be classified as small if they are less than 0.2 and large if they are more than 0.8 (the analogous values for the correlation effect size are \( r = 0.10 \) and \( r = 0.40 \)). These values, however, are not accepted by all researchers. As Rosenthal (1993) points out, “Mayo (1978) criticized Cohen (1977) for calling effect size large (\( d = 0.8 \)) when it accounted for “only” 14% of the variance” (p.132). Other critics can also be found in the work of Rimland (1979) about Smith and Glass (1977) even smaller effect sizes estimates.

To clarify this point, Rosenthal and Rubin (1982) introduced the concept of BESD. The BESD is an intuitive tool that provides the reader an easier understanding of the results camouflaged behind the usage of the effect size. The BESD addresses the question “What is the effect on the success rate (e.g., survival rate, cure rate, improvement rate, and selection rate) of the institution of a new treatment procedure, a new selection device, or a new predictor variable?” (Rosenthal, 1993).

According to this Table, an effect size of \( r = 0.30 \) for example which “only” corresponds to a variance of 9% can be responsible for a reduction in death rate from 65% to 35%. More generally it will be associated with an increase in success rate from 35% to 65%.
The distinction between small and medium effect sizes within this context is not as clear as having an $r$ of more or less than 0.2. As Rosenthal points out (1991) in the field of medicine, effect sizes as small as $r = 0.034$ have been considered big enough to make decisions about drugs as widespread as aspirin.

Table 3.2: Changes in Success Rates (BESD) corresponding to various values of $r^2$ and $r$ (Adapted from Rosenthal, 1993)

<table>
<thead>
<tr>
<th>Effect Sizes $r^2$</th>
<th>$r$</th>
<th>Equivalent to a Success Rate Increase From</th>
<th>To</th>
<th>Differences in Success Rates (a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.02</td>
<td>0.49</td>
<td>0.51</td>
<td>0.02</td>
</tr>
<tr>
<td>0</td>
<td>0.04</td>
<td>0.48</td>
<td>0.52</td>
<td>0.04</td>
</tr>
<tr>
<td>0</td>
<td>0.06</td>
<td>0.47</td>
<td>0.53</td>
<td>0.06</td>
</tr>
<tr>
<td>0.01</td>
<td>0.08</td>
<td>0.46</td>
<td>0.54</td>
<td>0.08</td>
</tr>
<tr>
<td>0.01</td>
<td>0.1</td>
<td>0.45</td>
<td>0.55</td>
<td>0.1</td>
</tr>
<tr>
<td>0.01</td>
<td>0.12</td>
<td>0.44</td>
<td>0.56</td>
<td>0.12</td>
</tr>
<tr>
<td>0.03</td>
<td>0.16</td>
<td>0.42</td>
<td>0.58</td>
<td>0.16</td>
</tr>
<tr>
<td>0.04</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>0.06</td>
<td>0.24</td>
<td>0.38</td>
<td>0.62</td>
<td>0.24</td>
</tr>
<tr>
<td>0.09</td>
<td>0.3</td>
<td>0.35</td>
<td>0.65</td>
<td>0.3</td>
</tr>
<tr>
<td>0.16</td>
<td>0.4</td>
<td>0.3</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td>0.25</td>
<td>0.5</td>
<td>0.25</td>
<td>0.75</td>
<td>0.5</td>
</tr>
<tr>
<td>0.36</td>
<td>0.6</td>
<td>0.2</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>0.49</td>
<td>0.7</td>
<td>0.15</td>
<td>0.85</td>
<td>0.7</td>
</tr>
<tr>
<td>0.64</td>
<td>0.8</td>
<td>0.1</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>0.81</td>
<td>0.9</td>
<td>0.05</td>
<td>0.95</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) the difference in success rates in a BESD is identical to $r$

Finally, once an effect size has been determined for all the studies it is necessary to determine an overall mean effect size, as well as for each subgroup. At this stage, it is
necessary to weigh the studies differently accordingly to their variance and within study sample size. This will permit the computation of an overall mean effect for which each study contributes differently depending on the size of its population and associated variance. The formula for the weighing factor is presented in Equation 5 (Lipsey & Wilson, 2001):

\[ w = \frac{1}{1 + \left[ ES^2 + 2(n_1 + n_2) \right]} \]  

(5)

where:  

\[ n = \frac{n_1 \times n_2}{n_1 + n_2} \]  

(6)

\( ES \) – is the effect size

\( n_1 \) and \( n_2 \) – are the number of elements in groups 1 and 2

**Corrections**

Several corrections were performed in accordance with meta-analysis theory. As far as weighting is concerned, the only type used was by the formula presented proportional to the inverse of the variance. Other proposed methods of weighing, such as study quality, are highly unreliable (Rosenthal, 1994). However, there are other corrections that should be taken into consideration. One of them deals with the fact that as the population value of \( r \) gets larger its distribution becomes more skewed. To deal with this Fisher (1928) suggested the \( Zr \) transformation (Equation 7):

\[ Zr = \frac{1}{2} \log_e \left[ \frac{1 + r}{1 - r} \right] \]  

(7)
Sampling error associated with the size of each study sample was suggested by Lipsey and Wilson (2001) as another reason for the correction of the effect size estimate. These researchers pointed out that studies with smaller populations have statistically larger measurement errors and suggested a correction (Equation 8) that takes into account the sample size:

$$ES_c = ES - \frac{3}{4N - 9}$$  \hspace{1cm} (8)

where: ES – is the uncorrected effect size

ES_c – is the corrected effect size

N – is the total number of elements in both groups

One other group of corrections suggested for meta-analysis research deals with study imperfections (Hunter & Schmidt, 1990; 1994). However these corrections require information (e.g., reliability of variables) not usually provided in most studies coded for a meta-analysis. Moreover, even when such information is provided for some of the studies being used it rarely is available for all of them. The researcher is then faced with the necessity to “decide if it is better to adjust some effect sizes while not adjusting others or to leave all of them unadjusted under the rationale that they are more comparable that way, even if less accurate” (Lipsey & Wilson, 2001). Furthermore, each adjustment is based upon assumptions about the data and purpose of the meta-analysis. These assumptions make many analysts uncomfortable (Lipsey & Wilson, 2001) as well as working with hypothetical ideal values obtained through corrections instead of the real
observed ones. For these reasons corrections associated with artifact adjustments are not used in this study.

All effect sizes will be computed using the described methodology. An example is presented in appendix B for the effect size level coding from the coding form.

**Treatment of the data/ Statistical methods**

The overall mean effect size (Equation 9) is calculated by dividing the sum of the products between each study individual effect size and its respective weighing factors by the total sum of the weighing factors:

\[
ES = \frac{\sum_i (w_iES_i)}{\sum_i w_i}
\]  

(9)

where:  

- \(ES\) - is the weighted mean effect size  
- \(ES_i\) – is the effect size of the \(i^{th}\) study  
- \(w_i\) – is the inverse variance weight for the effect size of study \(i\)

Confidence intervals were also computed for the overall mean effect sizes. The confidence interval is based on the standard error of the mean effect size and a critical value from the z-distribution Hedges and Olkin (1985) as well as Lipsey and Wilson (2001) provided us with the formula (Equation 10) for the computation of the standard error:

\[
SE_{\frac{1}{ES}} = \sqrt{\frac{1}{\sum w_i}}
\]  

(10)
For a given $\alpha$ it follows then that the formulas for the upper and lower limits of the confidence interval will be given by Equations 11 and 12.

\[
\tilde{ES}_L = \tilde{ES} - z_{(1-\alpha)} \left( \frac{SE}{\tilde{ES}} \right) 
\]

(11)

\[
\tilde{ES}_U = \tilde{ES} + z_{(1-\alpha)} \left( \frac{SE}{\tilde{ES}} \right) 
\]

(12)

where: $\tilde{ES}_U$ and $\tilde{ES}_L$ – are the upper and lower limits of the confidence interval

$z_{(1-\alpha)}$ – is the z value at the 1-$\alpha$ level of significance

$SE_{\tilde{ES}}$ – is the standard error associated with the average effect size

The next important step is the determination of the homogeneity (Equation 13) of the effect size estimates population for the overall sample (Hedges, 1982, Rosenthal & Rubin, 1982). This test is based on the Q statistic, distributed as a chi-square with k-1 degrees of freedom where k is the number of effect sizes.

\[
Q = \sum w_i \left( ES_i - \tilde{ES} \right)^2 
\]

(13)

where: $Q$ – is the Q statistic

$ES_i$ – is the effect size of the $i^{th}$ study

$\tilde{ES}$ – is the average weighted effect size

$w_i$ – is the inverse variance weight for the effect size of study $i$

Having developed an extensive list of possible variables from professional development programs correlated with student learning, the Fixed Effects Model for
meta-analysis as described by Hedges (1994) and Lipsey and Wilson (2001) was employed. This enables the researcher to investigate not only the overall effect size for the impact of professional development for science teachers on student learning, but also to differentiate the impact of the proposed variables.

This methodology can be compared to fixed effects ANOVA, being theoretically very similar. However, the presence of the multitude of studies with different standard deviations (and sizes) requires weighting to enable the comparison. The formula used to calculate the between group homogeneity is given by Equation 14.

\[ Q_{\text{BET}} = \sum w_j^2 \left( \frac{\sum w_j \bar{ES}_j}{\sum w_j} \right)^2 \]  

(14)

where: \( \bar{ES}_j \) – is the weighted mean effect size for each group \( j \)

\( w_j \) – is the sum of weights within each group

and

\[ Q_{\text{WIT}} = \sum w_i \left( ES_i - \bar{ES}_j \right)^2 \]  

(15)

where: \( \bar{ES}_j \) – is the weighted mean effect size for each group \( j \)

\( ES_i \) – is the effect size of the \( i^{th} \) study

\( w_i \) – is the inverse variance weight for the effect size of study \( i \)

or simply \( Q_{\text{WIT}} = Q_{\text{TOT}} - Q_{\text{BET}} \)  

(16)

\( Q_{\text{BET}} \), which has a chi-square distribution with \( j-1 \) degrees of freedom, is then compared with the critical \( Q_c \) with \( j-1 \) degrees of freedom at a certain \( \alpha \) level.
This type of fixed methods Meta-analysis basically implies two hypotheses being tested. First, by calculating the Q value for each group in one category, and the Q between and within groups’ we are testing if the variance for each group and for this combination of groups is superior to or significantly different from the overall variance. A statistically significant Q value indicates a heterogeneous distribution and warrants the researcher the validity of pursuing additional analyses (Lipsey & Wilson, 2001).

Secondly, by computing the effect size another hypothesis is being tested: is the observed effect size from each particular group (or overall) significantly different from zero?

A vote counting approach was also performed to further illustrate the direction of the observed effect. This methodology should not be interpreted in isolation as it may lead the reader towards wrong conclusions (Hedges, 1992). However, when interpreted with the backing up of the effect size analysis it may be a useful indicator of the effect direction.

**Vignettes**

Finally, in order to enrich and enhance the results of the meta-analysis, vignettes will be developed. These vignettes are driven by the quantitative results from the meta-analysis. These results are the backbone or framework from which the description usually present in a vignette emerges. This way, these vignettes should be viewed has more valid than vignettes usually presented only as the result of the philosophical beliefs of the author.
In the words of Campbell (1996) “A vignette is a short story without an ending. It is short, but not too short to present an issue. It is detailed but not so detailed that the underlying issue gets lost.” (p. 4). Vignettes have been used in several science education settings. They are present in the National Science Education Standards (NRC, 1990) to illustrate some of the curriculum recommendations and in Loucks-Horsley et al (1997) work describing different professional development activities for science and mathematics teachers. Moreover, they can even be used inside the classroom to raise science issues among the students (Campbell, 1996) (For an example of a vignette see appendix C).

Vignettes have the power to illustrate recommendations with “semi-real” situations and the ability to make quantitative results more easily understandable through exemplification. According to Campbell (1996, p.4)

There are three major steps to making up vignettes:

1. Determining issues or areas of concern for those who will be using the vignettes

2. Developing situations that are realistic and are relevant for those who will be using them

3. Testing the vignettes with groups similar to those who will be using them to ensure that the vignettes are clear and do provide people with an opportunity to deal with the issues you intended them to
To address the first two points the vignettes were developed focusing on the variables identified as the most significant for their impact on student learning. To test them, they were shared with a group of science education graduate students with teaching experience for feedback and enhancement.

