

EDUCATION POLICY ANALYSIS ARCHIVES

A peer-reviewed scholarly journal

Editor: Sherman Dorn

College of Education

University of South Florida

Volume 14 Number 5

February 23, 2006

ISSN 1068-2341

National Board Certification as Professional Development: What Are Teachers Learning?

David Lustick
University of Massachusetts Lowell

Gary Sykes
Michigan State University

Citation: Lustick, D., & Sykes, G. (2006). National board certification as professional development: What are teachers learning? *Education Policy Analysis Archives*, 14(5). Retrieved [date] from <http://epaa.asu.edu/epaa/v14n5/>.

Abstract

This study investigated the National Board for Professional Teaching Standards' (NBPTS) assessment process in order to identify, quantify, and substantiate learning outcomes from the participants. One hundred and twenty candidates for the Adolescent and Young Adult Science (AYA Science) Certificate were studied over a two-year period using the recurrent institutional cycle research design. This quasi-experimental methodology allowed for the collection of both cross-sectional and longitudinal data insuring a good measure of internal validity regarding observed changes between individual and across group means. Transcripts of structured interviews with each teacher were scored by multiple assessors according to the 13 standards of NBPTS' framework for accomplished science teaching. These scores provided the quantitative evidence of teacher learning in this study. Significant pre-intervention to post-intervention changes to these individual and group means are reported as learning outcomes from the assessment process. Findings suggest that the intervention had significant impact upon candidates' understanding of knowledge associated with science teaching with an overall effect



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-nd/2.5/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published jointly by the Colleges of Education at Arizona State University and the University of South Florida. Articles are indexed by H.W. Wilson & Co. Send commentary to Casey Cobb (casey.cobb@uconn.edu) and errata notes to Sherman Dorn (epaa-editor@shermamdorn.com).

size of 0.47. Standards associated with greatest gains include Scientific Inquiry and Assessment. The results support the claim that the certification process is an effective standards based professional learning opportunity comparable to other human improvement interventions from related domains. Drawing on qualitative data, we also explore three possible implications of teacher learning outcomes from certification upon classroom practice identified as Dynamic, Technical, and Deferred. These patterns suggest that more than one kind of learning may be taking place in relation to board certification. The discussion then considers the importance of this study for policy making and science teaching communities. Keywords: Teacher Learning; Professional Development; Certification; Science Education; National Board for Professional Teaching Standards (NBPTS).

The National Board and Public Policy

Evidence has begun to accumulate that demonstrates a relationship between National Board certification and student achievement (see Cavaluzzo, 2004; Goldhaber & Anthony, 2004; Vandevoort, Amrein-Beardsley, & Berliner, 2004). This relationship is relevant to policy because so many states and localities are providing financial support and incentives to encourage teachers to become Board certified. Whether such teachers in fact are “highly accomplished” as determined by their ability to promote higher student achievement is a matter of great interest at the moment. But if Board certified teachers are unusually capable, why may this be so? One answer is self-selection. Teachers who volunteer to undertake Board certification are superior teachers to begin with. In this case, certification is primarily a selection mechanism. Another answer is that the certification process itself constitutes a form of professional development that actually enhances teacher knowledge, skills, and dispositions in candidates regardless of whether or not they achieve certification. In this case, certification is a development process.

Between these two possibilities, the former is most likely, because certification is a relatively brief “treatment.” Although intensive and time-consuming, extending over a year of work, the certification process primarily calls on teachers to document their practice. Consequently, those who volunteer are most likely quite good teachers to begin with. Still, teachers prepare in study groups, learn from the materials they receive, and from the processes they undergo. Many teachers who are unsuccessful at achieving certification during the first attempt may re-take portions of the examination, so the process can extend over more than one year.¹ In light of the importance of Board certification as a significant policy effort to enhance the quality of teachers, the question of what teachers learn through the process is noteworthy. The study reported here supplies some preliminary evidence on this question.

The context for an inquiry into Board certification is the contemporary debate on professionalization as a policy choice. The heart of the professional premise is that teachers utilize expert knowledge in their work that may be codified and transmitted to practitioners. Such codification occurs in a number of places including the curriculum of preservice training, continuing education, and the standards and assessments that are used to select and screen entrants to the profession. As the National Commission on Teaching and America’s Future (1996) argued, the basis for professional standards today is a three legged stool that includes certification standards for entry under the jurisdiction of states, standards for teacher preparation programs established by NCATE;

¹ For this investigation, only first time candidates were included in the population pool. All retake candidates were removed from consideration.

and advanced standards for professional practice recently established by the National Board for Professional Teaching Standards (NBPTS). This model is represented in varying ways in most professions but is most highly elaborated in the medical field where advanced standards for practice are associated with specialization and have been developed by the various medical specialty boards.

Standards are one form for the representation and transmission of professional knowledge. Such knowledge is typically validated in relation to its efficacy in producing desirable outcomes, but due to the inevitably incomplete nature of professional knowledge, standards are the creation of consensus panels made up of experts in the relevant fields who draw on research based knowledge but also utilize judgment in formulating the standards. In this respect, the NBPTS has followed accepted practice in developing its specialized standards for each certificate area based on the work of consensus panels. Still, the work of validating professional standards is an ongoing process, and the present study contributes to this effort.

The policy of professionalization, as it enlists the resources and authority of the state is under challenge, however, from those who doubt that teaching involves codified expert knowledge that may be represented in various forms and used to discipline preparation, entry, evaluation, and advancement. Arguments made by opponents tend to stress that teachers employ ordinary knowledge and intelligence, typically acquired in university liberal arts programs and that screening for entry should involve just tests of basic skills, general aptitude, and knowledge of relevant subject matter.

In our society today, both views are advanced in a policy climate of skepticism about the quality of teaching, and these contending positions suggest quite different policy approaches to the enhancement of teaching. The first or professional view places more emphasis on the steady development of validated standards that underlie the “three legged stool.” The second or anti-professional approach places reliance on recruitment strategies that open teaching to applicants with diverse backgrounds and qualifications based on indicators of intelligence and general learning plus some modest ‘how to’ knowledge and practical experience.

The present study enters this policy debate by presenting evidence about what teachers are learning from the National Board certification process. Along the lines of the debate just indicated are two views. One suggests that teachers learn to be more reflective practitioners as a result of the process, supporting the professional claim. The other position argues that teachers simply learn how to master the assessment process in order to gain the incentives that states and districts are beginning to provide, such as additional pay. If the first view has force, then states and districts may be justified in providing public incentives. If the second view has force, then such allocations are unlikely to be worthwhile.

Framing the policy issue in these terms however only meets the threshold condition that teachers are learning something of value. Beyond the threshold are further questions about whether they use what they have learned in their practice and even further whether such use enhances student learning. Still, if the threshold condition is not met then further inquiry will be moot, as teachers are not undertaking any useful learning in the first place. So the present inquiry aims to provide the first test of this proposition—that board certification promotes useful learning among candidates and so is a worthwhile policy investment.

The National Board for Professional Teaching Standards has been setting standards for accomplished teaching and certifying teachers who measure up to those standards since 1993. During the past decade, the NBPTS has developed 27 areas of certification and has awarded certificates to more than 40,000 teachers (NBPTS, 2005). The certification process is rigorous; only one-half of applicants are successful. Yet, evidence suggests that candidates whether they pass or not, find the experience valuable to their professional growth. Candidates must complete an extensive portfolio that profiles their work with students, school, and community. In addition,

candidates take a computerized assessment that evaluates their content knowledge in their area of expertise. Together, the portfolio and the exam constitute the foundation of the certification experience.

Certification costs \$2300 per teacher, and as well as defraying application fees many states and local districts offer additional and often generous financial incentives to encourage teachers to become certified. All told, 48 states and 544 local districts offer some form of financial incentive or support for National Board certification. For example, North Carolina set aside more than \$26 million in 2003 to encourage and support teachers who pursue certification (Leef, 2003). There, incentives can add up to \$50,000 per teacher over the life of the 10-year certificate (NBPTS, 2004a). Nationwide, annual expenditures are substantial. In 2003, for example, approximately 16,000 teachers pursued certification for a total cost \$36.8 million that was paid almost entirely by state departments of education, local districts, and to a lesser extent, teacher unions. Eight of sixteen thousand achieved certification in that year (NBPTS, 2004a). If 75% of those who pass received a bonus, financial reward, or salary increase equivalent to \$2500², then an additional outlay of \$20 million needs to be added to the total public expenditures. For 2003 alone, taxpayers invested nearly \$57 million in National Board certification. This is a considerable sum, although only a miniscule portion of all funding for teacher professional development.³ In fiscally difficult times, states and local districts are now debating the merits of providing financial incentives and support for National Board certification (Griffin, 2003). Consequently, evidence on the effectiveness of this intervention is salient.

With growing numbers of teachers pursuing National Board certification and with increasing amounts of public dollars being used to fund and encourage the process, what kind of impact is certification having upon teacher quality? Does National Board certification provide teachers with an effective professional development? In other words, does the Board certification process provide opportunities for teachers to learn new knowledge and skills relevant to their work with students? Recent studies indicate that National Board teachers facilitate greater student achievement in their students (Goldhaber & Anthony, 2004 Vandevoort et al, 2004). However, very little quantitative evidence exists that indicate how and to what extent the certification process improves teacher quality.

The ‘What teachers are learning?’ question remains the least understood aspect of the professional development paradigm. According to Wilson and Berne (1999), research in this area has yet to “identify, conceptualize, and assess” what teachers are learning. Program evaluation research, in such professional development initiatives as the Eisenhower Professional Development Program for Math and Science Teachers, which focused upon teacher learning produced vague answers due primarily to limited methodologies such as surveys and self-reports (Mullens, Leighton, Laguarda, & O'Brien, 1996). Strikingly missing from the literature are empirical studies that address the questions surrounding professional development, teacher learning, and the impact upon teacher quality.

The study reported here makes a contribution to these issues, based on the use of a quasi-experimental design with several cohorts of candidates. While prior research has supplied evidence on the validity of the certification process (see Bond, Smith, Baker, & Hattie, 2000) and on its

² Financial incentives range from \$1000 to \$7500 annually for the life of the certificate (10 years). States like North Carolina offer a 12% increase in salary for the life of the certificate with successful completion of the process. For a complete review of the incentives and support offered by each state and more than 500 local districts, visit <http://www.nbpts.org/about/state.cfm>.

³ Determining the annual expenditures for professional development is a tricky and uncertain process. For detailed studies on how estimates are calculated, see Killeen, et al. (2002), and Odden, et al., (2002).

association with student outcomes, the current study takes up the question of what teachers might be learning from the process itself. If professional development is accepted as a primary means of improving student learning, then it becomes important to understand what teachers may or may not be learning from a specific professional development intervention. Before teachers can improve their work with students, they first must acquire new knowledge and skills to employ in the classroom. The results of this study present a systematic analysis of what teachers are learning from a specific intervention and how that learning might influence the quality of instruction. For the policy community, this study provides valuable quantitative knowledge that describes a learning opportunity for science teachers that may have dramatic impact upon student achievement and the learning experience.⁴

Description of the Intervention

The National Board for Professional Teaching Standards was established in 1987 through a grant from the Carnegie Corporation of New York as a means of defining, assessing, and recognizing accomplished teaching (NBPTS, 1991). The NBPTS has identified three critical aspects of certification: *standards*, establishing, reviewing, and refining standards of accomplished teaching through consensus about what teachers should know and be able to do; *assessment*, providing a valid and accessible means to evaluate teachers against the standards; and *professional development*, providing teachers with the opportunity to strengthen their practice through self-examination (Koprowicz, 1994). All standards, assessments, and scoring rubrics are based upon the five ‘core propositions of accomplished teaching’ that the Board developed and disseminated.⁵

The year long certification process for teachers has two main components: the construction of a detailed, reflective, and analytic portfolio over a four to six month span; and the completion of a content focused four hour computerized assessment.⁶ The portfolio for Adolescent and Young Adult Science has four sections that address the thirteen standards for AYA Science: Teaching a Major Idea in Science, Active Scientific Inquiry, Whole Class Discussion in Science, and Documented Accomplishments: Contributions to Student Learning.⁷ (For a description of these

⁴ Studies of effectiveness do not settle the matter. Cost-effectiveness studies are needed, that compare various kinds of professional development. Nevertheless, as a first order of business, inquiry into the effectiveness of professional development “treatments” is worthwhile. For further commentary, see Borko (2004).

⁵ The five core propositions state that

- 1) Teachers are committed to students and their learning.
- 2) Teachers know the subjects they teach and how to teach those subjects to students.
- 3) Teachers are responsible for managing and monitoring student learning.
- 4) Teachers think systematically about their practice and learn from experience.
- 5) Teachers are members of learning communities. (NBPTS, 1991 p. 13-14)

⁶ The assessment center ‘exercises’ have changed over the course of the last few years. Originally for example, the exam for AYA Science took two 8 hour days covering post-secondary level content material in science and pedagogical knowledge in the science classroom. Today, the exam has been reduced to a 4-hour session focused entirely on science content. For a complete description, go to: <http://www.nbpts.org/candidates/acob/nextgen/n20.html>, (NBPTS, 2004b).

⁷ It should be noted that starting in 2001, the ‘new’ format for portfolio construction was phased in over a two year period. The original portfolio required 6 entries. An Entry called “Assessment” was merged with the three other classroom-oriented entries and the Documented Accomplishments, Professional and Community were combined. Both formats were involved in this study, though which candidates had which form is unknown. The vast majority had to complete the new 4 entry version. Though the use of two

standards, see Table 3.) Using videotape, examples of student work, and artifacts representing professional accomplishments, teachers address questions in each section of the portfolio while constructing a presentation of their best practice⁸. The final product serves as evidence demonstrating the teacher's impact on classroom academic environment, student learning, and the school community. In this study, the identified components of certification also served as the curriculum and resources for teacher learning.

National Board Certification as Professional Development

The perception of National Board Certification (NBC) as a productive professional development may be due in no small part to the numerous endorsements received by a wide range of organizations. One such endorsement is from the Center for Research on Education, Diversity & Excellence:

We believe the process (NBC) represents sound professional development practice—it is focused on subject matter content and student learning, uses teacher self-reflection and inquiry linked to the teacher's own teaching situation and practice, and is highly collaborative. This kind of thorough, focused professional development is far too rare for most of California's teachers. (CREDE, 2003, p. 10)

How does National Board certification fit with current conceptions of effective professional development? The answer can be found by understanding the characteristics of quality professional development.

Over the last decade, ideas about what constitutes effective professional development for teachers have been changing (Little, 1993 & 1997; Ball & Cohen, 1995; Huberman 1993; Hargraves 1995; Darling-Hammond & McLaughlin, 1996; Stein and Brown 1997; Sykes, 1999). Such reexamination of professional development has been motivated by a pervasive dissatisfaction with traditional professional development, whose features are well known. Hawley and Valli (1999, p. 135) summarize these as follows:

1. *The Individually guided model*: individual teachers performing self-assessments and designing appropriate curriculum
2. *The Observer/assessment model*: principal or colleague observe teacher in class and then comment
3. *The Development/improvement model*: teachers involve themselves in whole school reform efforts
4. *The Training model*: teacher participation in course work, workshops, and conferences.

Teacher learning in these models has not been considered very effective due in part to the passive process where teachers are the recipients of knowledge and skills as defined by an outside authority such as a principal, visiting expert, or government administrator. The traditional model of

versions during part of this study could represent a confounding factor, it is unlikely since the two versions still address the same standards with the same types of tasks. The differences are more focused on reducing paperwork and streamlining the certification process rather than changing what it is teachers need to do. For example, Version 1 had two separate entries that addressed professional development history and community involvement respectively. In version 2, these entries were combined into one entry called 'Documented Accomplishments' that addresses both categories of professional activities.

⁸ For a thorough description of NBPTS, the process of certification, candidate requirements, and details of the process see Bailey and Helms (2000).

professional development is not constructed around any set of common standards or goals for the educators. For the most part, the experiences are isolated, extrinsically motivated, undisciplined, and leave little room to assess the accountability of results.

Hawley and Valli describe what they term a “consensus model” for improved professional development, oriented around seven principles:

1. Driven by goals and student performance
2. Involve teachers in the planning and implementation process
3. School based and integral to school operations
4. Organized around collaborative problem solving
5. Continuous and ongoing involving follow-up and support
6. Information rich with multiple sources of teacher knowledge and experience
7. Provide opportunities for developing theoretical understanding of the knowledge and skills learned. (Hawley & Valli, 1999, p. 137)

Part of a comprehensive change process that includes issues of student learning, the consensus model of professional development sees the teacher as an active learner and the process of learning embedded in practice. The model also emphasizes the role of reflection and professional discourse as effective means of teacher learning. Both traditional and emerging models of professional development add something meaningful to an understanding of how teaching may improve as a result of National Board certification. Ingvarson (1998, p. 133) explained, “In principle, both systems are essential and each should be complementary to the other, like two pillars holding up the same building.”

This view of professional development suggests that the process of National Board certification is an effective form of professional development. The process is completely voluntary as per the Consensus Model. It encourages professional discourse and collegiality as described in elements of both the traditional and consensus models. It encourages teachers to examine their work both inside and outside the classroom while embedding the collection of data on practice within the practice itself, played out over a considerable length of time. In addition, Board certification has well defined standards of performance and a well-specified goal as a result of participation. It is focused on both process and content and it incorporates meaningful attention to student learning as part of the work of self-assessment.

As National Board certification grows and matures, its impact may be felt beyond the teachers directly involved together with the students they teach. It may challenge many of the fundamental or traditional assumptions about what professional development looks like and how it is implemented. Ingvarson (1998, p. 134) writes,

Steadily increasing numbers of education authorities are accepting Board certification as evidence of professional development... The hope is that a new infrastructure of professional learning will develop around the incentive of Board certification, and there are signs that this is happening.

According to Reichardt (2001) “National Board certification provides a vision of good teaching and serves as a tool to direct individual teacher professional development,” and that there is “emerging evidence of the effectiveness of National Board certification as a method to improve teacher quality”. The study reported here is a first effort to test the proposition that board certification indeed is worthwhile professional development.

Evidence for National Board Certification as Effective Professional Development

The NBPTS has maintained that the process of recognizing accomplished teachers should “provide opportunities for candidates to develop professionally” (ETS, 1999). Some anecdotal evidence supports this objective. Whether they pass or fail, many teachers say they feel better about themselves as professionals and believe they are better practitioners because of their efforts. However, what teachers *feel* and *believe* may be quite different from what they *learn*. Therefore, inquiring about the precise nature of the learning outcomes from Board certification becomes important.

Anecdotal reports support the contention that National Board certification serves as effective professional development. Claims to this effect have regularly appeared (Tracz, et al., 1995; Kowalski, Chittenden, Spicer, Jones, & Tocci, 1997). Numerous teachers have testified to the benefits of National Board certification for their practice (Bailey & Helms, 2000; Gardiner, 2000; Jenkins, 2000; Chase, 1999; Benz, 1997; Haynes, 1995; Marriot, 2001; Roden, 1999; Wiebke, 2000). These teachers use such terms as “enlightening” (Mahaley, 1999) or “revitalizing” (Areglado, 1999) to describe their experiences with National Board certification. These accounts provide insights into the value of the Board certification experience, but tell little about what candidates actually may be learning.

Surveys have been conducted that expand upon testimonial accounts and provide more extensive interpretations of what National Board Certified teachers are learning from the assessments. For example, the NBPTS issued two reports based upon survey data that provided a national profile of Board Certified teachers and their feelings of “becoming a better teacher” from the certification process (NBPTS, 2001a, 2001b). These surveys report that among the more than 5,600 teachers who returned a completed survey (53% return rate), 92% felt that they were better teachers as a result of certification and 96% rated the certification process as a(n) “Excellent,” “Very Good,” or “Good” professional development experience (NBPTS, 2001b, p. 2). Such results are indicative yet leave open questions regarding the validity of self-reports and the particulars of how the process of National Board certification achieves particular outcomes.

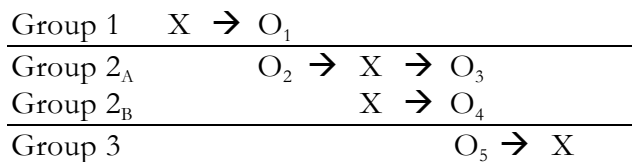
Other studies structured around a support group of candidates involved in the certification procedures provide more fine-grained evidence that candidates learn from NBC by participating in extended professional communities (Burroughs, Schwartz, & Hendricks-Lee, 2000; Manouchehehri, 2001; Rotberg, Futrell, & Lieberman, 1998). Studies also have found a value to the NBPTS materials such as the standards documents and portfolio instructions as important sources of teacher learning (Kowalski et al, 1997; Rotberg, et al., 1998). These investigations provide insight into the means and ends of Board certification, but do not pin down actual learning in any detail. Recent commentary on professional development identifies a need to complement small-scale, qualitative inquiry on teacher learning with quantitative investigations that attempt to clarify, identify, and substantiate specific outcomes (Borko, 2004; Floden, 2001; Crawford & Impara, 2001; Garet, Porter, Desimone, Birman, & Yoon, 2001; Knapp, 2003). This study responds to these calls.

Research Design

The measurement of teacher learning in this study required three components: a uniform curriculum serving as the “intervention,” a viable means of assessment, and a method that fit the cohort nature of National Board certification. For the curriculum, we chose the tasks and materials for AYA Science certification due to the lead author’s experience with this particular certificate. To

convert observations into measurable data, we relied on the procedures and rubrics the National Board employs to assess candidates for certification. To measure teacher learning outcomes that result from a specified treatment we turned to the logic of quasi-experimental design. The aim is to specify the association between certification (the independent variable) and teacher learning (the dependent variable). Crucial to such a design is the random selection of subjects and their random assignment to treatments. Since potential learning from National Board certification begins with a self-selected population, an experimental approach is not feasible (Campbell & Stanley, 1963; Cook & Campbell, 1979). In response, we chose a quasi-experimental design that accounts for the voluntary, self-selected nature of the subjects' participation while maintaining the pre-post collection of data. Titled the "Recurrent Institutional Cycle Design" (RICD), it controls (to the extent possible) for non-random threats to internal validity while providing a means of establishing some degree of causality between the treatment and observed results (Campbell & Stanley, 1963).

The RICD has been used for treatments that recur on a cyclical schedule where one group of individuals is finishing and another group is just beginning (Campbell & McCormick, 1957; Shavelson, Webb, & Hotta, 1987; Jimenez, 1999). Numerous studies in the social and medical sciences have used some variation of the RICD to address questions pertaining to program effectiveness including the effects of an intervention on leadership development (Lafferty, 1998) and on employment (Juin-jen, 1999). As Figure 1 illustrates, the RICD allows for cross sectional data to be collected from different groups at the same time; and longitudinal data from the same group over time.



X—Intervention (Board certification process)

O—Data Collection (Interviews)

Figure 1

RICD design for study. In this diagram of the research design, time is measured along the x-axis and groups of subjects are along the y-axis. The observation references are important in interpreting Table 4.

Over approximately 15 months from August 2002 to November 2003, data were collected from three groups sampled from three consecutive cohorts of AYA Science candidates. The design allows for the comparison of the pre and post measures between groups (cross sectional) and within groups (longitudinal). Pre to post gain scores test the relationship between the intervention and specified learning outcomes. Group 2 was divided randomly to create Group 2A and 2B. The two subgroups were needed to test for the effect of data collection on observed results. Since Group 2A had both pre and post observations (denoted 2A–Pre and 2A–Post respectively) and Group 2B only post, the comparison between the two post groups would allow us to consider any impact (if any) the interview process for pre observations may have had on the assessed scores (effect of testing).⁹

⁹ A comparison of Group 2A-Post and Group 2B revealed no significant differences indicating no effect of testing..

The RICD involves some restrictions, particularly concerning external validity (Campbell & Stanley, 1963). Generalization of outcomes is limited to teachers pursuing certification in AYA Science with the current set of 13 standards and not to all science teachers or other certificates. As well, quasi-experimental research designs also involve threats to their internal validity from a number of sources. While the RICD addresses threats due to history, testing, selection, mortality, and instrumentation, it cannot guard against maturation effects (Campbell & Stanley, 1963, Merriam & Simpson, 1995; Cook & Campbell, 1979). However, since the research is focused on the acquisition of a series of “complex and highly specialized skills and knowledge sets, it seems unlikely that just growing older or more experienced would be a significant influence on outcomes” (Campbell and Stanley, 1963, p. 59).

The effects of history or a test-retest effect cannot explain observed cross sectional differences between cohorts. The biggest threat to such observed differences however could be due to differences in recruitment (selection) from one year to the next. To account for this possible confounding explanation, demographic information was collected from each cohort and its respective group to provide a profile for comparing how each cohort and group compares on the characteristics of gender, years of teaching, school context, and students’ ability. Finally, since the instruments used to measure differences were unchanged throughout the study, it is unlikely that this could be a threat to internal validity. For discussion of additional limitations and caveats, see Appendix A.

Hypothesis

Our primary hypothesis states that the post-data (Observations 1 and 4) will demonstrate gains when compared with the pre-data (Observations 2 and 5). In this hypothesis, every participant from each of the three cohorts is considered simultaneously with four out of the five available observations. The alternative hypothesis (H_{Alt}) is that post scores from Group 1 and Group 2B will be greater when compared to the pre scores from Group 2A (Pre) and Group 3. The null hypothesis (H_{Null}) is that post scores from Group 1 and Group 2B will not be greater when compared to the pre scores from Group 2A (Pre) and Group 3.