**Pilot Study**

To evaluate the validity of the proposed research design, a meta-analysis with a sample of ten studies was conducted. The goal of the study was to find supportive evidence for the relationship between professional development and student learning. The research question was: Are Professional Development programs in science effective in promoting/enhancing student learning in science? This was crucial for the continuation of the study and for the exploration of the second question: What are the characteristics of the most and least effective programs?

Only quantitative studies on professional development for mathematics and science teachers reporting their impact on student learning, either directly in the form of an effect size or providing enough evidence and data that enabled us to compute the expected effect, were considered eligible. The subjects participating in the study were students in the K-12 education system. The key dependent variable is student learning. The independent variable is the teachers’ participation in a professional development program. For the purpose of this pilot study the students learning was coded as a function of their achievement and/or their attitudes towards science. The time frame established for the studies was from 1990 to the present (restricting the meta-analysis to only
“modern” studies). Both published and unpublished studies were considered eligible, including refereed journals, non-refereed journals, dissertations, government reports, and technical reports. The impact of the professional development programs on student learning was considered significant if both the overall effect size $d$, as established by Cohen (1988), of the sampled studies were greater than 0.30, and if the combined studies were heterogeneous yielding a significant result with $\alpha = 0.05$.

Table 3.3: Summary of the sampled studies for the pilot study

<table>
<thead>
<tr>
<th>Study</th>
<th>$d$</th>
<th>$r$</th>
<th>$Zr$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adelman, N. (1998)</td>
<td>0.293</td>
<td>0.145</td>
<td>0.146</td>
</tr>
<tr>
<td>Baylor, A. &amp; Ritchie, D. (2002)</td>
<td>0.681</td>
<td>0.322</td>
<td>0.335</td>
</tr>
<tr>
<td>Corcoran, T. &amp; Matson, B. (1998)</td>
<td>0.556</td>
<td>0.268</td>
<td>0.276</td>
</tr>
<tr>
<td>Goertz, M. &amp; Carver, R. (1998)</td>
<td>0.700</td>
<td>0.330</td>
<td>0.343</td>
</tr>
<tr>
<td>Goertz, M., Massel, D. &amp; Corcoran, T. (1998)</td>
<td>0.335</td>
<td>0.165</td>
<td>0.167</td>
</tr>
<tr>
<td>Marek, E. &amp; Methven, S. (1991)</td>
<td>0.112</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>Parke, H. &amp; Coble, C. (1997)</td>
<td>0.183</td>
<td>0.091</td>
<td>0.091</td>
</tr>
<tr>
<td>Raghavan, K., Cohen-Regev, S. &amp; Strobel, S. (2001)</td>
<td>0.041</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>Rubin, R., &amp; Norman, J. (1992)</td>
<td>0.523</td>
<td>0.252</td>
<td>0.259</td>
</tr>
<tr>
<td>Zucker, A. &amp; Marder, C. (1998)</td>
<td>0.740</td>
<td>0.347</td>
<td>0.364</td>
</tr>
</tbody>
</table>

where: $d = (M_1 - M_2) / s_w$

$r = d / \sqrt{d^2 + 1/(pq)}$, (p and q represent the proportions of the experimental and control groups)

$Zr = \frac{1}{2} \log_e \left[ \frac{(1 + r)}{(1 - r)} \right]$

Table 3.3 includes a summary of the sampled studies, including the magnitude of the traditional effect sizes – $d$, as established by Cohen (1988), the analogous correlation effect sizes – $r$, and the adjusted effect size – $Zr$ as suggested by Rosenthal (1991). The sample of studies analyzed yield an average adjusted correlation effect size of $\text{Avg} (Zr) =$
0.204 which corresponds to a traditional (Cohen, 1988) effect size of $d = 0.417$. This is considered a medium effect size (Lipsey, & Wilson, 2001). This combination of studies is heterogeneous ($p < 0.0001$).

This pilot study supports a relationship between the teachers’ participation in professional development programs and their students’ learning in science, with a medium sized effect (Lipsey & Wilson, 2001) due to that participation.

**Summary**

Meta-analysis has been determined as the best way to answer the proposed research questions. Its ability to merge results across studies and differentiate between the impacts of different variables is extremely useful in this case. The variables elected for investigation are grounded in the work of past researchers, identified as representative of the characteristics usually distinctive of professional development programs for science teachers. In particular, the fixed methods approach, chosen to be used in this case was explained. Vignettes were also introduced as one more powerful way of enriching the interpretation of the quantitative results attained with the meta-analysis. Finally, to evaluate the validity of the proposed research design the pilot study and its results were described.
Chapter 4
Data Analyses

Introduction

In this chapter the analysis of the collected data is described. First, a description of the sample and of the coding reliability is presented. Second, a vote counting procedure as well as a calculation of effect size for the overall impact of professional development programs for science teachers on student learning is presented to address the first research question. Finally, to answer the second research question, a more detailed effect size computation for the different variables coded in this investigation is discussed.

Description of the Sample

The final sample was composed of 35 different studies. Two of these studies presented data from two different levels and therefore were included twice, generating a total of 37 data points for the calculation of the effect size. To validate the coding procedure, inter coder reliability was calculated and found to be 93%.

Some of the variables initially intended to be coded were reported by less than 20% of the studies and will, therefore, not be presented in this analysis. These include: socio economic status and ability level of the students; size of the school where the teachers worked and its location; amount of funding; initial needs assessment; years of experience, educational background, and gender of the participating teachers; and
experimental mortality. Researchers involved with most of these studies when contacted informed the researcher that they did not collect such information because it was not considered relevant for their studies.

Table 4.1: Study characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Studies (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade Level</td>
<td></td>
</tr>
<tr>
<td>Elem</td>
<td>5</td>
</tr>
<tr>
<td>MS</td>
<td>7</td>
</tr>
<tr>
<td>HS</td>
<td>6</td>
</tr>
<tr>
<td>All levels</td>
<td>9</td>
</tr>
<tr>
<td>NA</td>
<td>10</td>
</tr>
<tr>
<td>Unit of Analysis</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>33</td>
</tr>
<tr>
<td>Teacher</td>
<td>4</td>
</tr>
<tr>
<td>Control for pre-existing differences</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>31</td>
</tr>
<tr>
<td>No</td>
<td>6</td>
</tr>
<tr>
<td>Number of students involved</td>
<td></td>
</tr>
<tr>
<td>&lt;100</td>
<td>6</td>
</tr>
<tr>
<td>100-300</td>
<td>7</td>
</tr>
<tr>
<td>300-1000</td>
<td>11</td>
</tr>
<tr>
<td>&gt;1000</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 4.1 presents other relevant study characteristics. The total number of students represented in the collections of studies for this meta-analysis is 84,651. Table 4.2 contains the authors of the coded studies and the respective effect size both in $d$ and $r$ forms, without any corrections.
Table 4.2: Studies and calculated effect size

<table>
<thead>
<tr>
<th>Studies</th>
<th>d</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adelman (1998)</td>
<td>0.29</td>
<td>0.14</td>
</tr>
<tr>
<td>Austin &amp; Hirstein (1997)</td>
<td>-0.10</td>
<td>-0.05</td>
</tr>
<tr>
<td>Clayton (1989)</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>Corcoran &amp; Matson (1998)</td>
<td>0.56</td>
<td>0.27</td>
</tr>
<tr>
<td>Davison (2000)</td>
<td>-0.54</td>
<td>-0.26</td>
</tr>
<tr>
<td>Dugger &amp; Johnson (1992)</td>
<td>3.45</td>
<td>0.86</td>
</tr>
<tr>
<td>Dugger &amp; Meier (1994)</td>
<td>2.94</td>
<td>0.83</td>
</tr>
<tr>
<td>Fishman, Marx, Best &amp; Tal (2003)</td>
<td>0.15</td>
<td>0.08</td>
</tr>
<tr>
<td>Fletcher &amp; Barufaldi (2003)</td>
<td>0.22</td>
<td>0.11</td>
</tr>
<tr>
<td>Goertz &amp; Carver (1998)</td>
<td>0.70</td>
<td>0.33</td>
</tr>
<tr>
<td>Goertz, Massel &amp; Corcoran (1998)</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>Hounshell &amp; Liggett (1976)</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>Iskandar (1991)</td>
<td>2.87</td>
<td>0.82</td>
</tr>
<tr>
<td>Lawrenz &amp; McCreath (1988)</td>
<td>0.62</td>
<td>0.29</td>
</tr>
<tr>
<td>Lee (2001)</td>
<td>0.33</td>
<td>0.16</td>
</tr>
<tr>
<td>Lu (1993)</td>
<td>1.71</td>
<td>0.65</td>
</tr>
<tr>
<td>Mackinnu (1991)</td>
<td>1.62</td>
<td>0.63</td>
</tr>
<tr>
<td>Marek &amp; Methven (1991)</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>McComas (1989)</td>
<td>1.01</td>
<td>0.45</td>
</tr>
<tr>
<td>Numedhal (1992)</td>
<td>0.41</td>
<td>0.20</td>
</tr>
<tr>
<td>O'Sullivan, Piper &amp; Carbonari (1981)</td>
<td>0.37</td>
<td>0.18</td>
</tr>
<tr>
<td>Otto &amp; Schuck (1983)</td>
<td>1.24</td>
<td>0.53</td>
</tr>
<tr>
<td>Parke &amp; Coble (1997)</td>
<td>-0.22</td>
<td>-0.11</td>
</tr>
<tr>
<td>Raghavan, Cohen-Regev, &amp; Strobel (level 1) (2001)</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>Raghavan, Cohen-Regev, &amp; Strobel (level 2) (2001)</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>Ross, Rolheiser &amp; Hogaboam-Gray (1998)</td>
<td>-0.46</td>
<td>-0.23</td>
</tr>
<tr>
<td>Rubin &amp; Norman (1992)</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>Scarborough &amp; White (1994)</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>Schneider (2001)</td>
<td>1.31</td>
<td>0.55</td>
</tr>
<tr>
<td>Thelen &amp; Litsky (1972)</td>
<td>1.15</td>
<td>0.50</td>
</tr>
<tr>
<td>Vieira (1999)</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>Welch &amp; Walberg (1970)</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>Wilson &amp; Garibaldi (level 1) (1976)</td>
<td>0.14</td>
<td>0.07</td>
</tr>
<tr>
<td>Wilson &amp; Garibaldi (level 2) (1976)</td>
<td>0.55</td>
<td>0.27</td>
</tr>
<tr>
<td>Wolfe (2002)</td>
<td>-0.16</td>
<td>-0.08</td>
</tr>
<tr>
<td>Yager &amp; Weld (1999)</td>
<td>0.75</td>
<td>0.35</td>
</tr>
<tr>
<td>Zucker &amp; Marder (1998)</td>
<td>0.74</td>
<td>0.35</td>
</tr>
</tbody>
</table>

In order to answer the first research question, Are professional development programs effective in enhancing student learning in science?, two approaches were used.
First a simple vote counting methodology investigating the direction of the observed effect sizes was conducted. Second we calculated the overall effect size for this sample of studies and its significance.

**Vote counting**

Vote counting is one of the simplest approaches to investigate the overall direction of the observed effect sizes. In this case, 31 out of the 37 studies that composed our sample reported effect sizes in the direction of the experimental group. That means that 86% of the studies in our sample reported a positive impact of professional development for science teachers on student learning.

**Fixed effects model effect sizes**

The calculation of effect sizes was performed using the corrections and formulas described and presented in Chapter 3. Because the computation of an effect size using a large number of studies requires extreme precision and repetition of formulas, the statistical package SPSS was employed to perform them. However, because SPSS does not permit direct computations of effect sizes it was necessary to employ a Macro program subroutine to do so. The Macro used to enable this computation is presented in appendix D, was developed by Lipsey and Wilson (2001), and is available online at http://mason.gmu.edu/~dwilsonb/ma.html.

To ensure that the Macro was working properly the first computation was also conducted in a Microsoft Excel subroutine developed by this researcher. The results were
exactly the same. Table 4.3 presents the overall effect size for this collection of studies in $Zr$ form. Table 4.4 presents the same results but in the more traditional correlation effect size ($r$) format.

Table 4.3: Overall Effect size ($Zr$) and descriptive statistics (output from the SPSS program)

<table>
<thead>
<tr>
<th>***** Meta-Analytic Results *****</th>
</tr>
</thead>
<tbody>
<tr>
<td>--- Distribution Description ---</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>37.000</td>
</tr>
<tr>
<td>--- Fixed Effects Model ---</td>
</tr>
<tr>
<td>Mean ES</td>
</tr>
<tr>
<td>Fixed</td>
</tr>
<tr>
<td>--- Homogeneity Analysis ---</td>
</tr>
<tr>
<td>Q</td>
</tr>
<tr>
<td>558.4485</td>
</tr>
</tbody>
</table>

A few aspects of these Tables should be emphasized. In Table 4.3 the distribution description shows the reader the total number of studies being used in this computation, the range of observed values (minimum and maximum) and the corresponding weighted standard deviation. Employing a fixed effects model as explained in Chapter 3 the overall mean effect size for this sample is provided, as well as the corresponding 95% confidence interval, the standard error, and the associated $z$ and $p$ values. Finally, the homogeneity analysis is performed to find if whether the various effect sizes that are averaged into a mean value all estimate the same population effect size (Lipsey & Wilson, 2001). A
statistically significant \( Q \), as the one attained, indicates a heterogeneous distribution, warranting the validity of additional analyses.

Table 4.4: Overall Effect size (r)

<table>
<thead>
<tr>
<th>Mean r</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2202</td>
<td>0.2074</td>
<td>0.2337</td>
<td>0.0070</td>
</tr>
</tbody>
</table>

Table 4.4 presents some of the same results as Table 4.3 but in a correlation effect size \( r \) metric instead of the corrected effect size \( Z_r \) reported in Table 4.3. Therefore, it can concluded that this sample of studies has an overall significant effect size \( (r = 0.22) \). Moreover, it should be considered a medium effect size (magnitude from 0.2 to 0.4; Lipsey & Wilson, 2001) as discussed in Chapter 3. Even when considering the 95% confidence interval, the effect size is always greater than 0.2. If Rosenthal’s (1991) Binomial Effect Size Display (BESD) metric is employed, this value represents an increase in the success rate from 0.39 to 0.61 (view Table 3.2, Chapter 3). The distribution of observed correlation effect sizes \( r \) is also presented in Figure 4.1. In this particular case the increase in “success” rate can be interpreted as representing the higher probability of improved student learning for those students whose teachers have been involved in professional development programs.
To further investigate the direction of the observed effect size, differentiation has been made between the studies achieving the maximum confidence in the estimation of their effect size, and all the others where some level of estimation was necessary. Estimation was necessary whenever incomplete data was reported for a particular study, and we were unable to obtain it from the author. The most common example of estimation, for this study, was the assumption of equal group sizes when only one of the group sizes was reported. These results are presented in Tables 4.5 and 4.6. Any coded study where some type of estimation was involved did not attain maximum confidence rating in the effect size computation (see variable 36 appendixes A and B). By doing so
one can determine whether the presence of studies where some estimation was involved in the calculation of effect size under or overestimates the overall result.

Table 4.5: Effect size (Zr) and descriptive statistics for the group of studies with maximum confidence in the estimation of their effect size (output from the SPSS program)

<table>
<thead>
<tr>
<th>Distribution Description</th>
<th>N</th>
<th>Min ES</th>
<th>Max ES</th>
<th>Wghtd SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.000</td>
<td>-.267</td>
<td>1.439</td>
<td>.258</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed &amp; Random Effects Model</th>
<th>Mean ES</th>
<th>-95%CI</th>
<th>+95%CI</th>
<th>SE</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>.3685</td>
<td>.3233</td>
<td>.4136</td>
<td>.0230</td>
<td>15.9858</td>
<td>.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Homogeneity Analysis</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>125.1195</td>
<td>13.000</td>
<td></td>
<td>.0000</td>
</tr>
</tbody>
</table>

The analysis of Tables 4.5 and 4.6 shows that both groups present medium effect sizes in the direction of the experimental group. Moreover, the fact that the group of studies with the maximum confidence in the estimation of the effect size, presents an even higher overall weighted effect size, leads to the conclusion that, at the most, our final effect size may be underestimated. Therefore, all the attained results, if significant for the overall sample, would also be significant if only “perfect” studies had been used. For this reason all the subsequent analysis has been done using the entire sample.

To answer the second research question, “What are the characteristics of the most and least effective programs?”, the different characteristics of such programs were coded. Differentiation between their impacts was done using the fixed effect model to calculate their corresponding effect sizes.
Table 4.6: Effect size (Zr) and descriptive statistics for the group of studies with less than the maximum confidence in the estimation of their effect size (output from the SPSS program)

<table>
<thead>
<tr>
<th>N</th>
<th>Min ES</th>
<th>Max ES</th>
<th>Wghtd SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>23.000</td>
<td>-.111</td>
<td>1.158</td>
<td>.146</td>
</tr>
</tbody>
</table>

----- Fixed & Random Effects Model

<table>
<thead>
<tr>
<th>Mean ES</th>
<th>-95%CI</th>
<th>+95%CI</th>
<th>SE</th>
<th>Z</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>.2094</td>
<td>.1949</td>
<td>.2239</td>
<td>.0074</td>
<td>28.3114</td>
</tr>
</tbody>
</table>

----- Homogeneity Analysis

<table>
<thead>
<tr>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>390.1705</td>
<td>22.0000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

The characteristic type of treatment was defined according to the work of Loucks-Horsley et al. (1998), and its results are presented in Tables 4.7 and 4.8. One should note, when analyzing the data in Table 4.7, the significance level attained for the homogeneity analysis. Both the between and within group variances are small, yielding correspondingly large Q’s; and p’s < 0.0001.

Table 4.7: Effect size (Zr) and descriptive statistics for the type of treatment variable (output from the SPSS program)

<table>
<thead>
<tr>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>212.9166</td>
<td>12.0000</td>
</tr>
<tr>
<td>Within</td>
<td>1046.4089</td>
<td>89.0000</td>
</tr>
<tr>
<td>Total</td>
<td>1259.3256</td>
<td>101.0000</td>
</tr>
</tbody>
</table>
Table 4.7: Effect size (Zr) and descriptive statistics for the type of treatment variable (continued)

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshop</td>
<td>60.2932</td>
<td>14.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Curric. Dev</td>
<td>111.4902</td>
<td>4.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Curric. Repl</td>
<td>87.5852</td>
<td>4.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Curric Impl</td>
<td>483.5094</td>
<td>19.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Inquiry</td>
<td>5.6237</td>
<td>4.0000</td>
<td>.2291</td>
</tr>
<tr>
<td>Action Research</td>
<td>11.8002</td>
<td>4.0000</td>
<td>.0189</td>
</tr>
<tr>
<td>Case discussion</td>
<td>5.6694</td>
<td>6.0000</td>
<td>.4612</td>
</tr>
<tr>
<td>Study Groups</td>
<td>10.7603</td>
<td>7.0000</td>
<td>.1494</td>
</tr>
<tr>
<td>Coaching and mentoring</td>
<td>3.6047</td>
<td>6.0000</td>
<td>.7300</td>
</tr>
<tr>
<td>Partnerships</td>
<td>56.4109</td>
<td>4.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Partnerships with scient</td>
<td>3.9703</td>
<td>6.0000</td>
<td>.6807</td>
</tr>
<tr>
<td>Technology for PD</td>
<td>153.7157</td>
<td>6.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Developing PD</td>
<td>51.9757</td>
<td>5.0000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Es</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>Z</th>
<th>p</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshop</td>
<td>0.1503</td>
<td>0.0103</td>
<td>0.1301</td>
<td>0.1705</td>
<td>14.5900</td>
<td>0.0000</td>
<td>15</td>
</tr>
<tr>
<td>Curric. Dev</td>
<td>0.3902</td>
<td>0.0316</td>
<td>0.3283</td>
<td>0.4520</td>
<td>12.3619</td>
<td>0.0000</td>
<td>5</td>
</tr>
<tr>
<td>Curric Rep.</td>
<td>0.4105</td>
<td>0.0336</td>
<td>0.3447</td>
<td>0.4763</td>
<td>12.2248</td>
<td>0.0000</td>
<td>5</td>
</tr>
<tr>
<td>Curric Impl</td>
<td>0.2384</td>
<td>0.0079</td>
<td>0.2230</td>
<td>0.2538</td>
<td>30.3255</td>
<td>0.0000</td>
<td>20</td>
</tr>
<tr>
<td>Inquiry</td>
<td>0.0853</td>
<td>0.0262</td>
<td>0.0339</td>
<td>0.1368</td>
<td>3.2513</td>
<td>0.0011</td>
<td>5</td>
</tr>
<tr>
<td>Action Research</td>
<td>0.0109</td>
<td>0.0539</td>
<td>-0.0948</td>
<td>0.1166</td>
<td>0.2022</td>
<td>0.8398</td>
<td>5</td>
</tr>
<tr>
<td>Case discussion</td>
<td>0.0906</td>
<td>0.0405</td>
<td>0.0112</td>
<td>0.1701</td>
<td>2.2361</td>
<td>0.0253</td>
<td>7</td>
</tr>
<tr>
<td>Study Groups</td>
<td>0.0606</td>
<td>0.0396</td>
<td>-0.0171</td>
<td>0.1383</td>
<td>1.5296</td>
<td>0.1261</td>
<td>8</td>
</tr>
<tr>
<td>Coaching and mentoring</td>
<td>0.1236</td>
<td>0.0257</td>
<td>0.0733</td>
<td>0.1739</td>
<td>4.8120</td>
<td>0.0000</td>
<td>7</td>
</tr>
<tr>
<td>Partnerships</td>
<td>0.2207</td>
<td>0.0115</td>
<td>0.1982</td>
<td>0.2431</td>
<td>19.2490</td>
<td>0.0000</td>
<td>5</td>
</tr>
<tr>
<td>Part. with scient</td>
<td>0.1413</td>
<td>0.0121</td>
<td>0.1175</td>
<td>0.1650</td>
<td>11.6375</td>
<td>0.0000</td>
<td>7</td>
</tr>
<tr>
<td>Technology for PD</td>
<td>0.1908</td>
<td>0.0118</td>
<td>0.1678</td>
<td>0.2139</td>
<td>16.2410</td>
<td>0.0000</td>
<td>7</td>
</tr>
<tr>
<td>Developing PD</td>
<td>0.2431</td>
<td>0.0223</td>
<td>0.1993</td>
<td>0.2868</td>
<td>10.8969</td>
<td>0.0000</td>
<td>6</td>
</tr>
</tbody>
</table>
The Q and p values for each group within this category indicate that most of the coded groups are heterogeneous (as evident through their high Q’s and small p’s). Higher effect sizes were attained by the treatments “Curriculum Development” and “Curriculum Replacement”, very close to being considered high (more than 0.4). Medium effect sizes were achieved by the treatments “Curriculum Implementation”; “Partnerships” and “Developing Professional Development”. All the other types of treatment were significant (at the α = 0.05 level) even though small effect sizes (less than 0.2) with the exception of “Action Research” and “Study Groups” that attained non-significant results.