By identifying, quantifying, and substantiating observed differences among groups on each of the National Board’s thirteen standards, this study provides evidence of teacher learning. This operational definition for ‘learning’ within the context of this investigation allows for identifying effects of intervention on thirteen dimensions of practice and a rich analysis into what the observed ‘learning’ might mean. The experience of each candidate with the intervention of National Board certification serves as the independent variable. The dependent variable is represented by the assessed scores of each candidate on each of the thirteen standards of accomplished secondary science teaching.

Study Population

This investigation focused upon the population of secondary science teachers who applied to undertake National Board certification.¹⁰ Applicants must be certified in the state in which they teach, currently teach at least two classes in the area of certification, and have at least 3 years of full

¹⁰ AYA Science was chosen since the lead author has first hand experience having achieved National Board certification in AYA Science in 1998.

time teaching experience. For this study, any teacher who registered for the certification process had to verify these requirements before being accepted as a candidate for National Board certification. The sample for this study was drawn from this pool of self-selected candidates for AYA Science certification. The population pool included all registered teachers for AYA certification for the years 2001–2002, 2002–2003, and 2003–2004. For each year this represents approximately 450–650 teachers. From this pool, approximately half the teachers were randomly invited to participate. The final list of participants was determined on a first to reply basis. Recruitment of teachers ended once each of the three groups reached the target of 40 teachers; however, recruitment procedures from year to year were not perfectly even. Effects due to variations in recruitment remain an important limitation to this study (details in Appendix A).

Group to Cohort comparisons indicated a high degree of similarity and fair representation, though the information available on the entire cohort included only age, gender, and geographic location. In analyzing the similarities and differences between Groups 1, 2A, 2B, and 3, twelve characteristics were compared. Table 1 provides a summary of these comparisons. Of those areas that showed difference (i.e., years of experience, learning of, incentive for, and support for National Board certification), the most important would appear to be years of experience. Groups 1, 2A, and 2B have an average of 15.3 years of experience whereas Group 3 has an average of 11.0 years of experience. Though Group 3 is significantly different from the other groups we would argue that there is very little qualitative difference between teachers who have 11 years versus teachers who have 15 years of experience. According to the literature on teacher effectiveness, both lengths of time fall into the category of ‘veteran’ teacher (Stronge, 2002; Darling-Hammond, 2000). The other identified between-group differences (how teachers learned of National Board certification, their incentive for pursuing Board certification, and the types of support provided for the process) are most likely due to the smaller group sizes for 2A and 2B. When both groups are combined, the differences are not significant. Overall, more than 90% of all teachers in this study received some form of financial incentive and support for pursuing National Board certification.

Table 1
Summary of Group to Group Comparisons on Demographic Variables

| Demographic Characteristic | Group 1–Post | Group 2A–Pre | Group 2B–Post | Group 3–Pre |
|------------------------------|--------------|--------------|---------------|-------------|
| Grades | ND | ND | ND | ND |
| Content | ND | ND | ND | ND |
| School | ND | ND | ND | ND |
| Region | ND | ND | ND | ND |
| Students | ND | ND | ND | ND |
| Gender | ND | ND | ND | ND |
| Years Teaching | ND | ND | ND | Different |
| Class Size | ND | ND | ND | ND |
| Length of Profiles (WORDS) | ND | ND | ND | ND |
| Learn of National Board | ND | Different | ND | ND |
| Incentive for National Board | ND | ND | Different | ND |
| Support for National Board | ND | ND | Different | ND |

ND: no significant difference.

Instrumentation

Interview protocols and assessment rubrics constitute the two forms of instrumentation used in this study.¹¹ The goal for the structured interview was to reproduce on a smaller scale the portfolio construction experience candidates complete in the certification process. Trained assessors then scored the transcribed interview as if the transcripts were complete portfolios in miniature. The coding of the transcripts by assessors provided a form of assessment that measured a candidate's weight of evidence regarding each of the thirteen standards of accomplished teaching. This evidence was then converted into a score on a 4-point scale so that each interview yielded thirteen scores of teacher knowledge, which in turn formed the basis for the pre-post comparisons.

The structured interview developed for this study is based (in part) on an approach developed by Kennedy, Ball, McDiarmid, and Williamson (1993) to track changes in teacher knowledge over the course of teacher education. These investigators state that one possible way of identifying changes in what teachers know is by "presenting teachers with hypothetical teaching situations" (Kennedy et al., 1993, p. 7). They continue, "If the situations were standardized," then "the amount of irrelevant, idiosyncratic differences in responses" could be reduced, and "the detailed, contextualized information about teachers' perceptions of practice" (ibid) would be increased. Focusing the protocols on teaching situations and standardizing them for all study participants "allows for researchers to see how the various aspects of expertise—knowledge, beliefs, attitudes about learning, teaching, and subject matter were drawn on to make teaching decisions" (ibid).

The interview for this study had six sections representing the different parts of the portfolio. Each section (or scenario) was modeled after one of the four mandatory portfolio entries. In addition, background and school context information also was collected. Table 2 summarizes the similarities and differences between the structured interview protocols and the portfolio entries.

Table 2
Comparison of Structured Interview and Portfolio Entry

| Required Aspects | Structured Interview Protocols | NBPTS Portfolio Entry/ Standard |
|------------------------|---|---|
| Introductory Questions | Teacher Background School Context Student Profile | Teacher Background School Context Student Profile |
| Scenario #1 | Teaching a Major Idea in Science Over Time | Teaching a Major Idea in Science Over Time |
| Scenario #2 | Scientific Inquiry | Scientific Inquiry |
| Scenario #3 | Best Practice | Assessment Center Tasks |
| Scenario #4 | Whole Class Discussion | Whole Class Discussion |
| Scenario #5 | Community, Professional Development, and Leadership | Community, Professional Development, and Leadership |

To assess the quality of individual teacher responses to the structured protocols in the interview, we used the rubrics and scoring procedures developed by the NBPTS for candidate

¹¹ For a discussion of the NBPTS rubrics and assessment procedures, please see Educational Testing Service, ETS (1999).

portfolio assessment. Experienced and knowledgeable National Board assessors for AYA Science were contracted to apply the assessment tools to the interview transcripts in a manner that paralleled their application to portfolio entries submitted to the National Board for certificate evaluation. The scoring rubrics are based on the same thirteen standards of accomplished secondary science teaching as the portfolio prompts that candidates must address in the presentation of their practice.

Standards to be Assessed

Teams of experienced science teachers, academics, researchers, and educational leaders developed the standards of accomplished teaching used in the AYA Science certificate. The standards are field tested regularly, and every five years are re-evaluated and adjusted based on input from the science education community. They represent an expert consensus on what constitutes ‘accomplished teaching’ in science. Table 3 provides an overview of the thirteen standards as separated into their four sets¹².

Table 3
Standards for AYA Science Interview Protocols

| Standard set | Standard | |
|---|----------|----------------------------------|
| Preparing the Way for Productive Student Learning | I. | Understanding Students |
| | II. | Knowledge of Science |
| | III. | Instructional Resources |
| Advancing Student Learning | IV. | Science Inquiry |
| | V. | Goals & Conceptual Understanding |
| | VI. | Contexts of Science |
| Establishing Favorable Context Learning | VII. | Engagement |
| | VIII. | Equitable Participation |
| | IX. | Learning Environment |
| Supporting Teaching and Learning | X. | Family & Community Outreach |
| | XI. | Assessment |
| | XII. | Reflection |
| | XIII. | Collegiality & Leadership |

Interrater Reliability

Interrater reliability is a measure of the degree to which raters agree in their assessment of each standard for each candidate. Assessors included the lead author (a former National Board assessor) and two National Board portfolio assessors each with 3 years of assessment experience, who also had served as assessor trainers for the AYA Science entries. To improve the agreement among the three assessors, the study provided a scoring guide, resource documents, and one to one training.

Assessors were trained for three days based on procedures used by the National Board to prepare its assessors to score ‘live’ portfolios. Prior to actual scoring, assessors practiced their assessment skills on model entries (previously scored portfolios) where they learn to understand and agree upon a common framework for scoring. The scoring rubric judged the family of scores on a

¹² See Appendix D for definitions of each standard.

scale from one to four. For this study, we provided the same sort of training, but on a much smaller scale. Assessors received a binder with 125 pages of training materials to guide their scoring. After working through all exercises (approximately 4 hours of work) and returning the completed score sheets for practice entries assessors were evaluated for their level of agreement and accuracy. An hour-long conversation with each assessor followed where feedback was provided and calibrations made. Only after successfully finishing the training, were 'live' entries given to the Assessors for scoring.

The assessor training was designed around ideas that have been shown to reduce rater effects. Objectives for the training materials and procedures included familiarizing judges with the measures used in the study, ensuring that the assessors understand the sequence of operations they must perform for each entry, and providing direction on how questions regarding data may be resolved or interpreted (Rudner, 1992). That all three assessors were quite familiar with the scoring rules and procedures of the National Board increased confidence in their ratings. Still, the measures of reliability among raters reflect the difficulty and complexity of the task. Comparison of ratings by three assessors on thirteen standards reveals a fairly wide range of variability. Inter-rater reliability analysis for this study indicates a fair to moderate relationship among assessor's scores for the same interview. A Pearson correlation of .458 existed among the three raters. (For a complete description of the inter-rater reliability statistics see Appendix B.)

Reasons for this moderate level of agreement can be traced to the overlapping nature of some of the standards used. For example, "*Knowledge of Students*" states that teachers know how to assess their students' learning. The standard for "*Assessment*" also emphasizes the assessment of student understanding. So an assessor has more than one category in which to place evidence pertaining to assessment. Unlike the work of assessors in portfolio assessment for the National Board, assessors on this project were instructed to determine thirteen distinct measures as opposed to the one overall measure required of the Board's process. With such complexity and the opportunity to place evidence in multiple categories, the moderate inter-rater reliability is understandable. Still, the conclusions from this study must be qualified by this reliability concern. (See Appendix A for further details on this matter.)

Data Collection

Data collection began with receipt of the Consent Form, which included some basic questions of the candidates. This study was conducted under the assumption that data would be collected in clearly identifiable pre and post intervention conditions. Post observations were made after a candidate completed and submitted a portfolio and had taken the assessment center exams and before they received word from the National Board about the outcome. This timing was achieved successfully in all Post-Observation cases. Pre-observations were made after a candidate paid the non-refundable registration fee and before significant amounts of portfolio work had been completed. (See Appendix A for a more detailed discussion of the problems associated with the pre-observations.)

Each candidate was sent an identical interview packet containing a sealed six-minute video clip of a whole class discussion in science, student artifacts, and classroom situations to be discussed during the interview. The specific questions they would be asked about these materials were not included in the packet. During an extended telephone interview (ranging from 40 to 90 minutes), teachers examined and analyzed the artifacts, thought about and responded to the interview questions, and watched the videotape for the first time. After the audio taped interview was

transcribed, a 'processed' version of the transcription¹³ was then scored by at least two assessors using the rubrics and standards of the National Board certification process. For each transcript, an assessor provided one score for each of the thirteen standards. The thirteen assessed scores for each candidate were then aggregated to the group level so that means representing different observations could be compared for significant differences at the overall, set, and individual standard level of analysis.

Results

The Flowchart for analyzing the results is presented in Figure 2 which provides a branching schematic for the decision making process. In this approach, testing continues only when significant differences are identified at the Overall, Sets, and then the Standards levels of analysis. The Recurrent Institutional Cycle Design for this study yielded a total of five observations (see Figure 1). Table 4 presents the Combined Pre-Post Comparison (H1) using four out of the five observations, pooling data from all 118 participants.¹⁴ All data sets except for O3 (Group 2A–Post) are included in this comparison. Throughout this analysis, one-tailed Contrast t-tests were used to determine significance.¹⁵ A comparison of

This study asked, "Does National Board certification lead to significant learning in teachers undertaking the process?" If so, the post observations would be greater than pre observations. Table 4 provides the results. There are 114 degrees of freedom indicating every participating teacher in the study was taken into account for this comparison. The value of the contrast has a *p* value of .009, which is significant at an alpha level of 0.05. The corresponding effect size of this observed difference is 0.473, which according to Cohen's effect size metric for behavioral sciences is a 'moderate' indication that there are meaningful differences between pre and post group scores (Cohen, 1977).

¹³ The 'processed' transcript was a crucial step in this study. The assessors' task was meant to resemble their experience with 'live' portfolios as much as possible. To hand them a 'raw' transcript would have impeded their ability to adequately and fairly evaluate the written words. The National Board is a stickler for format (font, margins, spacing, etc) and the researchers wanted the interview transcripts to resemble a real entry as much as possible. The aim of the processing was to improve the appearance, but not the meaning or intent of the words. See Appendix C for an example of this process.