Table 4.8: Effect size (r) for the type of treatment variable

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean r</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshop</td>
<td>0.1492</td>
<td>0.0103</td>
<td>0.1294</td>
<td>0.1689</td>
</tr>
<tr>
<td>Curric. Dev.</td>
<td>0.3715</td>
<td>0.0316</td>
<td>0.3170</td>
<td>0.4235</td>
</tr>
<tr>
<td>Curric Rep.</td>
<td>0.3889</td>
<td>0.0336</td>
<td>0.3317</td>
<td>0.4433</td>
</tr>
<tr>
<td>Curric Impl</td>
<td>0.2340</td>
<td>0.0079</td>
<td>0.2194</td>
<td>0.2485</td>
</tr>
<tr>
<td>Inquiry</td>
<td>0.0851</td>
<td>0.0262</td>
<td>0.0394</td>
<td>0.1360</td>
</tr>
<tr>
<td>Action Research</td>
<td>0.0109</td>
<td>0.0538</td>
<td>-0.0945</td>
<td>0.1161</td>
</tr>
<tr>
<td>Case discussion</td>
<td>0.0904</td>
<td>0.0405</td>
<td>0.0112</td>
<td>0.1685</td>
</tr>
<tr>
<td>Study Groups</td>
<td>0.0605</td>
<td>0.0396</td>
<td>-0.0171</td>
<td>0.1374</td>
</tr>
<tr>
<td>Coaching and mentoring</td>
<td>0.1230</td>
<td>0.0257</td>
<td>0.0732</td>
<td>0.1722</td>
</tr>
<tr>
<td>Partnerships</td>
<td>0.2172</td>
<td>0.0115</td>
<td>0.1956</td>
<td>0.2384</td>
</tr>
<tr>
<td>Part. with scient</td>
<td>0.1404</td>
<td>0.0121</td>
<td>0.1170</td>
<td>0.1635</td>
</tr>
<tr>
<td>Technology for PD</td>
<td>0.1885</td>
<td>0.0118</td>
<td>0.1662</td>
<td>0.2107</td>
</tr>
<tr>
<td>Developing PD</td>
<td>0.2384</td>
<td>0.0223</td>
<td>0.1967</td>
<td>0.2792</td>
</tr>
</tbody>
</table>

The most relevant trend seems to be the greater effect size achieved by the groups dealing with curriculum reform. They represent the 1st, 2nd, and 4th largest effect sizes.
observed. Two of the intended variables did not appear in the coding procedure. The treatments “Examining student work” and “Immersion in the world of scientists and mathematicians” are therefore absent from this analysis.

Table 4.9: Effect size (Zr) and descriptive statistics for the theoretical dimension variable (output from the SPSS program)

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>457.8663</td>
<td>5.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Within</td>
<td>456.6624</td>
<td>41.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Total</td>
<td>914.5286</td>
<td>46.0000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science as Inquiry</td>
<td>25.0946</td>
<td>7.0000</td>
<td>.0007</td>
</tr>
<tr>
<td>Physical science</td>
<td>134.6811</td>
<td>11.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Life science</td>
<td>1.2487</td>
<td>4.0000</td>
<td>.8700</td>
</tr>
<tr>
<td>Earth and space</td>
<td>3.5832</td>
<td>4.0000</td>
<td>.4653</td>
</tr>
<tr>
<td>STS</td>
<td>111.3423</td>
<td>6.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Unifying concepts</td>
<td>180.7125</td>
<td>9.0000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>Z</th>
<th>p</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science as Inquiry</td>
<td>0.9060</td>
<td>0.0175</td>
<td>0.5620</td>
<td>0.1249</td>
<td>501695.0</td>
<td>0.0000</td>
<td>8</td>
</tr>
<tr>
<td>Physical science</td>
<td>0.1503</td>
<td>0.0152</td>
<td>0.1205</td>
<td>0.1802</td>
<td>9.8725</td>
<td>0.0000</td>
<td>12</td>
</tr>
<tr>
<td>Life science</td>
<td>0.0784</td>
<td>0.0234</td>
<td>0.0325</td>
<td>0.1244</td>
<td>3.3484</td>
<td>0.0008</td>
<td>5</td>
</tr>
<tr>
<td>Earth and space</td>
<td>0.0337</td>
<td>0.0534</td>
<td>-0.0710</td>
<td>0.1384</td>
<td>0.6302</td>
<td>0.5285</td>
<td>5</td>
</tr>
<tr>
<td>STS</td>
<td>0.4866</td>
<td>0.0219</td>
<td>0.4437</td>
<td>0.5296</td>
<td>22.2418</td>
<td>0.0000</td>
<td>7</td>
</tr>
<tr>
<td>Unifying concepts</td>
<td>0.5127</td>
<td>0.0214</td>
<td>0.4707</td>
<td>0.5547</td>
<td>23.9390</td>
<td>0.0000</td>
<td>10</td>
</tr>
</tbody>
</table>
The next variable reported is the theoretical dimension as defined in the National Science Education Standards (1996). The results for this variable are presented in Tables 4.9 and 4.10. The homogeneity analysis, again, validated the grouping into these categories. Moreover, all the groups in Table 4.9 are heterogeneous with the exception of “Life science” and “Earth and space”. The dimensions “Science and Society” and “Science and Technology”, always appeared together in studies dealing with what is commonly called “Science-Technology-Society (STS)”. Therefore, these two dimensions were merged into just one called STS. The dimension History and Nature of Science never appeared in our sample.

As we can see on Table 4.10 the theoretical dimensions “Science as Inquiry”, “STS”, and “Unifying concepts” all achieved large effect sizes (more than 0.4). The dimensions “Physical science” and “Life science” achieved small, but significant results. On the other hand, the dimension “Earth and Space” was the only one with a non-significant result.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean r</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science as Inquiry</td>
<td>0.7192</td>
<td>0.0175</td>
<td>0.5095</td>
<td>0.1243</td>
</tr>
<tr>
<td>Physical science</td>
<td>0.1492</td>
<td>0.0152</td>
<td>0.1199</td>
<td>0.1783</td>
</tr>
<tr>
<td>Life science</td>
<td>0.0782</td>
<td>0.0234</td>
<td>0.0325</td>
<td>0.1238</td>
</tr>
<tr>
<td>Earth and space</td>
<td>0.0324</td>
<td>0.0533</td>
<td>-0.0709</td>
<td>0.1375</td>
</tr>
<tr>
<td>STS</td>
<td>0.4515</td>
<td>0.0219</td>
<td>0.4167</td>
<td>0.4851</td>
</tr>
<tr>
<td>Unifying concepts</td>
<td>0.4720</td>
<td>0.0214</td>
<td>0.4388</td>
<td>0.5040</td>
</tr>
</tbody>
</table>
The next characteristic presented is the additional context variables as defined in Chapter 3. The results are presented in Tables 4.11 and 4.12.

Table 4.11: Effect size (Zr) and descriptive statistics for the Additional context variable (output from the SPSS program)

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific content</td>
<td>39.4380</td>
<td>5.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Pedagogy</td>
<td>15.7886</td>
<td>6.0000</td>
<td>.0149</td>
</tr>
<tr>
<td>PCK</td>
<td>174.3354</td>
<td>10.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Technology</td>
<td>139.2215</td>
<td>8.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Testing, assessment</td>
<td>221.1075</td>
<td>10.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Standards</td>
<td>288.3965</td>
<td>10.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>other</td>
<td>.0000</td>
<td>.0000</td>
<td>-9.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Es</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>Z</th>
<th>p</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific content</td>
<td>0.1455</td>
<td>0.0173</td>
<td>0.1117</td>
<td>0.1793</td>
<td>8.4336</td>
<td>0.0000</td>
<td>6</td>
</tr>
<tr>
<td>Pedagogy</td>
<td>0.1324</td>
<td>0.0126</td>
<td>0.1076</td>
<td>0.1572</td>
<td>10.4776</td>
<td>0.0000</td>
<td>7</td>
</tr>
<tr>
<td>PCK</td>
<td>0.4213</td>
<td>0.0195</td>
<td>0.3831</td>
<td>0.4595</td>
<td>21.6090</td>
<td>0.0000</td>
<td>11</td>
</tr>
<tr>
<td>Technology</td>
<td>0.1485</td>
<td>0.0110</td>
<td>0.1269</td>
<td>0.1700</td>
<td>13.4757</td>
<td>0.0000</td>
<td>9</td>
</tr>
<tr>
<td>Testing, assessment</td>
<td>0.3130</td>
<td>0.0110</td>
<td>0.2913</td>
<td>0.3346</td>
<td>28.3423</td>
<td>0.0000</td>
<td>11</td>
</tr>
<tr>
<td>Standards</td>
<td>0.2956</td>
<td>0.0144</td>
<td>0.2673</td>
<td>0.3238</td>
<td>20.5249</td>
<td>0.0000</td>
<td>11</td>
</tr>
<tr>
<td>other</td>
<td>0.1460</td>
<td>0.0138</td>
<td>0.1190</td>
<td>0.1729</td>
<td>10.6136</td>
<td>0.0000</td>
<td>1</td>
</tr>
</tbody>
</table>
All these variables achieved significant and heterogeneous results, with “Pedagogical Content Knowledge” achieving the largest effect size. The reason for the -9 under the p column for the “other” group on Table Q, is due to the fact that only one study was coded as “other” (one study reported a strong emphasis on Equity as part of the professional development program) making the degrees of freedom for that variable zero and producing an impossible result ($+\infty$) for $p$ that the SPSS output reported as -9.

Table 4.12: Effect size (r) for the additional context variable

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean r</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific content</td>
<td>0.1445</td>
<td>0.0173</td>
<td>0.1112</td>
<td>0.1774</td>
</tr>
<tr>
<td>Pedagogy</td>
<td>0.1316</td>
<td>0.0126</td>
<td>0.1072</td>
<td>0.1559</td>
</tr>
<tr>
<td>PCK</td>
<td>0.3980</td>
<td>0.0195</td>
<td>0.3654</td>
<td>0.4297</td>
</tr>
<tr>
<td>Technology</td>
<td>0.1474</td>
<td>0.0110</td>
<td>0.1262</td>
<td>0.1684</td>
</tr>
<tr>
<td>Testing, assessment</td>
<td>0.3032</td>
<td>0.0110</td>
<td>0.2833</td>
<td>0.3226</td>
</tr>
<tr>
<td>Standards</td>
<td>0.2873</td>
<td>0.0144</td>
<td>0.2611</td>
<td>0.3129</td>
</tr>
<tr>
<td>other</td>
<td>0.1450</td>
<td>0.0138</td>
<td>0.1184</td>
<td>0.1712</td>
</tr>
</tbody>
</table>

Even though all of these variables have attained significant results there is a difference in the magnitude of those results. “Pedagogical Content Knowledge” achieved the largest effect size (almost 0.4, the level at which correlation effect sizes start being considered large). “Testing and Assessment” and “Standards” also achieved medium sized effect sizes. All the other proposed variables only attained small, albeit still significant, effect sizes.

Finally, the results addressing the variables dealing with the duration of the program are presented. The results for the spread variable are presented on Tables 4.13...
and 4.14; while the results for the total duration variable are presented on the Tables 4.15 and 4.16.

Table 4.13: Effect size (Z_r) and descriptive statistics for the Spread variable (output from the SPSS program)

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>50.3581</td>
<td>3.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Within</td>
<td>508.0905</td>
<td>33.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Total</td>
<td>558.4485</td>
<td>36.0000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1 week</td>
<td>7.1975</td>
<td>6.0000</td>
<td>.3030</td>
</tr>
<tr>
<td>Up to 1 month</td>
<td>12.1393</td>
<td>5.0000</td>
<td>.0329</td>
</tr>
<tr>
<td>Up to 6 months</td>
<td>30.6979</td>
<td>4.0000</td>
<td>.0000</td>
</tr>
<tr>
<td>Up to 1 year</td>
<td>458.0558</td>
<td>18.0000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Es</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>Z</th>
<th>p</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1 week</td>
<td>0.0953</td>
<td>0.0434</td>
<td>0.0103</td>
<td>0.1804</td>
<td>2.1963</td>
<td>0.0281</td>
<td>7</td>
</tr>
<tr>
<td>Up to 1 month</td>
<td>0.1376</td>
<td>0.0221</td>
<td>0.0942</td>
<td>0.1810</td>
<td>6.2140</td>
<td>0.0000</td>
<td>6</td>
</tr>
<tr>
<td>Up to 6 months</td>
<td>0.1341</td>
<td>0.0231</td>
<td>0.0889</td>
<td>0.1793</td>
<td>5.8165</td>
<td>0.0000</td>
<td>5</td>
</tr>
<tr>
<td>Up to 1 year</td>
<td>0.2507</td>
<td>0.0080</td>
<td>0.2350</td>
<td>0.2663</td>
<td>31.4167</td>
<td>0.0000</td>
<td>19</td>
</tr>
</tbody>
</table>

In Tables 4.14 and 4.16, one may note that the longer the duration and spread of the program the larger the attained effect size. Moreover, programs of shorter duration present more homogeneous results.
Table 4.14: Effect size (r) for the spread variable

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean r</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1 week</td>
<td>0.0950</td>
<td>0.0434</td>
<td>0.0103</td>
<td>0.1785</td>
</tr>
<tr>
<td>Up to 1 month</td>
<td>0.1367</td>
<td>0.0221</td>
<td>0.0939</td>
<td>0.1790</td>
</tr>
<tr>
<td>Up to 6 months</td>
<td>0.1333</td>
<td>0.0231</td>
<td>0.0887</td>
<td>0.1774</td>
</tr>
<tr>
<td>Up to 1 year</td>
<td>0.2456</td>
<td>0.0080</td>
<td>0.2308</td>
<td>0.2602</td>
</tr>
</tbody>
</table>

As far as the spread is concerned, all the groups attained significant results, but only the group with spreads over six months achieved a medium effect size. The importance of duration is also emphasized by the results from Table 4.16; only the group with over 100 contact hours had a medium effect size.

Table 4.15: Effect size (Zr) and descriptive statistics for the total duration variable (output from the SPSS program)

------- Analog ANOVA table (Homogeneity Q) -------

<table>
<thead>
<tr>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>88.0627</td>
<td>4.0000</td>
</tr>
<tr>
<td>Within</td>
<td>470.3859</td>
<td>32.0000</td>
</tr>
<tr>
<td>Total</td>
<td>558.4485</td>
<td>36.0000</td>
</tr>
</tbody>
</table>

------- Q by Group -------

<table>
<thead>
<tr>
<th>Group</th>
<th>Q</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>8hrs</td>
<td>1.6259</td>
<td>3.0000</td>
<td>.6535</td>
</tr>
<tr>
<td>9-16</td>
<td>8.0085</td>
<td>2.0000</td>
<td>.0182</td>
</tr>
<tr>
<td>17-50</td>
<td>24.0611</td>
<td>6.0000</td>
<td>.0005</td>
</tr>
<tr>
<td>51-100</td>
<td>17.8860</td>
<td>6.0000</td>
<td>.0065</td>
</tr>
<tr>
<td>over 100</td>
<td>418.8044</td>
<td>15.000</td>
<td>.0000</td>
</tr>
</tbody>
</table>

------- Effect Size Results by Group -------

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Es</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>Z</th>
<th>p</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>8hrs</td>
<td>-0.0250</td>
<td>0.0707</td>
<td>-0.1636</td>
<td>0.1136</td>
<td>-0.3535</td>
<td>0.7237</td>
<td>4</td>
</tr>
<tr>
<td>9-16</td>
<td>0.0907</td>
<td>0.0534</td>
<td>-0.0139</td>
<td>0.1953</td>
<td>1.6989</td>
<td>0.0893</td>
<td>3</td>
</tr>
<tr>
<td>17-50</td>
<td>0.1289</td>
<td>0.0376</td>
<td>0.0552</td>
<td>0.2026</td>
<td>3.4295</td>
<td>0.0006</td>
<td>7</td>
</tr>
<tr>
<td>51-100</td>
<td>0.1242</td>
<td>0.0157</td>
<td>0.0935</td>
<td>0.1549</td>
<td>7.9269</td>
<td>0.0000</td>
<td>7</td>
</tr>
<tr>
<td>over 100</td>
<td>0.2629</td>
<td>0.0082</td>
<td>0.2468</td>
<td>0.2790</td>
<td>32.0058</td>
<td>0.0000</td>
<td>16</td>
</tr>
</tbody>
</table>
Furthermore, the two groups with smaller durations did not even attain significant results. A possible interaction between the reported variables was also explored. However, no significant results were found.

Table 4.16: Effect size (r) for the total duration variable

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean r</th>
<th>SE</th>
<th>-95% CI</th>
<th>+95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>8hrs</td>
<td>-0.0250</td>
<td>0.0706</td>
<td>-0.1622</td>
<td>0.1131</td>
</tr>
<tr>
<td>9-16</td>
<td>0.0905</td>
<td>0.0533</td>
<td>-0.0139</td>
<td>0.1929</td>
</tr>
<tr>
<td>17-50</td>
<td>0.1282</td>
<td>0.0376</td>
<td>0.0551</td>
<td>0.1999</td>
</tr>
<tr>
<td>51-100</td>
<td>0.1236</td>
<td>0.0157</td>
<td>0.0932</td>
<td>0.1537</td>
</tr>
<tr>
<td>over 100</td>
<td>0.2570</td>
<td>0.0082</td>
<td>0.2419</td>
<td>0.2720</td>
</tr>
</tbody>
</table>

Table 4.17 summarizes the results from the data analyses. It includes all the characteristics from the coded variables that attained at least a medium effect size ($r > 0.2$; Lipsey & Wilson, 2001). This enables us to answer the second research question, identifying the characteristics of professional development programs that have a larger impact on student learning.

The results achieved with this analysis will now be used as the skeleton for the construction of the Vignettes presented in the next chapter. In this way, any proposed Vignette for a professional development program is grounded in solid and significant quantitative results.
Table 4.17: The impact of key professional development characteristics on student learning

<table>
<thead>
<tr>
<th>Key Professional Development characteristics</th>
<th>Mean r</th>
<th>-95% CI</th>
<th>+95% CI</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum Replacement</td>
<td>0.3889</td>
<td>0.3317</td>
<td>0.4433</td>
<td>0.0336</td>
</tr>
<tr>
<td>Curriculum Development</td>
<td>0.3715</td>
<td>0.3170</td>
<td>0.4235</td>
<td>0.0316</td>
</tr>
<tr>
<td>Developing PD</td>
<td>0.2384</td>
<td>0.1967</td>
<td>0.2792</td>
<td>0.0223</td>
</tr>
<tr>
<td>Curriculum Implementation</td>
<td>0.2340</td>
<td>0.2194</td>
<td>0.2485</td>
<td>0.0079</td>
</tr>
<tr>
<td>Partnerships</td>
<td>0.2172</td>
<td>0.1956</td>
<td>0.2384</td>
<td>0.0115</td>
</tr>
<tr>
<td>Theoretical dimension</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science as Inquiry</td>
<td>0.7192</td>
<td>0.5095</td>
<td>0.1243</td>
<td>0.0175</td>
</tr>
<tr>
<td>Unifying concepts</td>
<td>0.4720</td>
<td>0.4388</td>
<td>0.5040</td>
<td>0.0214</td>
</tr>
<tr>
<td>Science/Technology/Society (STS)</td>
<td>0.4515</td>
<td>0.4167</td>
<td>0.4851</td>
<td>0.0219</td>
</tr>
<tr>
<td>Additional context variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCK</td>
<td>0.3980</td>
<td>0.3654</td>
<td>0.4297</td>
<td>0.0195</td>
</tr>
<tr>
<td>Testing, assessment</td>
<td>0.3032</td>
<td>0.2833</td>
<td>0.3226</td>
<td>0.0110</td>
</tr>
<tr>
<td>Standards</td>
<td>0.2873</td>
<td>0.2611</td>
<td>0.3129</td>
<td>0.0144</td>
</tr>
<tr>
<td>Time spread</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over six months</td>
<td>0.2456</td>
<td>0.2308</td>
<td>0.2602</td>
<td>0.0080</td>
</tr>
<tr>
<td>Total duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over 100 hours</td>
<td>0.2570</td>
<td>0.2419</td>
<td>0.2720</td>
<td>0.0082</td>
</tr>
</tbody>
</table>

Summary

This chapter includes the data analysis for this investigation, in a way that addresses the posed research questions. The results indicate a significant impact of professional development programs for science teachers on student learning. Moreover, the results also indicate a differentiation between the impact of several of the characteristics of professional development programs. In particular, the greater impact was attained by programs focused on curricular reform (development, replacement and
implementation), science as inquiry, pedagogical content knowledge, and with durations over 100 hours and spread over more than six months.
Chapter 5
Discussion, Implications and Further Research

Introduction

In this chapter the data presented in Chapter 4 to answer the initial research questions are compared and contrasted to the “state of the art” as described in Chapter 2. These findings are then presented in the form of vignettes directly developed from the data analysis. Implications for professional development programs for science teachers are discussed in light of these findings. Finally, suggestions for further research are presented.

Discussion

Evaluating student learning

The number of studies dealing with professional development for science teachers in the last 35 years is impressive. The simplest search using an online engine such as ERIC yields over 4000 references. However, the variety of the references is tremendous. If attention is concentrated only on studies using quantitative methodologies the number of references is approximately 200. Within these references, only 35 focused on student achievement as one of the measured outcomes.

It is important to assess the impact of professional development on student learning. The current climate of accountability requires students to achieve higher
standards, teachers to be held accountable for the students’ results and professional
developers to show that their programs had a positive effect on student achievement
(Bush, 2002; Guskey, 1998). This emphasis on data dealing with student achievement is
also present in some of the National Research Council reports. It is clear in “Scientific
Research on Education” (NRC, 2002) when this report calls for more “statistical
synthesis (that) provide means of aggregating data across studies” (p.144). It is also
evident in the publication “Educating Teachers of Science, Mathematics, and
Technology: New Practices for the New Millennium” (NRC, 2001) that calls for
“synthesizing data across studies and linking it to school practice” (p.121).

The effectiveness of professional development for science teachers in enhancing
student learning

The data analysis is conclusive when answering the first research question. There
was a significant effect on the learning by their students when the teachers participate in
professional development programs. This effect ($r = 0.22$) is considered medium in size
according to the literature (Lipsey & Wilson, 2001). If one employs Rosenthal’s (1991)
Binomial Effect Size Display metric it becomes even more evident. It represents an
increase in the success rate from 39% to 61%. By no means can one ignore the
significance of this result. This stresses the need to create ways to promote the
participation of teachers in such programs.

In some countries, professional development is mandatory, in others it is not. In
the U.S. it varies among states. Despite these differences, educators must try to find ways
to motivate teachers to enroll in these types of programs if we are serious about the quality of their teaching and the learning of their students. As Ponte (1994) discusses, mandatory enrollment may not be the best solution. Intrinsic motivation on the part of teachers is a very important factor. The portrayal of the teaching profession in a more professional light (Darling-Hammond, 1993) may encourage teachers to continue to grow professionally.

*The characteristics of the most and least effective programs*

This question was also addressed with success. The coding procedure employed grouped the studies into relevant and homogenous categories. Unfortunately, not all of the variables initially intended for codification were sufficiently reported on the sampled studies. This again reflects the poorly designed evaluation that has permeated much of the research dealing with professional development.

From within the coded variables interesting results were derived. The variables – type of treatment, theoretical domain, additional context variables, spread and time – all represented significant ways of grouping the studies and different magnitudes for the effect size attained for different characteristics.

*Type of treatment*

The variable type or treatment defined from the work of Loucks-Horsley et al. (1998) grouped the studies into 15 different types of professional development activities that are representative of the experiences of science teachers. Not all of the results were
statistically significant. And even between the ones which were, the magnitudes are varied.

Emphasis must be given to the high impact of the treatments dealing with curriculum replacement, implementation, and development. These treatments produced three of the top four effect sizes for the type of treatment variable. This result is aligned with the literature sustaining the support for teachers working with curriculum (Bencze & Hodson, 1999).

Two other characteristics represent at least medium effects ($r > 0.2$). Developing professional developers ($3^{rd}$ overall) and Partnerships ($5^{th}$ overall). Developing professional developers is a treatment of great interest and requires further research. This researcher believes that when developing professional development teams, the members must be very carefully selected. The composition of such a team was not controlled for in this analysis. This researcher believes that having a team composed of educators, scientists, and master teachers as suggested by Fletcher and Barufaldi (2002) will have a greater impact than a team composed only of administrators.