¹⁴ In Group 2A, one more subject than anticipated dropped out of the certification process (we anticipated six dropping out of the certification process and the study, but the actual number was seven) and technical problems with the tape recorder during one interview allowed for only 18 usable pre-post comparisons.

¹⁵ T-tests were used to compare two post groups with 2 pre groups. ANOVA analysis that compares the four groups without consideration of pairings, produces similar if not less significant results. Instead of nine significant standards, only the three most significant standards from the t-tests remain significant in the ANOVA analysis at the .05 level. Since the comparison is between pairs of groups and not four groups independent from each other, the contrasting paired t-test is most appropriate.

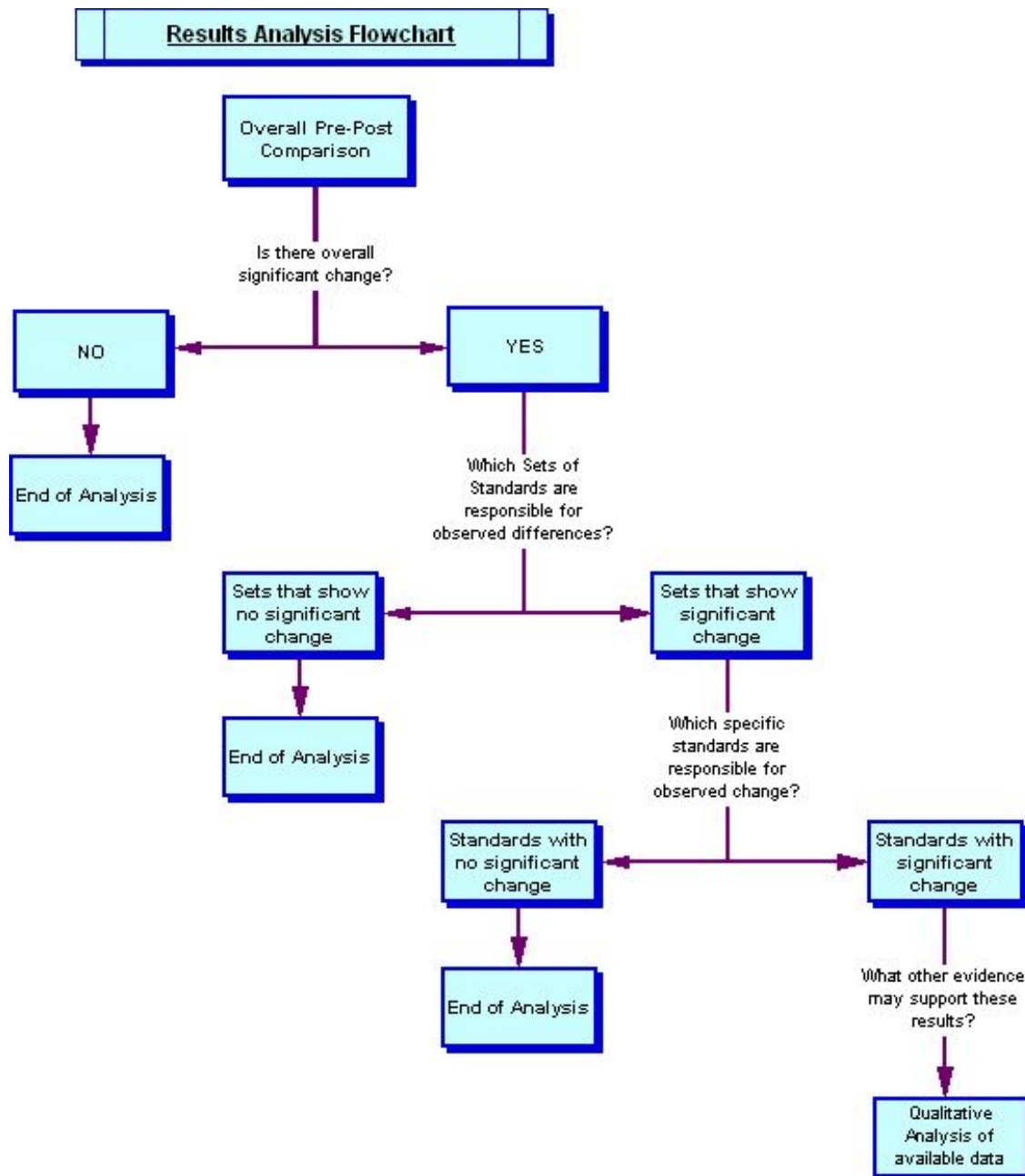


Figure 2
Analysis Flowchart

Table 4
Descriptive Statistics for all Groups in Hypothesis 1

| Cohort Group | Obs. | N | Mean | Std. Dev. | Std. Error | 95% Confidence Interval | | Min. | Max. |
|--------------|----------|------|------|-----------|------------|-------------------------|-------|------|------|
| | | | | | | Lower | Upper | | |
| 1 | 1 (post) | 40.0 | 2.81 | 0.45 | 0.07 | 2.67 | 2.96 | 1.65 | 3.69 |
| 2A | 2 (pre) | 18.0 | 2.64 | 0.56 | 0.13 | 2.36 | 2.92 | 1.70 | 3.62 |
| 2B | 4 (post) | 20.0 | 2.79 | 0.43 | 0.10 | 2.59 | 2.99 | 2.04 | 3.64 |
| 3 | 5 (pre) | 40.0 | 2.54 | 0.39 | 0.06 | 2.42 | 2.67 | 1.54 | 3.19 |
| Means | | 29.5 | 2.69 | 0.46 | 0.04 | 2.61 | 2.77 | 1.54 | 3.69 |

See Figure 1 for the sequence of observations and cohort groups.

Table 5
Test for Significance Overall for Hypothesis 1

| Standard | Value of Contrast | Std. Error | <i>t</i> | <i>df</i> | <i>p</i> (1-tailed) | Effect Size |
|----------|-------------------|------------|----------|-----------|---------------------|-------------|
| Overall | 0.423 | 0.176 | 2.400 | 114 | .009 | 0.473 |

Next, we can look more closely into the standards to pinpoint more specifically what teachers may have learned from the certification process. Four sets of standards for AYA Science require four separate t-tests for analysis so it becomes necessary to employ an adjustment procedure that takes into account the use of multiple t-tests, which increases the likelihood of committing a Type I error. To address this concern, a Bonferroni adjustment procedure was employed. This conservative adjustment reduces the risk of a family-wise error, while still allowing for the identification of observed differences (Yip, 2002; Homack, 2001). Because we wished to maintain a 0.05 alpha level for each of the four tests, significance was determined at an α level of .0125.

At this level, as Table 6 reveals, the contrasts for Set II (Advancing Student Learning) and Set IV (Supporting Teaching and Student Learning) were both found to be significant at $p = 0.008$ and $p = .005$ respectively. Set III (Establishing a Favorable Context for Student Learning) was found to be marginally significant at $p = .013$. Set II and Set IV had effect sizes of 0.482 and 0.524 respectively indicating the main areas of observed learning. Because significant differences were identified in three out of the four sets, we can examine each set of standards more closely to identify the specific standards that may be responsible for the observed learning.

Table 6
Results of Analysis at the Level of Sets of Standards

| Set | Value of Contrast | Std. Error | <i>t</i> | <i>df</i> | <i>p</i> (1-tailed) | Effect Size |
|--|-------------------|------------|----------|-----------|------------------------|-------------|
| I. Preparing the Way for Productive Student Learning | 0.304 | 0.176 | 1.727 | 114 | .043 | 0.341 |
| II. Advancing Student Learning | 0.494 | 0.202 | 2.442 | 114 | .008* | 0.482 |
| III. Establishing Favorable Context for Student Learning | 0.461 | 0.204 | 2.253 | 114 | .013 | 0.444 |
| IV. Supporting Teaching and Student Learning | 0.459 | 0.173 | 2.656 | 114 | .005* | 0.524 |

* $p < .0125$

Sets II, III, and IV have a total of 10 standards requiring 10 t-tests. Once again, a Bonferroni Adjustment procedure was used to reduce the chances of a Type I error due to repeated use of the test. Again, to maintain an overall α level of .05, the significance for each test was set at $\alpha = .005$. At this level, two standards are significant and two are marginal as shown in Table 7. *Scientific Inquiry* from Set II and *Assessment* from Set IV are significant at .001 and .002 with effect sizes of 0.606 and 0.596 respectively. The two standards that were marginally significant included *Goals and Conceptual Understanding* from Set II and *Reflection* from Set IV at $\alpha = .009$ and .007 respectively. Though marginally significant at the level of the Set, Set III did not have any standards significant at the .005 level.

What might account for the observed differences at the overall, sets, and standards levels of analysis? Are the gains observed in this study due to the intervention or something else? What percent of the observed variance can be attributed to possible covariates? To answer these questions, an analysis of covariance (ANCOVA) was conducted using potential confounding factors. Co-variables included gender, years of experience, class size, student type, school context, and geographic region. The results of this analysis are presented in Table 8.

Table 7

Analysis of Individual Standards

| Standard | Value of Contrast | Std. Error | <i>t</i> | <i>df</i> | <i>p</i> (1-tailed) | Effect Size |
|---|-------------------|------------|----------|-----------|------------------------|-------------|
| <i>II. Advancing Student Learning</i> | | | | | | |
| Science Inquiry | 0.667 | 0.217 | 3.073 | 114 | .001* | 0.606 |
| Goals and Conceptual Understanding | 0.494 | 0.206 | 2.392 | 114 | .009 | 0.472 |
| Contexts of Science | 0.321 | 0.279 | 1.151 | 114 | .126 | — |
| <i>III. Establishing Favorable Context for Student Learning</i> | | | | | | |
| Engagement | 0.477 | 0.208 | 2.294 | 114 | .012 | 0.452 |
| Equitable Participation | 0.448 | 0.242 | 1.851 | 114 | .033 | 0.365 |
| Learning Environment | 0.457 | 0.223 | 2.049 | 114 | .021 | 0.404 |
| <i>IV. Supporting Teaching and Student Learning</i> | | | | | | |
| Family and Community Outreach | 0.174 | 0.217 | 0.803 | 114 | .212 | — |
| Assessment | 0.647 | 0.214 | 3.022 | 114 | .002* | 0.596 |
| Reflection | 0.607 | 0.241 | 2.515 | 114 | .007 | 0.496 |
| Collegiality and Leadership | 0.409 | 0.180 | 2.276 | 114 | .012 | 0.449 |

* $p < .005$. There is no effect size calculated for standards where the significance is over .10.

In this ANCOVA, only “Student Type” is a significant covariate and possible confounding variable at the .05 level, $p = 0.026$. The teacher’s gender, years of experience, class size, school context, and geographic region did not co-vary with the observed gains in assessed scores. So, is ‘Student Type’ a viable alternative for explaining observed results?

Table 8
Analysis of Covariance

| Source of Variations | Degrees of Freedom | Sum of Squares | Mean Square | <i>F</i> | <i>P</i> |
|----------------------|--------------------|----------------|-------------|----------|----------|
| Model | 14 | 5.39 | 0.38 | 2.12 | .017 |
| Cohort Group | 3 | 1.78 | 0.59 | 3.26 | .025 |
| Pre-Post Comparison | 1 | 0.65 | 0.65 | 3.59 | .061 |
| Gender of Teacher | 1 | 0.27 | 0.27 | 1.49 | .226 |
| Years Experience | 1 | 0.00 | 0.00 | 0.01 | .939 |
| Class Size | 1 | 0.04 | 0.04 | 0.24 | .622 |
| Student Type | 3 | 1.75 | 0.58 | 3.21 | .026* |
| School Context | 2 | 0.52 | 0.26 | 1.42 | .247 |
| Region | 3 | 1.03 | 0.34 | 1.88 | .138 |
| Error | 101 | 18.36 | 0.18 | | |
| Corrected Total | 115 | 23.75 | | | |
| R-Square | 0.23 | | | | |

* Significant at the $p = 0.05$

The “Student Type” variable was derived from candidate responses to the question, “How would you generally describe your students?” Responses typically were quite general (i.e. low, below average, average, above average, high, varied, or mixed.). We then coded all responses to this question into four possible categories: Low, Average, High, or Varied. The ‘student type’ indicator is not based upon any standardized or objective source of data, but rather each teacher’s overall impression of their students. “Student Type” reflects the teacher’s perception of his or her students’ general ability rather than the actual ability level of students.

If we look at how the various pre and post groups rated their students’ abilities and compare their observed overall scores, an important pattern emerges. Teachers who rated students’ abilities lower tended to score lower in this study than peers who rated students’ abilities higher. However, this relationship holds true for both pre and post group observations with teachers in the post groups consistently demonstrating improved scores compared to teachers in the pre groups regardless of how they rated their students’ abilities. Figure 3 illustrates this result showing four parallel lines. The two lines lower on the x-axis are from the pre groups and the two lines higher on the x-axis are from the post groups. For each category of student ability, the scores from the post observations are higher than those from the pre observation. Because all four lines never intersect across all four student ability categories, we can conclude that there is no interaction between how a teacher rates student ability and their observed scores. In fact, the certification process becomes a more likely explanation for observed differences since the only possible confounding variable shows no interaction with the results.