The variable called Partnerships represents what Loucks-Horsley et al. (1998) entitled Professional Networks, while others may simply call it collaboration. Again, this is not a new finding in the literature, Barufaldi and Reinhartz (2001) stressed the power of collaboration and the importance of consciously promoting it.

Several other types of professional development attained small but still significant effect sizes. Specifically they include: workshops, inquiry, case discussions, coaching and mentoring, partnerships with scientists and technology for professional development. The
fact that these categories only achieved small effect sizes does not mean that they should be disregarded. One must remember that the definition of small effect size is not clear cut and is often under discussion. One should also remember that often, especially in medicine, correlation effect sizes as small as 0.034 have been considered big enough to make medical decisions. These treatments should not be discounted because they attained only small effect sizes. A small effect size is certainly more desirable than a negative one. Perhaps, however, one should use these types of treatments as secondary to those achieving larger effect sizes, when the goal of professional development is to improve student learning. The literature (Loucks-Horsley et al., 1998) suggests that most professional development programs do not rely simply on one component, or type of treatment. Therefore, the addition of these other treatments as complements should be regarded as natural and beneficial.

Two of the coded types of treatment – Action Research and Study Groups – did not attain significant effect sizes for their impact on student learning. Both of these approaches have received attention from researchers and are valuable within certain contexts. Action Research is even one of the strands represented in the National Association for Research in Science Teaching (2004) international conference. Nonetheless, if the focus of a professional development program is student learning, the results do not support the use of these techniques. They may be effective in promoting other outcomes of professional development, such as teacher retention or renewal which are important outcomes that a particular program may be focusing on. However, if the
focus is on student learning, it is recommended that one of the other treatments identified as having a significant impact on student learning be utilized.

Theoretical dimension

The variable theoretical dimension presented very interesting results. The most compelling of all is the very large effect size achieved by the programs focusing on science as inquiry ($r = 0.72$). This represents a dramatic increase in the success rate of students (Rosenthal, 1991), from 15% to 85%. Science as inquiry has received the attention of many researchers (NRC, 1996, 2001). This result reemphasizes its importance for the promotion of student learning.

Two other theoretical dimensions achieved large effect sizes: Science, Technology, and Society (STS), and Unifying concepts. STS has been the focus of several studies during the 1990’s to establish its validity with relevance to the work of Yager (1993, 1996, 1999). Unifying concepts have also demonstrated a positive impact on student learning as shown in the meta-analysis done by Hurley (2001), who investigated the integration of concepts dealing with science and mathematics. Physical Science and Life Science both achieved small though significant effect sizes.

One very interesting variable to evaluate would have been the alignment between the theoretical domains focused in the professional development program with the perceived needs of the teachers participating in the program. Unfortunately the data collected is not sufficient to pursue this line of inquiry.
Additional context variables

The grouping created by the characteristic additional context variables was significant for every category suggesting the relevance of these variables in professional development programs. Three of these variables attained medium effect sizes: Pedagogical Content Knowledge; Testing and Assessment, and Standards.

Pedagogical Content Knowledge, in particular, was the largest of the effect sizes ($r = 0.398$) within all of the additional suggested variables. This finding is also supported by the work of Shulman (1986), emphasizing the importance of giving teachers the opportunity to be, not only pedagogically sound, but also proficient in those pedagogical practices aligned with the particular content they teach.

The focus on testing and authentic assessment has also received attention from several researchers. Neill (1997) in his work commenting on the report, “Principles and Indicators for Student Assessment Systems” developed by the National Forum on Assessment (1995), emphasizes assessment has having as its main purpose the improvement of student learning. Even when it has other purposes assessment should still support student learning.

The other variables, Focusing on a specific content, Pedagogy, and the Use of technology, all attained small but significant effect sizes. This indicated that they do make a significant contribution to student learning.
Time

The variable time was defined in two different ways that attained similar results. Both the spread and the total duration variables led to larger effect sizes with the increase of the total number of hours and the longer the spread. For both of these variables, only at the last level, programs having more than 100 contact hours and ranging over a spread of more than 6 months, was the calculated effect size medium. Inversely, at the lowest level (up to 8 hours and less than a week spread) the effect size was minimal. In the case of the total duration, the group of studies coded as having less than 8 hours had a negative effect size on student learning.

Nevertheless, one unequivocal finding emerges from the analysis of these two variables. The longer the program and its spread throughout the year, the greater its impact on student learning. This result is aligned with previous findings (Lawrenz, 1984, Loucks-Horsley et al., 1998) calling for longer programs with continued support as opposed to one day, one-shot workshops.

Finally, all analyses of possible interactions between the coded variables returned no significant results. Therefore, one should accept that for this sample of studies there is no interaction between the professional development characteristics that were coded.

Findings portrayed in Vignettes

To represent possible applications of these findings three vignettes have been developed. These vignettes are grounded in the findings from the data analysis and are intended to enhance the findings as possible professional development programs with the
characteristics to be more relevant to improve student learning. They are not intended as recipes to be directly applied without adaptation. Professional development is highly dependent on local contexts that should guide the planning and development of each program.

Vignette A – Curriculum Development

Program Technology Improves Physical Science (TIPS) has been established for four years through a district initiative that is funded through a grant. The teachers participating in the program are grouped according to the grade level taught, which includes 3rd, 4th, and 5th grades. There are approximately four teachers in each grade level and they will be meeting throughout the summer and academic year. All participating teachers will experience a minimum of 112 contact hours throughout the year. There will be one 20-hour week workshop at the beginning of the Summer, and another at the end of the Summer. During the academic year, the teachers will meet twice a month for four hours during each meeting.

The first summer week begins with a discussion by the teachers about improving their own curriculum. A common area that emerges in the discussion is the topic of electricity. Most of the teachers feel that the students do not comprehend some of the major concepts affiliated with electricity. After identifying this area, the instructors develop a plan, with the collaboration of the participating teachers, to target the key topics in electricity in the curriculum in the upcoming year. The remainder of the first and
final summer workshop has the teachers exploring electrical concepts in-depth and examining various types of curricular units.

During the academic year, the group focused on developing new lessons in their electricity units. Each lesson focused on a topic in electricity, and utilized technology for the simulation component. Initially, the teachers developed lessons to pilot in their classes. Once a lesson was taught, it was modified and shared with another teacher. Each pair of teachers collaboratively assessed and modified the developed curriculum. During each lesson, they collected data on student performance, and then modified the lesson accordingly. Mid-way through the year, the teachers presented their lessons to the group in the form of a reformed physical science unit. The discussion about the newly developed lessons was lively and energetic. This discussion enabled the teachers to consider their development of the lessons and to contemplate the formulation of more reform-based lessons. Ultimately, the teachers experienced “new” approaches to “old” materials through the creation of curriculum.

In the final week of the TIPS program, the group explored their progress over the year and contemplated what topics they should address in the upcoming year. Their decisions were based upon student learning and their own observations regarding the development of the lessons. The teachers also contributed to the evaluation and reformulation of the TIPS program focused on the improvement of student learning through the experiences of the teachers.
Vignette B – Curriculum Replacement

Program Science Unleashes Nature (SUN) targets teachers from all school levels, and is financed through collaboration between local businesses and a federal grant. Its goal is for the participating teachers to develop replacement curriculum units focusing upon the integration of science technology and society (STS). The reformed curriculum will be used by the teachers in their respective classrooms during the school year. In addition, the teachers also developed new assessment instruments aligned with the newly developed units.

This year the teachers decided to develop replacement units in fluid mechanics. They chose this particular content because they recognized its ability to promote an STS approach and their lack of confidence for teaching it. Program SUN started with a three week summer workshop where the teachers worked for 20 hours a week (a total of 60 hours) to become proficient in the curricula dealing with fluid mechanics and the curricular approach integrating STS. Each teacher worked individually to develop one replacement unit for each semester. They could choose any link between fluid mechanics, technology and society that they judged relevant for their own students. They used dams, bridges, and submarines as meaningful links for the STS approach.

During the first trimester of the school year, each teacher had an opportunity to be visited during the implementation of the unit by one of the members of the professional development team. John Science, a 12th grade physics teacher was visited when he was teaching his new unit dealing with the flight of paper airplanes. He was videotaped
during the implementation of his unit and met with the observer for an hour after each of his classes for debriefing and discussion.

Between the first and second trimester the group met again for two days to share their experiences. They had an opportunity to view each others videotapes, share their knowledge, and receive feedback from the group. During the following trimester this process was repeated. The SUN program encompassed a total of 90 contact hours during eight months.

_Vignette C – Developing Professional Developers_

Program SIR (Scientific Inquiry Rules) is in place in a geographically large state where a large student population is widely dispersed. To address the local needs, the coordinators of program SIR focused on developing professional development teams throughout their state. Using this strategy they initially focused only on a small number of trainers who would then work in their local area, providing professional development to a larger number of teachers.

The goal of the program is the promotion of “Scientific Inquiry” as an approach to teaching and learning science. This content was chosen due to the recommendations present in the National Science Education Standards (NRC, 1996), and their recognition that this approach is still underrepresented in their state. The instructional teams, having elements from very different backgrounds, started by constructing a shared understanding of what “Scientific Inquiry” is, as defined by the National Science Education Standards. They also discussed the qualities of successful professional development programs for
science teachers, and types of activities to promote “Scientific Inquiry”. They used this opportunity to model appropriate professional development strategies by presenting, demonstrating, and supporting teacher learning and change to foster in-depth understanding of the content and pedagogical content knowledge required for effective teaching and learning of students and other educators.

The first phase occurred two weeks at the beginning of summer, meeting at a central location in their state. During these weeks they also initiated an online network to enable them to collaborate, discuss topics of interest, set and pursue common goals, share information and strategies, and identify and address common problems.

Upon completion of the first phase they returned to their own regions where each team immediately started working with a group of 15 to 20 teachers. They worked with these teachers for a total of 120 hours during the next year, partitioned in 40 hours during the summer, 40 hours during the school year, and another 40 hours in the beginning of the following summer. Different regions focused on different content, according to local needs, but they all focused on “Scientific Inquiry” within each particular subject.

During the implementation of their programs, the instructional teams made effective use of the online collaboration that they had created to support each other. This vehicle served as medium to ask questions, share success stories, and motivate each other.

In the first professional development vignette, program TIPS, the teachers are involved in one of the types of treatment that has shown a larger impact on student
learning – curriculum development. Moreover, collaboration, which is another type of activity that significantly impacts student learning, is also emphasized. Student learning is focused as one of the goals of the program and the participating teachers use student data to review and improve the new curricular units that they are developing. The use of technology is present as an additional variable due to its significant impact on student learning. As far as the duration of the program is concerned, it has been presented as having a very big spread, covering the entire academic year, giving continuous support to the teachers and with a total duration over 100 hours.

In the second vignette, Program SUN, the teachers are exposed to the type of treatment that was found to have a larger impact on student learning – curriculum replacement units. At the same time there is a clear focus on the STS theoretical dimension that was also found to have a large impact on student learning. Moreover, there is also an emphasis on the development of new assessment instruments, one of the additional context variables found to have a significant impact on student learning. The duration of this program is also spread over six months and with a total number of contact hours close to 100.

In the third vignette, program SIR, was developed around another type of professional development activity found to have a significant impact on student learning – developing professional developers. What is more, the main goal of this program is the promotion of “scientific inquiry”, as the theoretical dimension. “Scientific inquiry” is the variable in this study that was found to have the largest impact on student learning. Pedagogical content knowledge is another additional variable present in this study also
found to have a significant impact on student learning. This program is also spread throughout the school year and with a total duration over 100 hours.

**Implications**

The purpose of this study was to investigate the effects of professional development on student learning. The results demonstrated a significant effect on student learning as a result of professional development for science teachers. The first important implication of this study is that it is wise to invest in professional development for science teachers. This is aligned with Darling-Hammond’s arguments that teacher education does matter, in this case professional development for science teachers. It somewhat refutes Walsh’s (2001) position that teacher education has no effect on student learning.

Furthermore, the findings will assist present and future professional development providers in determining ways to maximize the impact of their programs on student learning. If the main goal of professional development is to promote and improve student learning, it is strongly suggested that trainers focus on programs with specific characteristics (see table 4.17), such as programs centered on activities dealing with science curricula. Science curriculum implementation, development, and replacement were found to be three of the most effective ways of improving student learning in science through professional development. Two other treatments that were found to be correlated with student learning were developing professional developers and creating professional networks enhance collaboration and partnerships among science teachers.
Focusing the content of the teacher development program on “Science as Inquiry” was found to be a successful way to promote student learning. Two other effective ways include the promotion of the link between Science-Technology-Society and focusing on Unifying concepts.

Among the other variables that emerged as significantly correlated with student learning are pedagogical content knowledge (Shulman, 1986), testing and assessment and focusing on the National Science Education Standards (NRC, 1996). All demonstrated above average impact on student learning

Finally, the programs should be sustained throughout the year. The group of programs that the teachers participated in for a period of more than six months achieved medium effect size. Also the total duration of such programs should be at least 100 hours.

The conceptual framework for this study centered on the idea that student learning should be the main focus of professional development; however, there are many additional goals of professional development other than student learning. Professional development influences other critical issues such as “changes in teacher knowledge and practice, implementation of new programs, changes in school culture, and development of teachers’ leadership abilities” (Loucks-Horsley & Matsumoto, 1999, p. 259). The impact of professional development on these critical issues should also be considered as well as the impact of these issues on student learning. This is a challenging task. For example, improved teacher knowledge should be assessed and related to student learning; however, professional development may have minimal impact on other issues such as changes in school culture. Nonetheless, even the minimal impact of professional
development, on other variables may have had an effect on student learning, but they were not detected in this study. Some of the strategies employed by professional development programs are embedded within them because philosophically they are part of the core values to promote change, evolution and, ultimately, student learning. Issues and challenges dealing with school culture and teacher leadership will always be addressed despite the difficulty of linking them directly to improvement of student learning. For political, economic, and social reasons, a market value based approach has often been the driving force in professional development programs, often hindering their ultimate success. Frequently commendable initiatives in the schools are eliminated because their importance cannot be substantiated with hard data. In the long run it may be advisable to invest in professional development utilizing those characteristics of professional development that have the potential to significantly enhance student learning even though the “return on investment” may require a longer period of time to reach fruition.

In addition to student learning as an important outcome of professional development, these programs must also include intensive and extensive follow-up activities. If the main goal of professional development is to improve student learning through teachers’ improved expertise, and if the change process is as complex as it has been presented, then short-term professional development programs are limited in their effectiveness. Continuous interactions with teachers are essential not only to nurture their change and evolution process, but also to collect valuable feedback from them and from their students to further improve professional development programs.
Further research

This study suggests several areas of possible further research. Many questions remain unanswered and new ones have emerged.

The first issue is how the findings from this study can be translated into practice in a real professional development program. Several key characteristics of a program to promote student learning have been identified but it is not clear how they interact in a real situation. It would be helpful and interesting to conduct a case study with a program identified as successful under this framework. This would enable the researcher to become familiar with the interactions behind the program that enable it to be successful. Such a closer inspection of the variables is not possible using a meta-analysis.

Another issue that is unclear deals with the proposed theoretical framework. This study used the framework to investigate the interaction between the three domains to determine how professional development impacts student learning. But one may ask what occurs in the teachers’ domain? How are teachers impacted by the professional development and how do they translate it into their own classrooms? It is necessary to further investigate the process of change that occurs in teachers as a consequence of professional development. Understanding this process will assist in maximizing its effectiveness.

Another important issue is the alignment among professional development program objectives, teacher change, and student learning. It is important to understand the perceptions of teachers and trainers concerning the professional development program
in which they are involved. Moreover, it is necessary to verify how the program objectives translate into the teachers’ classroom and ultimately the students’ learning.

Several questions remain unanswered concerning the issues of time. How much time is sufficient? How much spread for a certain duration of time? The data analysis showed that the groups involved in professional development with more than 100 total hours and more than six months of continuous support were effective. The question remains, however, as to what “point in time” during the professional development does it become effective. The issue of time is not specific to any one particular discipline. Another meta-analysis with professional development programs from a variety of disciplines focused on the time variable is desirable. A larger number of studies would permit grouping into smaller intervals and the use of a weighted least squares regression approach to find the time and spread that maximize program effectiveness, making the findings more generalizable to other disciplines.

Professional development is always context dependent and no findings or recommendations should be applied literally. Professional development programs must be adapted to every particular situation. In this study, for example, programs with more than 100 contact hours were found to be most effective. This finding indicates that professional development providers should carefully consider duration and spread if one of the intended outcomes is improved student learning.
Conclusions

Research can and should be conducted to evaluate the impact of professional development on student learning. However, most professional development programs do not take that step. They only focus on the teacher as the outcome measure. It is imperative that professional development programs make a concerted effort to evaluate their impact on students. As Hein (1997) proposes, student learning should always be seen as the ultimate goal of any educational process. Therefore, any professional development program should be driven, informed, and restructured according to student’s data. This is achieved by focusing the evaluation of the effectiveness of professional development programs at the most complex level, student learning, as suggested by Guskey (1998). A program can only be considered successful if it is able to document its impact on the intended outcome, and that outcome should ultimately improve student learning. For that reason, all professional development programs should include in their evaluation a research component looking at their impact on student learning. Only if research on professional development includes sufficient detail about its impact on students can one hope to use it to improve and design future professional development programs.

The findings of this study establish a clear relationship between professional development for science teachers and improved student learning. A framework for professional development was created that centers on student learning, and the results supported this relationship. This focus for professional development programs is supported by the empirical evidence in this study. Professional development for science
teachers positively impacts their student learning in science. However, there are a multitude of different program characteristics that lead to different impacts on students. Therefore, if one is serious about promoting student learning through professional development the program design must be carefully chosen to include those characteristics that will maximize its impact on students.

This study analyzed many program characteristics and the results revealed varying effects. Some types of professional development activities do not impact student learning. Those are the cases of “Action Research” and “Study Groups”. If student learning is the main goal of professional development than these types of activities should not be chosen. They may be effective and even recommendable in a particular case where another specific outcome is desired, but if student learning is the main focus these should not be the types of activities made available for teachers.

If the main outcome of a professional development program for science teachers is to enhance student learning, the characteristics of the program should be chosen from those identified in this study as being effective (see table 4.17). These include activities dealing with curriculum development, replacement, and implementation; developing partnerships and collaboration; and developing other professional developers.

Specific theoretical domains should also be targeted, in particular, scientific inquiry, STS and unifying concepts. All of these domains were found to have a large impact on student learning. Finally, professional development must be sustained through time. One day workshops do not have a significant impact on student learning. If one is serious about professional development a concerned policy effort must be made to
provide teachers with a significant number of contact hours sustained throughout the school year.
Appendix A

Coding Form
CODING FORM

Reference information

<table>
<thead>
<tr>
<th>STUDY LEVEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Study ID [STUDYID]</td>
</tr>
</tbody>
</table>

Sample Characteristics

<table>
<thead>
<tr>
<th>Sample Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Grade Level [GRADLEV]</td>
</tr>
<tr>
<td>1. Elem</td>
</tr>
<tr>
<td>2. MS</td>
</tr>
<tr>
<td>3. HS</td>
</tr>
<tr>
<td>0. NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SES [SES]</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. SES</td>
</tr>
<tr>
<td>1. Low</td>
</tr>
<tr>
<td>2. Med</td>
</tr>
<tr>
<td>3. High</td>
</tr>
<tr>
<td>0. NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ability level [ABILEV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Ability level</td>
</tr>
<tr>
<td>1. Low</td>
</tr>
<tr>
<td>2. Normal</td>
</tr>
<tr>
<td>3. Gift &amp; Tal</td>
</tr>
<tr>
<td>0. NA</td>
</tr>
</tbody>
</table>

School Characteristics

<table>
<thead>
<tr>
<th>Size [SIZE]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Size</td>
</tr>
<tr>
<td>1. 0-500 students</td>
</tr>
<tr>
<td>2. 500-1000</td>
</tr>
<tr>
<td>3. &gt;1000</td>
</tr>
<tr>
<td>0. NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type [TYPE]</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. type</td>
</tr>
<tr>
<td>1. Public</td>
</tr>
<tr>
<td>2. Private</td>
</tr>
<tr>
<td>0. NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location [LOCAT]</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. location</td>
</tr>
<tr>
<td>1. Rural</td>
</tr>
<tr>
<td>2. Urban</td>
</tr>
<tr>
<td>0. NA</td>
</tr>
</tbody>
</table>

Treatment Characteristics

<table>
<thead>
<tr>
<th>Type of treatment (choose all that apply) [TYPETREAT]</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Type of treatment</td>
</tr>
<tr>
<td>1. Workshop</td>
</tr>
<tr>
<td>2. Curric Dev</td>
</tr>
<tr>
<td>3. Curr Replac</td>
</tr>
<tr>
<td>4. Curr Impl</td>
</tr>
<tr>
<td>5. Inquiry</td>
</tr>
<tr>
<td>6. Action Research</td>
</tr>
<tr>
<td>7. Case disc</td>
</tr>
<tr>
<td>8. Developing PD</td>
</tr>
<tr>
<td>9. Study groups</td>
</tr>
<tr>
<td>10. Coaching and Ment</td>
</tr>
<tr>
<td>11. Partnerships</td>
</tr>
<tr>
<td>12. Examin. Student work</td>
</tr>
<tr>
<td>13. Immersion work scientists</td>
</tr>
<tr>
<td>14. Partnerships with scientists</td>
</tr>
<tr>
<td>15. Tech for PD</td>
</tr>
</tbody>
</table>

120
9. Theoretical dimensions (choose all that apply) [THEODIM]
   1. Science as inquiry
   2. Physical science
   3. Life science
   4. Earth and space
   5. Science and technology
   6. Science and society
   7. HNOS
   8. Unifying concepts

10. Additional context variables (choose all that apply) [ADDCONTVAR]
   1. Specific content
   2. Pedagogy
   3. PCK
   4. Technology
   5. Testing, assessment
   6. Standards
   7. Other (specify):

11. Housing [HOUSE]
   1. University
   2. School district
   3. School
   0. NA

12. Funding per teacher [FUND]

13. Initial Needs Assessment [NEEDASSES]
   1. Yes
   2. No
   0. NA

Teacher characteristics

14. Years of experience (dominant group) [YRSEXP]
   1. 0-5 years
   2. 5-10 yrs
   3. +10yrs
   0. NA

15. Education background (dominant group) [EDBACK]
   1. Bachelors in Ed
   2. Alt Cert
   3. Masters
   0. NA

16. Gender [GENDER]
   (%Female)
   0. NA

Design Characteristics

17. Unit of Analysis [UNITAN]
   1. Student
   2. Class
   3. Teacher
   4. School
   0. NA

18. Control for pre-existing differences [PREDIF]
   1. Yes
   2. No
   0. NA
19. Experimental mortality [EXPMORT]
   (%)  
   0 NA

20 A). duration of treatment (spread) [SPREAD]
   1 up to 1 week
   2 up to 1 month
   3 up to 6 months
   4 up to 1 year
   0 NA

20 B). duration of treatment (total) [TOTALDUR]
   1 8hrs
   2 9-16 hrs
   3 17-50hrs
   4 51-100hrs
   5 over 100hrs
   0 NA

Outcome characteristics

21. Type of criterion or outcome construct [OUTCOM]
   1 Achievement
   2 Attitudes

22. Congruence of treatments and outcome measure [CONG]
   1 yes
   2 no

23. Method of measurement [METHMES]
   1 Stand test
   2 Specific test
EFFECT SIZE LEVEL

Effect size information

24. Type of data effect size based on [ESTYPE]
   1 means and standard deviations
   2 t- or F-value
   3 chi-square
   4 frequencies or proportions, dichotomous
   5 frequencies or proportions, polychotomous
   6 other (specify):