This co-variation between observed scores and the teachers’ self reported impressions of student ability suggests that a teacher’s expectations for student success play an important role in predicting what kind of scores (both pre *and* post intervention) a teacher would receive in this study. Teachers with higher expectations for their students based on a more positive assessment of student abilities tend to perform better on the standards that assess their knowledge and skills.

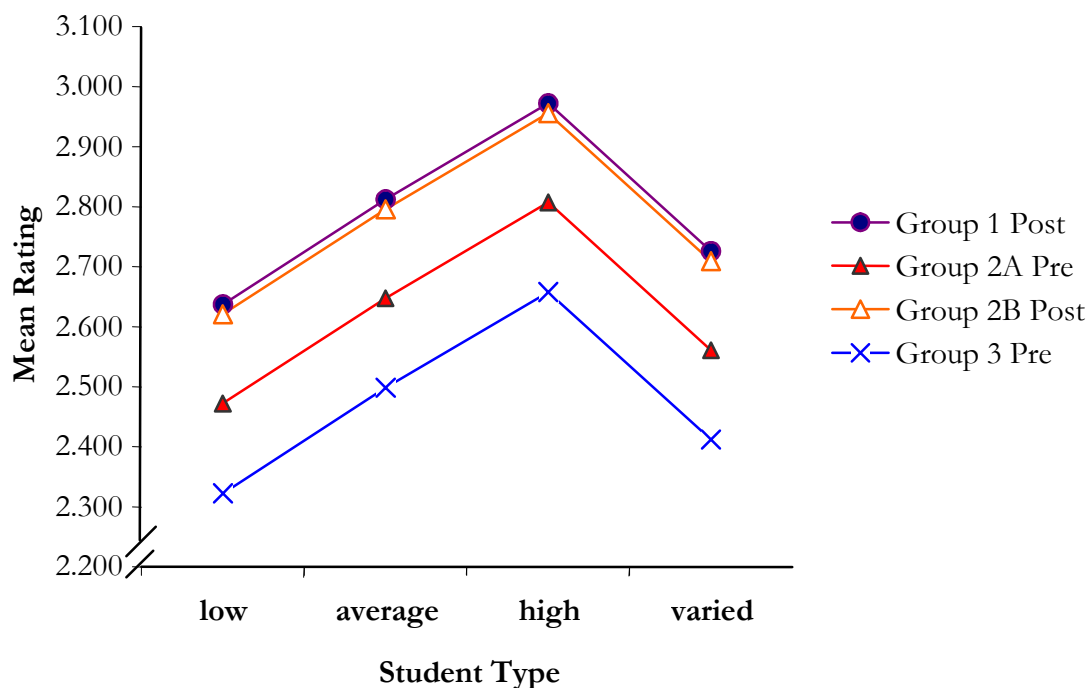


Figure 3
Mean teacher rating by student type

Two standards then—*Scientific Inquiry* and *Assessment*—demonstrated the most improvement as determined by the assessment process. How might this result be interpreted? To help answer this question, we turn to qualitative data from the open-ended interview questions. At the end of each of the 138 interviews conducted with 120 teachers for this study, participants were asked to generally address their experience with National Board certification.¹⁶ The candidates responded with answers that provided detailed evidence regarding their overall, positive, and/or problematic aspects of the experience. A quantitative comparison of all 78 post-interview responses to this prompt provides a means of comparing how teachers opted to discuss aspects of the experience. For example, in their response, did teachers focus on issues of Engagement, Reflection, or Knowledge of Students?

Analysis of these data using a grounded theory approach and coding scheme based upon the language and meaning of each of the 13 standards supplies a gauge for determining which standards (if any) participating teachers found most significant to their own learning (Glaser, 1995). The results strongly support the quantitative results. The three standards commented on most by teachers (*Scientific Inquiry*, *Assessment*, and *Reflection*) corresponded with the observed significant (or marginally significant) gains. Figure 4 overlays the number of candidate comments regarding specific

¹⁶ 18 out of 20 Teachers in Group 2A were interviewed twice resulting in a total of 138 interviews. Of these, 78 were post interviews (40 from Group 1, 20 from Group 2B, and 18 from Group 2A-Post) and 60 were pre interviews (40 from Group 3, and 20 from Group 2A-Pre).

standards with the observed gain scores. The areas of greatest change correspond quite closely with the reported learning opportunities afforded by National Board certification in Adolescent and Young Adult Science. The convergence of these two sources of evidence supports the conclusion that the certification process promotes productive teacher learning in the areas of Scientific Inquiry, Assessment, and Reflection.

Observed Gains and Teacher Comments

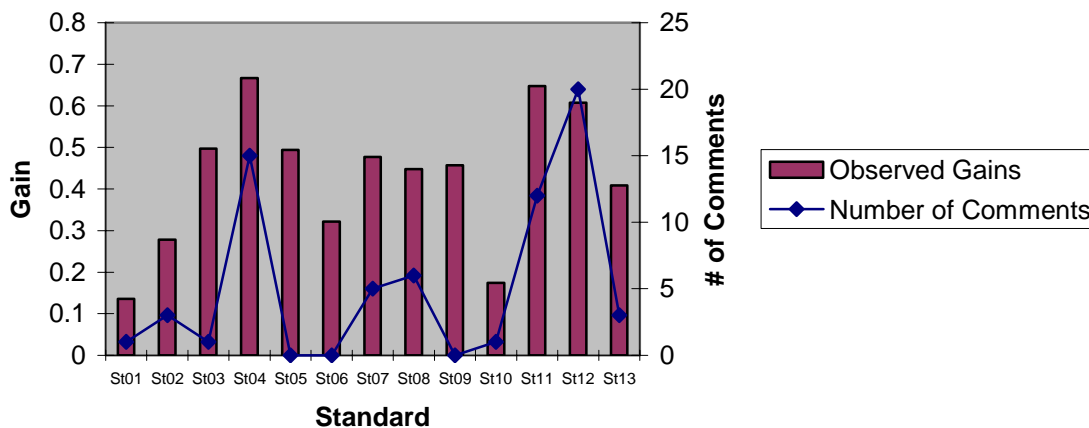


Figure 4
Comparison of Observed Gains with Teacher Comments

Scientific Inquiry: Evidence for Learning

The National Board defines the Scientific Inquiry standard as a scale that pertains to a teacher's ability to develop in students the mental operations, habits of mind, and attitudes that characterize the process of inquiry (see Appendix D for operational definitions for all 13 standards of accomplished science teaching). Is it reasonable that the greatest teacher learning was identified with the *Scientific Inquiry* standard? To answer this question, we return to the comments made by the teachers who shared their ideas about the certification process. The open-ended interview data allow for further probes into the nature of teacher learning. When we discuss *Scientific Inquiry*, what does it mean? How does the Board conceptualize Scientific Inquiry? In its discussion of the standard, the NBPTS states that,

It is not a basic goal of science instruction to fill students with as much information as possible; rather, it is to help students acquire the mental operations, habits of mind and attitudes that characterize the process of scientific inquiry—that is, to teach them how scientists question, think, and reason.
(NBPTS, 1997, p. 31)

According to the standard, the best way for teachers to reach this goal is to have students take an “active role” in their learning by arranging frequent opportunities for “hands on” science activities and “open-ended investigations” complete with post-activity time for reflection and analysis. Teachers’ understanding of teaching with scientific inquiry is reflected in the choices and decisions made in their planning, lesson management, and assessment. Teachers choose age- and skill-level appropriate classroom activities that are as much ‘minds on’ as they are ‘hands

on.’ Other indications that a teacher effectively employs scientific inquiry in the classroom pertain to questioning style, wait-time after asking questions, discussion management, and an acceptance of the “unpredictable consequences of an activity and student-centered pedagogy” (NBPTS, 1997, p. 32).

Though the definition of scientific inquiry may be broad and have different meanings to different teachers, the National Board defines it with specific characteristics and observable qualities. An illustrative list of skills, dispositions, habits of mind, and pedagogical approaches outlines what a teacher should know and be able to use effectively in the science classroom. Consequently, pre-to-post improvement in scores on *Scientific Inquiry* supports the claim that teachers are learning to align their practice more closely with the National Board’s conception of scientific inquiry and teaching. The interview evidence generally supports this claim.

For many teachers who commented on this standard, it is apparent that the National Board’s version of ‘scientific inquiry’ constituted a new approach to teaching science. For example, when Sharon, a teacher from Wyoming, was asked what she learned from the certification process, her response expresses a recurring theme among many candidates¹⁷. She said:

I would point as an example to the increased use of inquiry within my classroom. I think that that has had some strong benefits in terms of helping students to think about what science is. And science is a process and not just a memorization of facts to spew them back out at the teacher. And I do think that doing the National Boards has helped me to incorporate that more into the classroom.
(Group 2A–Post, Teacher #12)

Mike, another teacher from Wyoming, concurred:

[The] National Board is making me look at how much scientific inquiry I’m doing where the students are actually doing the inquiry versus me just regurgitating.
(Group 2B, Teacher #32)

For both of these teachers, National Board certification allowed them to revisit, rethink, and re-try a process that they were already familiar with in the science classroom.

For other teachers, scientific inquiry represented a new way to teach. In attempting to fulfill the requirements of the portfolio, they were directed to teach according to the scientific inquiry method. As Susan, a teacher from Arkansas commented,

I had a real tough time coming up with and dealing with the inquiry base process. And I found out that other science teachers that had gone through the National Board Certification—some who got it and some who didn’t—had a tough time with that. It’s very difficult to not want to jump in and help the kids. And to see them sort of struggling and kind of thinking what they needed to do type thing. And even though, you know, I would—I had to be very careful with my questions so that they would think of what they had to do next without me giving them an idea as to what to do. (Group 2A–Post, Teacher #17)

Susan is describing the issues associated with trying to teach in a new way. The efforts were demanding and went against her existing tendencies and habits of mind. She needed to be much more self-conscious regarding how students were asked questions and how responses to students were formulated. For an experienced teacher with well-rehearsed “scripts” for interacting with students, the standards for scientific inquiry proved to be a difficult challenge, but probably a significant learning experience.

Learning associated with scientific inquiry can be traced to the portfolio requirements for National Board certification. The portfolio describes in detail the National Board’s version of

¹⁷Pseudonyms are used for all teachers.

scientific inquiry. The prompts teachers need to address throughout the entry on “*Active Scientific Inquiry*” pertain to how decisions are made, actions taken, and evidence collected in support of student learning. This framework for reflection and analysis provides the curriculum from which teachers plan, construct, and implement lessons.

Assessment: Evidence for Learning

Assessment in educational circles has a wide range of meanings, from standardized tests that provide a particular view of what students may understand to more immediate and classroom based activities. The National Board defines its “*Assessment*” standard as, “the process of using formal and informal methods of data-gathering to determine students’ growing scientific literacy, understanding, and appreciation” (NBPTS, 1997, p. 45). The results from this process are then used to inform instructional decisions (Gallagher, 2000); hence, effective assessment relies heavily on a teacher’s strong understanding of content.

Evidence in support of learning related to *Assessment* is substantial. The theme emphasized is the repeated use of focused, detailed, and extensive evidence around student learning. For many teachers, this is a new practice. Through the efforts to construct a meaningful portfolio, issues of student assessment are explored in rich detail. For example, Karen, a teacher from Kentucky, reports,

Well, like I used to just grade a test. You know based on how the grades were on the test that would kind of be my indication of if the kids learned or not. And now, I just see that there’s all types of assessment and that how a student does on your test is going to have an influence on your teaching and how you instruct. You know I never looked at that as a tool for changing my instruction.
(Group 1, Teacher #9)

Karen expresses a deeper understanding and appreciation for assessment that did not exist prior to the certification process. Her practice is enriched by assessment, as it becomes a tool for improving student learning instead of a requirement at unit’s end. Another effect of the National Board’s emphasis on learner assessment also provoked Karen to change her view of students in the learning process. She continues,

I realized that you really need to look at a student’s individual needs and style—you know like for a long time I taught just college bound kids and you kind of think that they are just all the same. And they are not....it just made me individualize more. (Group 1, Teacher #9)

Karen describes a significant adjustment regarding her approach to teaching and learning, from a teacher-centered stance to a more student-centered appreciation. Assessment facilitated this change by allowing the teacher the opportunity to more closely examine what individual students are actually learning in relation to her teaching.

The qualitative data also suggest that the learning associated with the assessment standard is even more profound than that observed in relation to the scientific inquiry standard. For example, once the power of detailed and intensified assessment of student learning is experienced, it changes the way a teacher thinks about practice. Rita, another teacher from Kentucky, says,

And it truly—and it has already carried over to this year—you know when kids don’t write very well—you almost dread reading their writing. But I find myself really wanting to read their lab reports and stuff. And I feel like what I say to them on their papers—I definitely give them more feedback. But my feedback is

more direct. So I feel like I analyze their work better than I did before. (Group 1, Teacher #41)

In this example, the teacher “looks forward” to a task that she previously “dreaded.” Assessment has not only improved her understanding of student ideas and reasoning, but has also led to an improved appreciation for effective engagement through appropriate and complete feedback.

These responses suggest that for some of the teachers sampled, a rather profound shift has taken place, perhaps pointing to the kind of “self-sustaining, generative” learning now recognized in the literature as a relatively rare event (Franke, Carpenter, Fennema, Ansell, & Behrend, 1998; Franke, Carpenter, Levi, & Fennema, 2001). In this case, we cannot know how long such effects may last, nor how such reported insights actually may affect teaching practice. But in the annals of teacher learning, the reports of these teachers are remarkable in themselves against the backdrop of so many teachers’ dismal accounts of their professional development experiences. The process of Board certification appears to have been a transformative experience for at least some teachers on some dimensions of their practice.

Interpreting Learning Outcomes

Intriguing as this evidence is of learning from Board certification, we need to look further to uncover some important differences in how teachers experienced the process. For example, one study found evidence that professional growth from National Board certification resulted secondarily in changes in classroom behavior of candidates (Kowalski, et al., 1997). Another account, however, portrays teachers regarding Board certification as a bureaucratic process undertaken for extrinsic reasons (i.e., additional income) (Ballou, 2003). Consistent with this view, teachers might simply “jump through hoops” while de-coupling the process from their teaching, much as they are reported to do with university-based Master’s degrees coursework (Murnane, Singer, Willett, Kemple, & Olsen, 1991). Such a result would hardly constitute a worthwhile form of professional development. Did we see this phenomenon in our interviews? Alternatively, as with other forms of learning, might influences of certification yield delayed effects? Some research on pre-service teacher education indicates such results (e.g., Grossman, et al., 2000). In this reckoning, teachers may mull over what they have learned, only gradually introducing new ideas into their practice. Such possibilities suggest that there may be a range of outcomes, not all of one kind. Indeed, this is what we found. In particular, we identify three qualitatively different learning responses, which we label “dynamic,” “technical,” and “deferred.”

“Dynamic learning” refers to self-reports of immediate, meaningful change in a teacher’s beliefs, understandings, and actions in the classroom. Roughly half of all teachers interviewed post-intervention fell into the dynamic learning category¹⁸. For example, Jasmine, a teacher from Tennessee, provides a glimpse of this when she states:

The analytical part of learning doesn’t just end with sending in your paperwork to National Certification. It’s something then that you can’t help but continue to do. The questions that I had to answer in written form pop into my head now all the time. (Group 2A–Post, Teacher #3)

¹⁸ We provide rough estimates of the proportions of teachers falling into the “dynamic,” “technical,” and “deferred” categories but note that coding on this aspect of the study was carried out by the lead author alone, so no reliability estimates are reported. Consequently we underscore the proper caution here.

Lynn, a teacher from Florida, presents another example of this orientation when she claims that the experience of serious reflection from the certification experience is “carried over and you are just different. You think about it [teaching-learning] differently” (Group 1, Teacher #12).

Both Lynn and Jasmine seem to have internalized the National Board framework of reflection and action. The skills they acquired pertaining to the act of reflecting on practice and student learning persist well after the certification process is over. Jasmine “can’t help” but continue with the same approach that was conveyed by repetition and focus through the portfolio prompts. Lynn describes the effects of learning as ‘carrying over’ into the current semester.

Many comments from teachers focus on the new skills and knowledge gained from certification and the immediate impact on their teaching of scientific inquiry. For example, Shirley, a teacher from New York, is still involved with portfolio-like activities. She states,

I’m still doing—you know—more inquiry based labs, more projects. My classroom has become much more student centered and more—having students be more analytical and critical thinkers. (Group 1, Teacher #36)

She has learned an appreciation for a more ‘student centered’ approach, for more ‘analytical and critical’ thinking in the students. A teacher from California provides another example. She describes how gains in self-confidence provided the necessary strength to try a new type of lesson with her class. She says,

The skills that I gained last year in focusing on good teaching and the things that they make you focus on in your portfolio—I gained a lot of confidence in those skills and they became more natural and easy for me. Whereas I might not have taken that risk had I not gone through the process. (Group 1, Teacher #40)

Connie, another teacher, hints that prior to certification the skills in question were present, but remained weak or underdeveloped. These are skills and dispositions associated with conducting a more student-centered class where the teacher does not dispense knowledge, but helps students create their own understandings. The teacher says that these skills became “more easy” which implies that they were present before certification, which provided an opportunity to develop them further. She is making a direct connection between National Board certification in the previous semester and new lessons implemented in the current term.

Dynamic learning might be described with another teacher’s description of learning from National Board certification. Paul, a teacher from Massachusetts, says,

If you can get a kid to think about a subject that you are teaching—if you can get a kid to internalize it—then he’ll have it forever. It’s the same thing I think with adults. (Group 1, Teacher #4)

Dynamic learning may be ‘internalized’ but the important element is that the teacher acts upon that new knowledge, skill, or understanding to consciously and deliberately try to improve the learning experience in his class. How long this improved teaching will continue is open to question. “Forever,” as Paul suggests? Or, might he revert to old ways after a few weeks or months? Longitudinal investigations are needed to answer such questions.

A second interpretation, “technical learning” indicates an emphasis on acquiring techniques useful in obtaining certification, but does not necessarily carry over into teaching itself. Teachers are learning how to be better candidates for National Board certification, but not necessarily how to be better teachers. An analysis of teacher comments suggest that roughly one quarter of teachers interviewed post-intervention fell into this category. With technical learning, the certification process is essentially de-coupled from the teacher’s actual practice in the classroom. Chris, a teacher from South Carolina, reflects on this prospect in comments about other teachers whom she has observed:

I think that while for some teachers it’s going to make them better teachers, for some teachers they are going to have to do during that year and not change what

they are really doing—that they are putting on a show and they just do that well. And whether you maintain [that kind of teaching] afterwards is what I question. On how much that’s being done. (Group 1, Teacher #7)

Mathew, a teacher from Virginia echoes this observation when he says,

I haven’t put on a ‘dog and pony show’ this year where I’m inventing all of these terrific lessons that I didn’t have before. (Group 2B, Teacher #31)

Both perceive some of their colleagues as not being honest with the spirit of self-reflection, self-realization, and professional development that is part of the National Board certification process. These ‘dog and pony’ teachers put their efforts into impressing the assessors who evaluate the portfolios. They orient to the certification process, not to the improvement of their own teaching; they respond to the incentives associated with certification rather than to intrinsic motivations for professional development. For example, Margaret from Maryland describes her experience with these words:

I felt like it was more of a—more of an exercise in trying to find out what they were looking for. And so I spent more of my energy doing that than in actually reflecting on my own practice and writing about it. (Group 2B, Teacher #33)

Margaret’s comment suggests that the requirements of certification interfered with the quality of her experience. She interprets the tasks as pleasing the assessors rather as addressing her genuine learning needs or those of her students.

The emphasis here is on developing good strategies for passing the tests, for writing the way the “readers want you to write;” and for picking up tips on how to successfully manage the process. Such learning is only (at best) tangentially related to classroom practice. Another teacher, Sarah from Florida, contributes an additional insight. She says,

I think a lot of it is, what hoops can you jump through and how well, how good are you at writing, at saying, what you need to say to prove based upon the rubric? Are you good at being able to work through that? If you are, that’s great.

But someone might be a very good teacher and just based upon what they submit, it might not be evidence of what they are doing. (Group 1, Teacher #29)

Accomplished teachers, according to this view, may not be good at proving they are accomplished. National Board certification may be too restrictive in format and style to fit some teacher’s modes of communication. Sarah’s comment also underlies the importance of the technical knowledge needed to succeed at communicating one’s practice effectively through the portfolio. The choice of artifacts, decisions regarding lessons, actions taken during the taping of a class, and the details of how to analyze student work, all contribute to technical learning. While all candidates to a certain degree need to address the demands of the technical components of certification, the problem seems to arise when the technical learning overshadows the intended emphasis on self-reflection and student learning.

For one teacher, this problem had no resolution. Jerry, a science educator from Florida, came up with a scientific metaphor in his response:

Frankly, I found that it was so difficult. It’s sort of like Heisenberg’s Principle. You can either know where the electron is and not know what it’s doing or know what it’s doing and not know where it is. I could either teach as an effective teacher or I could go through this procedure to prove I’m an effective teacher. (Group 2A–Post, Teacher #19)

The implication here is that, “I can’t do both.” The teacher could not devote the time and energy required to communicate the quality of practice effectively through the construction of the portfolio and teach with the same intensity that he was accustomed to prior to the demands of the certification process.

So for some teachers the certification process actually seemed to be a diversion from their teaching—in favor of “jumping through hoops”—rather than a stimulus for reflecting on or learning about teaching. So conceived, learning from certification had value that was narrowly instrumental at best.

Yet a third possibility presented itself, that a genuine form of learning related to good teaching might be deferred to a time when teachers had more opportunity to reflect and to consider how to use what they had learned from the process. Such *deferred learning* holds out the possibility for genuine influences on practice at some future time. Approximately one quarter of teachers interviewed post-intervention fell into this category. Sharon, a teacher from Wyoming, illustrates this prospect when she says,

Now that I've completed everything—it's all turned in—the stressful parts of it are gone; and I have the opportunity to sort of look back and observe and see how some of the things have been incorporated into my teaching—I think that it was particularly useful. (Group 2A—Post, Teacher #12)

Sharon is describing how the stress associated with technical learning interferes with the possibility of dynamic learning. Her remark that, “I have the opportunity to sort of look back and observe and see how some of the things have been incorporated into my teaching” is a meta-cognitive act, where the teacher thinks about how she thinks about her practice. This teacher is not claiming that she is more reflective, more student-centered, more focused on planning or assessment or engagement. No specific outcome is identified—yet. The teacher sees this kind of analysis as removed from the “self”. At this point, the teacher is unaware of how her practice may have changed. She needs distance from the intense experience of certification and time to examine her practice to recognize differences in values, decision-making, and beliefs that may have arisen in response to the certification process.

Deferred learning also may be related to uncertainty. To the extent that a teacher is uncertain if learning took place as a result of National Board certification, the possibility exists that a learning outcome might be realized some time in the future. In describing whether certification affected his practice, Mathew, a teacher from Virginia, comments, “I'm not sure that it's changed, at this point, how I taught” (Group 2B, Teacher #31). By qualifying the statement with “at this point,” he leaves open the possibility that lessons learned from the experience may be realized at a future point in time. As teachers reflect on the process after the fact, many may be considering how to make use of things they learned, exploring discrepancies between their preferred methods and what they perceive to be preferred by the National Board. Such reflection can move in two directions—to reaffirm a teacher's commitment to her existing preferences, or to provoke some change. To the extent that certification unsettled some teachers' thinking, it holds the possibility of ushering change—but only the possibility.

Discussion

In the current climate of policy debate, single study results are often promoted, sometimes in the press, sometimes by policy advocates, as definitive resolutions to complex questions. We reject such oversimplification in drawing conclusions from this study for policy and practice. Instead, we frame the implications along these lines, first for policy, then for the practice of professional development.

Teacher learning has become a policy variable in the context of a wide range of efforts to improve education. From its inception, National Board certification was promoted as a professional development opportunity, part of the ongoing effort to professionalize teaching as a policy strategy.

If standards serve as a critical carrier for the knowledge base of teaching, then standards-based practice clearly must become a hallmark of teaching if it is to realize the promise of professionalism. The certification process involves many of the hallmarks of effective professional development, but chiefly as it represents the use of standards in practice. What teachers learn from the process is to evaluate their own practice in the light of objective, external standards. But this process may or may not in fact yield the outcomes that the National Board and its proponents have sought. This study provides one—and the first—indication that teachers are undertaking worthwhile learning, bolstering the position of the advocates for professionalism as a policy choice. The immediate implication is that public investment in Board certification is warranted. Certainly, however, we underscore how slender is the evidence presented here. Limitations on this study include the restriction to one certification area; “learning” measured via a telephone rating task; reliabilities in the low end of the range; a moderate overall effect size; and qualitative evidence indicating that some but not all teachers took positive advantage of the process to improve their teaching. These limitations urge caution in any sweeping conclusions, but we choose to emphasize the generally positive results, notwithstanding the limitations noted.