25. Direction of effect [DIREFECT]
   1 treatment group
   2 neither
   3 control group
   0 cannot tell or statistically insignificant report

Sample size

26. treatment group sample size [NTREAT]
27. control group sample size [NCONTROL]

Means and standard deviations

28. treatment group mean [TREATMEAN]
29. control group mean [CONTROLMEAN]
30. treatment group standard deviation [TREATSD]
31. control group standard deviation [CONTROLSD]

Significance tests

32. t-value [TVAL]
33. F-value [FVAL]
34. Chi-square value [CHISQVAL]

Calculated Effect size

35. Effect size [ES]
36. Confidence rating in effect size computation [ESCONF]
   1 highly estimated
   2 moderate estimation
   3 some estimation
   4 slight estimation
   5 no estimation
Appendix B

Example of coding for the effect size level
EFFECT SIZE LEVEL

Effect size information

24. Type of data effect size based on [ESTYPE]
   1 means and standard deviations
   2 t- or F-value
   3 chi-square
   4 frequencies or proportions, dichotomous
   5 frequencies or proportions, polychotomous
   6 other (specify):

25. Direction of effect [DIREFECT]
   1 treatment group
   2 neither
   3 control group
   0 cannot tell or statistically insignificant report

Sample size

26. treatment group sample size [NTREAT]
   108
   27. control group sample size [NCONTROL]
   153

Means and standard deviations

28. treatment group mean [TREATMEAN]
   16.3
   29. control group mean [CONTROLMEAN]
   14.4
   30. treatment group standard deviation [TREATSD]
   5.2
   31. control group standard deviation [CONTROLSD]
   5.6

Significance tests

32. t-value [TVAL]
33. F-value [FVAL]
34. Chi-square value [CHISQVAL]

Calculated Effect size

35. Effect size [ES]
   0.352
   36. Confidence rating in effect size computation [ESCONF]
      1 highly estimated
      2 moderate estimation
      3 some estimation
      4 slight estimation
      5 no estimation
Appendix C

Sample Vignette
Computers in Geometry Class


A group of high school mathematics teachers has been meeting twice a month at their school for a seminar with mathematicians and mathematics educators from a nearby university. These teachers have been using computers in their geometry classes for the past year and a half, and the seminar provides them with opportunities to discuss what is happening in their classroom as they think about new ways of teaching and learning.

The computer software allows students and their teachers to construct geometric shapes and to make measurements of lengths and angles and computations based on these measurements, thus providing an environment for open-ended exploration and discovery of patterns and relationships.

Although the teachers have been excited about their use of computers in geometry, many have voiced frustration in trying to make decisions about appropriate tasks for their students. Some teachers have been most comfortable focusing student attention on specific relationships, while others are dissatisfied with activities structured to lead students toward a particular “discovery.” At times, many teachers have felt their own knowledge of geometry inadequate to deal with questions and conjectures that arise from open-ended explorations.

Gloria described a task she assigned her class early in the year:

I wanted my students to learn that the sum of the angle measurements in a triangle is 180 degrees, so I had them construct a lot of triangles on the computer and record the angle measures. The software made it possible to collect a lot of data quickly and make a generalization. I thought my students would remember the relationship better if they discovered it themselves.
Rich talked about the same task:

I was really reluctant to use that activity because it didn’t seem like exploration. It made me feel that I would be directing the students toward a single result and not really taking advantage of the technology. But when Gloria told me about some of the things her students came up with, I thought it might lead in some interesting directions. I was amazed at what happened. My students didn’t just see what I thought they would see; many of them went off in all sorts of directions exploring other shapes. One even asked about a circle! I wasn’t quite sure where to go with that question, but it certainly seemed intriguing—and it took us into lots of other ideas when we discussed it in the seminar.

Constanza remembered a lesson that was especially important to her:

One of my students had constructed a shape on the computer screen that he said looked three-dimensional. We took off on a discussion of geometric models and representations of shapes, something I hadn’t really expected them to get in that lesson. As we were talking about two- and three-dimensional shapes, Jan asked about a line. Well, a lot of the students thought that was boring, but Raoul held up a paper clip and said he thought it was two-dimensional. And another student said that if you bent the paper clip it would be three-dimensional.

That sent off a bunch of conjectures, with the students coming up with good reasons for why the bent paper clip could be considered one-, two-, or three-dimensions. There was a lot more there than I had anticipated, and I thought it would be a great topic for discussion in the seminar. It made us think a lot about representations and how we describe and define geometric shapes.

The seminar has been a place where teachers can share their struggles with their colleagues and university faculty and develop meaningful activities for their students. For
many of the teachers, one of the most valuable aspects of the seminar has been the opportunity to extend their own understanding of geometry.
Appendix D

SPSS Macro for meta-analysis
*--------------------------------------------------------------------------
* Macro for SPSS/Win Version 6.1 or Higher
* Written by David B. Wilson (dwilson@crim.umd.edu)
* Meta-Analyzes Any Type of Effect Size
* To use, initialize macro with the include statement:
* INCLUDE "[drive][path]MEANES.SPS".
* Syntax for macro:
* MEANES ES=varname /W=varname /PRINT=option.
* E.g., MEANES ES=D /W=IVWEIGHT.
* In this example, D is the name of the effect size variable
* and IVWEIGHT is the name of the inverse variance weight
* variable. Replace D and IVWEIGHT with the appropriate
* variable names for your data set.
* /PRINT has the options "EXP" and "IVZR". The former
* prints the exponent of the results (odds-ratios) and
* the latter prints the inverse Zr transform of the
* results. If the /PRINT statement is omitted, the
* results are printed in their raw form.
*--------------------------------------------------------------------------
preserve
set printback=off
define meanes (es=!charend('/') /w=!charend('/')
/print = !default('RAW') !charend('/'))
preserve
set printback=off mprint=off
*--------------------------------------------------------------------------
* Enter matrix mode and get data from active file
*--------------------------------------------------------------------------
matrix
get x /file * /variables = !es !w /missing omit
*--------------------------------------------------------------------------
* Compute variables needed to calculate results
*--------------------------------------------------------------------------
compute k = nrow(x).
compute es = make(k,1,-99).
compute es(1:k,1) = x(1:k,1).
compute w = make(k,1,-99).
compute w(1:k,1) = x(1:k,2).
release x.
*--------------------------------------------------------------------------
* Compute random effect variance component and new weight
*--------------------------------------------------------------------------
compute c = ((csum((es&**2)&*w)-csum(es&*w)**2/csum(w))-(k-1))
/(csum(w)-csum(w&**2)/csum(w)).
do if (c < 0).
.compute c = 0.
.end if.
compute w_re = 1/(c + (1/w)).
*--------------------------------------------------------------------------
* Calculate summary statistics
*-------------------------------------------------------------
compute df = k - 1.
compute mes = csum(es*^w) / csum(w).
compute mes_re = csum(es*^w_re)/csum(w_re).
compute sem = sqrt(1/csum(w)).
compute semre = sqrt(1/csum(w_re)).
compute les = mes - 1.95996*sem.
compute ues = mes + 1.95996*sem.
compute les_re = mes_re - 1.95996*semre.
compute ues_re = mes_re + 1.95996*semre.
compute q = csum((es**2)*w)-csum(es*^w)**2/csum(w).
do if (df>0).
  compute p = 1- chicdf(q,df).
end if.
compute z = mes/sem.
compute z_re = mes_re/semre.
compute pz = (1-cdfnorm(abs(z)))*2.
compute pz_re = (1-cdfnorm(abs(z_re)))*2.
compute sd = sqrt(q*csum(w)**-1).
*-------------------------------------------------------------
* Transform Output if Requested
*-------------------------------------------------------------
!IF (!print !eq 'EXP'|!print !eq 'exp'|!print !eq 'Exp') !THEN.
compute mes = exp(mes).
compute les = exp(les).
compute ues = exp(ues).
compute mes_re = exp(mes_re).
compute les_re = exp(les_re).
compute ues_re = exp(ues_re).
compute sem = -9.9999.
compute semre = -9.9999.
!IFEND.

!IF (!print !eq 'IVZR'|!print !eq 'ivzr'|!print !eq 'Ivzr'
    |!print !eq 'IvZr') !THEN.
compute mes = (exp(2*mes)-1)/(exp(2*mes)+1).
compute les = (exp(2*les)-1)/(exp(2*les)+1).
compute ues = (exp(2*ues)-1)/(exp(2*ues)+1).
compute mes_re = (exp(2*mes_re)-1)/(exp(2*mes_re)+1).
compute les_re = (exp(2*les_re)-1)/(exp(2*les_re)+1).
compute ues_re = (exp(2*ues_re)-1)/(exp(2*ues_re)+1).
compute sem = -9.9999.
compute semre = -9.9999.
!IFEND.

*-------------------------------------------------------------
* Create Output Matrices
*-------------------------------------------------------------
compute table1 = make(1,4,-99).
compute table1(1,1) = k.
compute table1(1,2) = mmin(es).
compute table1(1,3) = mmax(es).
compute table1(1,4) = sd.

compute table2 = make(2,6,-99).
compute table2(1,1) = mes.
compute table2(1,2) = les.
compute table2(1,3) = ues.
compute table2(1,4) = sem.
compute table2(1,5) = z.
compute table2(1,6) = pz.
compute table2(2,1) = mes_re.
compute table2(2,2) = les_re.
compute table2(2,3) = ues_re.
compute table2(2,4) = semre.
compute table2(2,5) = z_re.
compute table2(2,6) = pz_re.

compute table3 = make(1,3,-99).
compute table3(1,1) = q.
compute table3(1,2) = df.
compute table3(1,3) = p.

*-------------------------------------------------------------
* Print summary statistics
*-------------------------------------------------------------
print /title "***** Meta-Analytic Results *****".
print table1
/title '------- Distribution Description' +
' ---------------------------------'
/clabel "N" "Min ES" "Max ES" "Wghtd SD"
/format f11.3.
print table2
/title '------- Fixed & Random Effects Model' +
' -----------------------------'
/clabel "Mean ES" "-95%CI" "+95%CI" "SE" "Z" "P"
/rlabel "Fixed" "Random"
/format f9.4.
print c
/title '------- Random Effects Variance Component' +
' ------------------------'
/rlabel 'v' = '/format f10.6.
print table3
/title '------- Homogeneity Analysis' +
' -------------------------------------'
/clabel "Q" "df" "p"
/format f11.4.
print
/title 'Random effects v estimated via noniterative' +
' method of moments.'.
!IF (!print !eq 'EXP'|!print !eq 'exp'|!print !eq 'Exp') !THEN .
print
/title 'Mean ES and 95% CI are the exponent of the' +
' computed values (Odds-Ratios).'.
(Lipsey and Wilson, 2001)
Bibliography


American Association for the Advancement of Science (1989). *Science for all Americans: A Project 2061 report on literacy goals in science, mathematics, and technology*. Washington, DC.


VITA

Luis Fonseca Tinoca was born in Lisbon, Portugal on January 18, 1975. He is the son of António Cândido da Silva Tinoca and Maria Benedita Pereira da Fonseca Tinoca. After completing his high school work at the Colégio Militar in Lisbon in 1993, he entered the Instituto Superior Técnico da Universidade Técnica de Lisboa. In 1995 he transferred to the Faculdade de Ciências da Universidade de Lisboa where he received the degree of licentiate in Physics and Chemistry education in 1999.

In 1998, Luis started working as a physics and chemistry teacher at the Fernando Lopes Graça High School in Parede, Portugal. In 1999 he was hired as a physics and chemistry teacher by the Dom Domingos Jardo Middle School in Mira Sintra, Portugal. In 2000 he entered the Graduate School of The University of Texas at Austin to pursue his Doctorate of Philosophy in Science Education.

In 2000 he married to Claudia Ramos Tinoca of Nampula, Mozambique. One son, Rafael Ramos Tinoca was born in August of 2001, and one daughter, Carolina Ramos Tinoca was born in November of 2003, both in Austin, Texas.

Permanent Address: Estrada A-da-Maia, n50, 4E, 1500-004 Lisboa, Portugal

This dissertation was typed by the author.