Turning next to implications for professional development, research is just beginning to explore what teachers are learning from professional development. According to Wilson and Berne (1999), research in this area has yet to “identify, conceptualize, and assess” what teachers are learning. Determining the effects of professional development on teaching and learning is notoriously difficult. Consider some of the problems. Teachers might acquire new knowledge or skills yet choose not to deploy them in their practice. Or they might make changes initially, but revert gradually to old ways. Or the changes they make might not enhance their practice. Some prior scholarship reveals teachers importing only certain aspects of reforms into their teaching with uncertain overall and long-term effects. Describing such problems, investigators have resorted to such metaphors as “hybrids” to indicate the distinctive mix of grafting the new onto the old (see, for example, Cohen, 1990; Cuban, 1993). Furthermore, “change” does not automatically mean “improvement.” The latter term requires a value judgment as well as an empirical result. In consequence, many problems attend any summary conclusions about teacher learning from professional development experiences. The study reported here cannot resolve such issues authoritatively. Results require qualified interpretation, which we offer along these lines.

First, the National Board standards represent a broad consensus within the science education community that, in Joseph Schwab’s evocative terms, Science is a “narrative of inquiry,” not simply a “rhetoric of conclusions” (Schwab, 1974). Instruction that aspires to teach students the methods of science is a critically important issue at the dawn of the 21st century. Consequently, the underlying values represented by the National Board standards constitute a professional consensus; what these standards “teach” about science instruction are eminently defensible.

Then, the overall effect size of 0.47 derived from multiple comparisons falls within the moderate to strong range based on several comparative criteria. In the field of science education, for example, a meta-analysis from the early 1980s serves as one comparison. Enz, Blecha, and Horak (1982) reviewed research projects that investigated the effects of professional development in science education on participating teachers and/or their students. In the sixteen studies gathered from 1973–1980, the overall average effect size for science in-service projects was 0.84. Our result falls below this average, which would be regarded as on the high end of the range.

Turning to our qualitative data, this study is also significant in identifying which aspects of the National Board’s standards appeared to exert the greatest influence. Other studies will be required to confirm these results, including examination of other certificate areas, but we offer one conjecture by way of explanation. A certain logic would suggest that the greatest learning is likely to occur where there was the largest discrepancy between “standards-based practice” and pre-existing

teaching. If this is roughly true, then our study indicates that many secondary science teachers are not emphasizing principles and practices of the scientific method in their instruction, nor using assessment feedback in ongoing instructional decision-making. This conjecture points to these aspects of practice as needful of improvement in both pre-service and in-service education for science teachers.

Finally, the study clearly uncovered a mix of what we called “dynamic,” “technical,” and “deferred” learning. This too seems quite plausible. Some teachers might regard Board certification as a genuine learning opportunity, others might undertake it for the extrinsic rewards, and still others might learn from the process in a gradually evolving manner. Mixed motives and outcomes are more nearly the norm in human affairs than singular or pristine results. In fact, the different categories of learning described in this study support the conclusion that Board certification provides the opportunity for teachers to learn about specific aspects of their work. How that learning impacts practice remains unclear however.

Teachers in this study demonstrated significant learning in the areas of Scientific Inquiry and Assessment regardless of whether they were successful at achieving Board certification or the characteristics of their particular school setting. Therefore, it would appear that the benefits of Board certification go beyond the immediate financial rewards successful candidates receive to take the form of improved knowledge and understanding of science instruction for both those who achieve and those who do not achieve certification. If teacher learning is considered an important component to improving teacher quality and ultimately student achievement, then these results point to the possibility that the process of Board certification may positively impact the quality of instruction (as defined by the National Board) and students’ learning experiences regarding two vital areas of instruction. Further research on this relationship is needed to pinpoint the degree to which science teaching improves, the duration of those changes, and the impact of changes in practice upon student achievement. On balance, though, we are inclined to read the overall pattern of results in support of National Board certification as a worthwhile form of professional development. The caveats, as always, are important, but so is the preponderance of the evidence.

References

- Areglado, N. (1999). I became convinced. *Journal of Staff Development*, 20(1), 35–37.
- Bailey, D. L. & Helms, R. G. (2000). *The national board certified teacher. Fastback 470*.
Bloomington, IN: Phi Delta Kappan Education Foundation, pp 8–47.
- Ball, D., & Cohen, D. (1995). *Developing practice, developing practitioners: Toward a practice-based theory of professional education*. New York: National Commission on Teaching and America's Future.
- Ballou, D. (2003). Certifying accomplished teachers: A critical look at the National Board for Professional Teaching Standards. *Peabody Journal of Education*, 78(4), 201–219.
- Benz, J. (1997). Measuring up: A personal journey through National Board Certification in Art. *Art Education*, 50(5), 20–24, 49–50.
- Bond, L., Smith, T., Baker, W., & Hattie, J. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*.
Greensboro, NC: Center for Educational Research and Evaluation at the University of North Carolina at Greensboro.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(8), 3–15.
- Burroughs, R., Schwartz, T. A., & Hendricks-Lee, M. (2000). Communities of practice and discourse communities: Negotiating boundaries in NBPTS certification. *Teachers College Record*, 102(2), 344–374.
- Campbell, D. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*.
Hopewell, NJ: Houghton Mifflin Company.
- Campbell, D. & McCormack, T. (1957). Military experience and attitudes toward authority. *American Journal of Sociology*, 62, 482–490.
- Cavaluzzo, L. (2004, November). *Is National Board Certification an effective signal of teacher quality?* Alexandria, VA: The CNA Corporation.
- Chase, B. (1999). *NEA Today*, 2(99), 17.
- Cohen, D. (1990). Revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 311–330.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York, Academic Press.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin Company.

- Crawford, J. & Impara, James. (2001, 4th edition). Critical issues, current trends, and possible futures in quantitative methods. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 133–173). Washington, D.C.: American Educational Research Association.
- CREDE, (2002). *The five standards for effective pedagogy*. Center for Research on Education, Diversity & Excellence. Retrieved May 23, 2004 from: <http://www.crede.ucsc.edu/standards/standards.html>.
- Cuban, L. (1993, 2nd Edition.). *How teachers taught: Constancy and change in American classrooms, 1890–1990*. New York: Teachers College Press.
- Darling-Hammond, Linda (2000). Teacher quality and student achievement: a review of state policy evidence. *Educational Policy Analysis Archives*, 8(1). Retrieved February 10, 2004, from <http://epaa.asu.edu/epaa/v8n1/>.
- Darling-Hammond, Linda. (2002, September 6). Research and rhetoric on teacher certification: A response to "Teacher Certification Reconsidered," *Education Policy Analysis Archives*, 10(36). Retrieved February 10, 2004, from <http://epaa.asu.edu/epaa/v10n36.html>.
- Darling-Hammond, L. and McLaughlin, Ida (1996). Policies that support professional development in an era of reform. In M. W. O. McLaughlin, Ida (Ed.), *Teacher Learning* (1st ed.) (pp. 202–218). New York: Teachers College Press.
- Educational Testing Service. (1999). *Technical analysis report*. Princeton, NJ: National Board for Professional Teaching Standards.
- Enz, J., Blecha, M. & Horak, W. (1982). *Review and analysis of reports of science inservice projects: Recommendations for the future*. Paper presented at the annual Meeting of the National Science Teachers Association, April, 1982, Chicago, IL.
- Floden, R. (2001, 4th edition). Research on effects of teaching: A continuing model for research on teaching. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 3–16), Washington, D.C.: American Educational Research Association.
- Franke, M., Carpenter, T., Fennema, E., Ansell, E., & Behrend, J. (1998). Understanding teachers' self-sustaining, generative change in the context of professional development. *Teaching and Teacher Education*, 14(1), 67–80.
- Franke, M., Carpenter, T., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, 38(3), 653–689.
- Gallagher, J. (2000). Teaching for understanding and application of science knowledge. *School Science and Mathematics*, 100(6), 310–18.
- Gardiner, S. (2000). I leave with more ideas than I can ever use. *Journal of Staff Development*, 21(4), 14–17.

- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–45.
- Goldhaber, D. & Anthony, E. (2004). Teacher quality and student achievement. Urban Diversity Series, *Center on Reinventing Public Education* Retrieved November 6, 2004, from http://www.crpe.org/workingpapers/pdf/NBPTSquality_report.pdf.
- Glaser, B. G. (1995) A look at grounded theory: 1984–1994. In Glaser, B. G. (Ed.). *Grounded theory, 1984–1994* (pp. 3–17). Mill Valley, CA: Sociology Press.
- Griffin, K. (2003, February 8). Bonuses going to teachers with national certification, *The Daily Oklahoman*, p. 1.
- Grossman, P. L., Valencia, S. W., Evans, K., Thompson, C., Martin, S., & Place, N. (2000). Transitions into teaching: Learning to teach writing in teacher education and beyond. *Journal of Literacy Research*, 32(4), 631–662.
- Hargraves, A. (1995). Development and desire: A postmodern perspective. In T. R. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 9–34). New York: Teachers College, Columbia University.
- Hawley, Willis D. & Valli, Linda (1999). The essentials of effective professional development. In Linda Darling-Hammond and Gary Sykes (Eds.), *Teaching as the learning profession* (pp. 127–150). San Francisco: Jossey-Bass.
- Haynes, D. D. (1995). One teacher's experience with National Board assessment. *Educational Leadership*, 52, 58–60.
- Homack, S. R. (2001). Understanding what ANOVA post hoc tests are, really. *Paper presented at the annual meeting of the Southwest Educational research Association*, February 1–3, New Orleans, LA.
- Huberman, M. (1993). The model of the independent artisan in teachers' professional relations. In J. Little and M. McLaughlin (Ed.), *Teachers' work: Individuals, colleagues, and contexts* (pp. 11–50). New York: Teachers College Press.
- Ingvarson, L. (1998). Professional development as the pursuit of professional standards: The standards-based professional development system. *Teaching and Teacher Education*, 14, pp. 127–140.
- Jenkins, K. (2000). Earning board certification: Making time to grow. *Educational Leadership*, 57, 46–48.
- Jimenez, P. M. (1999). Psychosocial intervention with drug addicts in prison. Description and results of programme. *Intervencion Psichosocial*, 8(2), 233–250.

- Juin-jen C., Chung-cheng, L. & Ching-chong, L. (1999). The employment and wage effect of shifting to an indirect tax when firms set wages efficiently, *Economic Record*, 75, 156–166.
- Kennedy, M., Ball, D., McDiarmid, G., & Williamson, X. (1993). *A study package for examining and tracking changes in teacher's knowledge*. East Lansing, MI: National Center for Research on Teacher Education, Michigan State University.
- Killeen, K., Monk, D., Plecki, M. (2002). School district spending on professional development: Insights available from national data, 1992–1998. *Journal of Education Finance*, 28, Summer, 25–50.
- Knapp, M. (2003). Professional development as a policy pathway. In R. Floden (Ed.). *Review of Research in Education*, 27 (pp. 109–158). Washington, D.C.: American Educational Research Association.
- Koprowicz, C. L. (1994). *What state legislators need to know about the National Board for Professional Teaching Standards.*, Denver, CO: National Conference of State Legislatures
- Kowalski, K., Chittenden, E., Spicer, W., Jones, J., & Tocci, C. (1997). *Professional development in the context of National Board for Professional Teaching Standards certification: Implications beyond certification*. Paper presented at the annual meeting of the American Educational Research Association, March 24–28, Chicago, IL. (ERIC Document Reproduction Service No. ED 412257).
- Lafferty, B. D. (1998). *An empirical investigation of a leadership development program*. Washington, DC, 248 leaves. (Unpublished Thesis, Ed.D. George Washington University).
- Leef, G. C. (2003). National Board Certification: Is North Carolina getting its money's worth? *A Policy Report from the North Carolina Education Alliance*. Retrieved February 6, 2005, <http://www.johnlocke.org/acrobat/policyReports/leefcertification2.pdf>.
- Little, J. W. (1993). Teacher's professional development in a climate of education reform. *Education Evaluation and Policy Analysis*, 15(2), 129–151.
- Little, J. W. (1997). *Excellence in Professional Development and Professional Community* (Working Paper). Berkeley, CA: University of California, Berkeley for US Department of Education, Blue Ribbon Schools Program.
- Mahaley, D. (1999, September/October). One teacher's account. *The Clearing House*, p. 5.
- Manouchehri, A. (2001). Collegial interaction and reflective practice. *Journal of the Association of Teacher Educators*, 22, 86–97.
- Marriot. (2001, February/March). Increased insight. *Reading Today*, p. 7.
- Merriam, S., & Simpson, E. (1995). *A Guide to research for educators and trainers of adults* (2nd ed.). Malabar, FL: Krieger Publishing Company.

- Mullens, J. E., Leighton, M. S.; Laguarda, K. G., O'Brien, E. (1996). *Student learning, teaching quality, and professional development: Theoretical linkages, current measurement, and recommendations for future data collection*. Working Paper Series. Washington, DC: Policy Studies Associates.
- Murnane, R. J., Singer, J. D., Willett, J. B., Kemple, J. J., & Olsen, R. J. (1991). *Who will teach? Policies that matter*, Cambridge, MA: Harvard University Press.
- National Board for Professional Teaching Standards. (1991). *Toward high and rigorous standards for the teaching profession* (3rd ed.). Washington, DC: Author.
- National Board for Professional Teaching Standards. (1997). *Instruction manual for adolescent and young adult (AYA) science certification*, Fairfax, VA: Author.
- National Board for Professional Teaching Standards. (2001a). *The impact of national board certification on teachers (Survey results)*. Arlington, VA: Author.
- National Board for Professional Teaching Standards. (2001b). *I am a better teacher (Survey results)*. Arlington, VA: Author.
- National Board for Professional Teaching Standards. (2001c). *Adolescence and young adult science standards*. Retrieved December 1, 2005, from http://www.nbpts.org/pdf/aya_science.pdf.
- National Board for Professional Teaching Standards. (2003). *Q & A: Questions and answers for teachers about national board certification*. Arlington, VA: Author.
- National Board for Professional Teaching Standards. (2004a) *State and local incentives for certification*. Retrieved on August 8, 2004 from <http://www.nbpts.org/about/state.cfm>.
- National Board for Professional Teaching Standards (2004b). *The 5 core propositions of accomplished teaching*. Retrieved August 8, 2004, from http://www.nbpts.org/standards/five_core.html.
- National Commission on Teaching and America's Future (1996). *What matters most: Teaching for America's future*. New York: Teachers College, Columbia University.
- Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002). A cost framework for professional development. *Journal of Education Finance*, 28(1), 51–74.
- Reichardt, R. (2001). *Toward a comprehensive approach to teacher quality*. Policy Brief No. 13, Mid-Continent Research for Education and Learning, Aurora, CO. Retrieved July 20, 2005, from http://www.mcrel.org/pdfconversion/policybriefs/pb_towardacomprehensive.html.
- Roden, J. (1999). Winners and winners: What I learned from not earning National Board Certification. *Teaching and Change*, 6(4), 416–419.

- Rotberg, I. C., Futrell, M. H., & Lieberman, J. M. (1998). National Board certification: Increasing participation and assessing impacts. *Phi Delta Kappan*, 79, 462–466.
- Rudner, L. M. (1992). Reducing errors due to the use of judges. *Practical Assessment, Research & Evaluation*, 3(3). Retrieved August 8, 2004, from <http://ericae.net/pare/getvn.asp?v=3&n=3>.
- Schwab, J. J. (1974). Decision and choice: The coming duty of science teaching. *Journal of Research in Science Teaching* 11(4), 309–317.
- Sclafani, Susan (2005). *The U.S.-China perspective*. Presentation to the US-China Conference on Secondary Education Reform, Michigan State University, East Lansing, MI, July 18, 2005.
- Stronge, J. H. (2002). *Qualities of effective teachers*. Alexandria, Va.: Association for Supervision and Curriculum Development.
- Shavelson, R., Webb, N., & Hotta, J. (1987). The concept of exchangeability in designing telecourse evaluations. *CADÉ: Journal of Distance Education/Revue de l'enseignement a distance*. Retrieved November 3, 2004, from <http://cade.athabasca.ca/vol2.1/shavelson.html>.
- Stein, M. K. & Brown, C. A. (1997). Teacher learning in a social context: integrating collaborative and institutional processes with the study of teacher change. In B. Nelson (Ed.), *Mathematics teachers in transition* (pp. 155–191.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sykes, G. (1999). Teacher and student learning. In Linda Darling-Hammond & Gary Sykes (Eds.), *Teaching as the learning profession* (pp. 151–180). San Francisco: Jossey-Bass Publishers.
- Tracz, S., Sienty, S., Todorov, K., Snyder, J., Takashima, B., Pensabene, R., Olsen, B., Pauls, L., & Sork, J. (1995). *Improvement in teaching skills: Perspectives from National Board for Professional Teaching Standards field test network candidates*. Paper presented at the annual meeting of the American Educational Research Association, April, 1995, San Francisco, CA.
- Vandevoort, L. G., Amrein-Beardsley, A. & Berliner, D. C. (2004, September 8). National board certified teachers and their students' achievement. *Education Policy Analysis Archives*, 12(46). Retrieved November 6, 2004, from <http://epaa.asu.edu/epaa/v12n46/>.
- Wiebke, K. (2000, Nov/Dec). My journey through National Board certification. *Instructor*, 110, 26–27.
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In A. Iran-Nejad & P. D. Pearson (Eds), *Review of research in education* (Vol. 24, pp. 173–209). Washington, DC: American Education Research Association.

- Yip, J. (2002, May). Two short notes on statistics. *Porcupine! On line Journal*, 25, May 2002.
Retrieved November 10, 2004, from <http://www.hku.hk/ecology/porcupine/por25/25-misc-stat.htm>.

Appendix A

Study Limitations and Caveats

The limitations on this study are not numerous, but they are important. Most notable was the inconsistency in recruitment protocols for each of the three groups. In particular, the identification of subjects for Group 2 was problematic. The main concern was the delayed group formation resulting in candidates participating in the study with differing degrees of experience with the intervention. The lack of a true 'pre-status' for Group 2A was shown to be problematic on two accounts.

First, institutional and procedural obstacles hampered the timely and efficient recruitment of participants for Group 2A–Pre. The delays resulting from these obstacles resulted in a less than ideal 'pre-intervention' status for data collection. We suspect that as teachers spent more time with the intervention, their assessed scores in this study improved. To explore this possibility, we performed a series of correlation studies that compared the observed data collected and the status of the candidate (amount of experience with the intervention) at the time of the interview. The results indicate a weak to moderate relationship between the two variables ($r = 0.40$). The relationship was strong enough to support the conclusion that had the interviews for Group 2A–Pre occurred at an earlier point or prior to the certification process, most likely the data collected from Group 2A–Post would have resulted in greater pre to post gains and a larger effect size for Group 2A.

Another way this problem may have been avoided would have been to increase the size of Group 2. In this study, the pre and post observations of Group 2A played a more vital role in this analysis than any of the other observations, yet it was the smallest group, with only 18 teachers. Groups 1 and 3 each had 40 teachers. To address this problem, it would have been better to increase the size of Group 2 to 60 teachers allowing for 30 each in Group 2A and Group 2B. The Group 2 increase would then be balanced by a reduction in the size of Groups 1 and 3 to 30 teachers each. Such an alteration would have maintained the same number of teachers in the study, but would better reflect the relative influence each observation contributes to the overall analysis.

Consistent use of random selection and random assignment also would have improved this study. Only Groups 1 and 2B were randomly selected for data collection. Due to unavoidable time pressures, the pre-groups were interviewed on a first come, first served basis. Improved recruitment procedures would allow for more consistent use of random selection for all three groups and random assignment within Group 2 to either 2A or 2B. Due to the inconsistent use of random selection and the quasi-experimental nature of this research, the conclusions from this study are generalizable only to secondary science teachers who volunteer for National Board certification. Hence, external validity is limited to this population of teachers.

Finally, inter-rater reliability was less than ideal for measurement instruments. Though a reliability of .458 between assessors is considered low to moderate in social research, there is reason to believe that with greater resources of time and money this reliability measure could be improved. Had assessors met together for a weekend to wrestle with issues specific to this research, many of the observed inconsistencies could in all likelihood have been avoided. For example, during the calibration process, assessors practice scores were compared against a standard. If they consistently scored above the standard, then the trainer and assessor discussed the issue so that the appropriate adjustments could be made. To make this process more effective, the scoring practices of the assessor need to be more intensive before the collection of data, then revisited periodically throughout the study to maintain consistency between all three assessors. The best way to improve

agreement among raters on such a complex assessment activity is by investing more time in developing a common and agreed upon framework.

Appendix B

Inter-rater Reliabilities

Inter-rater reliability is a measure of the degree to which raters agree in their assessment of each standard for each candidate. An examination of Figure B-1 reveals some important observations. First, measures of different standards have different levels of reliability, though these differences occupy a rather small range. With a maximum average reliability of 0.632 and a minimum average reliability of 0.259, the mean value of 0.458 demonstrates that more measures fall toward the lower than the higher end of the range.

Table B-1 provides descriptive statistics comparing the ratings of the three assessors. Each measure is consistent across all three individuals with the exception of Assessor 1 who scored every transcript while the Assessors 2 and 3 each scored slightly more than half of the total number. It is important to note that if Assessor 1's data is removed from the study, an analysis of data from Assessors 2 and 3 yield nearly identical results, i.e. Scientific Inquiry and Assessment remain the two standards that demonstrate significant pre to post intervention gain.

Table B-1: Assessor Descriptive Statistics

| Measure | Assessor #1 | Assessor #2 |
|--------------------------|-------------|-------------|
| Mean | 2.76 | 2.59 |
| Standard Error | 0.06 | 0.07 |
| Median | 2.82 | 2.62 |
| Mode | 2.85 | 2.52 |
| Standard Deviation | 0.64 | 0.57 |
| Sample Variance | 0.42 | 0.32 |
| Kurtosis | -0.52 | -0.48 |
| Skewness | -0.16 | -0.06 |
| Range | 2.79 | 2.40 |
| Minimum | 1.21 | 1.42 |
| Maximum | 4.00 | 3.83 |
| Sum | 364.77 | 170.75 |
| Count | 132.00 | 66.00 |
| Largest (1) | 4.00 | 3.83 |
| Smallest (1) | 1.21 | 1.42 |
| Confidence Level (95.0%) | 0.11 | 0.14 |

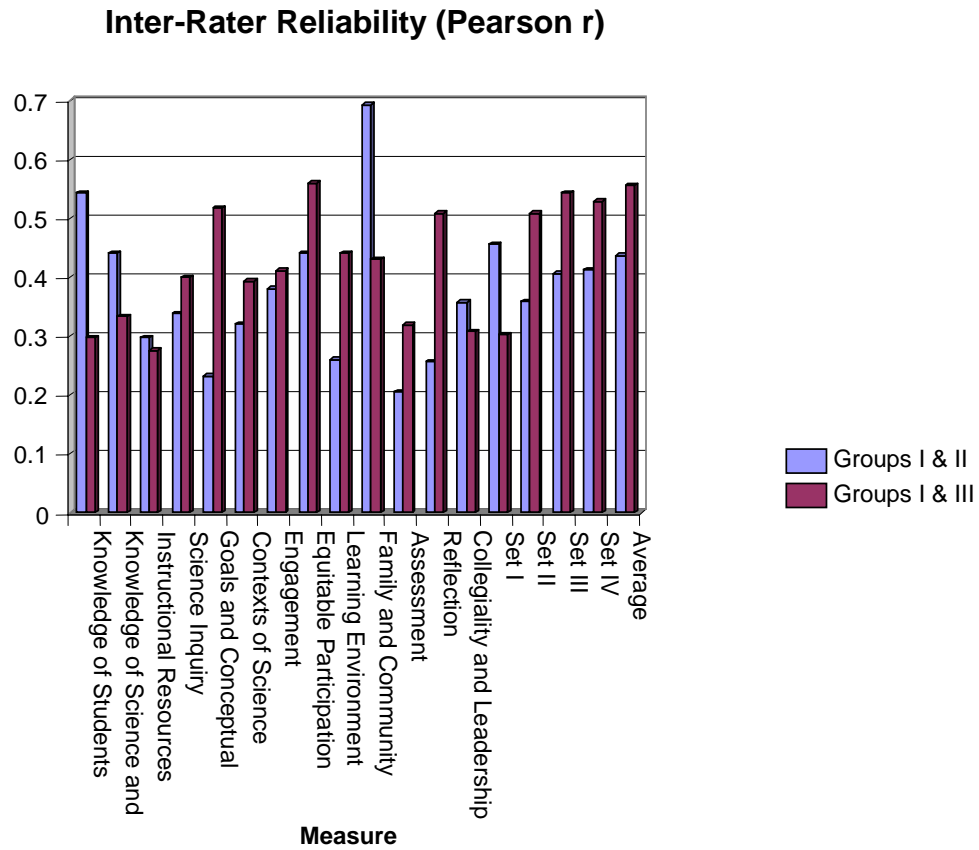


Figure B-1
Comparison of Interrater Reliability (Pearson r) across standards

Appendix C

Processing Example

To make transcripts easier to understand and to allow for the content of the teacher's response to be the sole focus of the assessment, transcripts were processed to make them all appear visually the same and to remove any evidence of the interviewer.

Original Version

- (1) Do you think that the teacher has been effective in facilitating and supporting meaningful scientific discussion where students explore and have the opportunity to understand important scientific ideas? What does this teacher do or not do that supports your answer?**

Assessor—"Scientific ideas?"

"Yeah—for example the fact that there—changes in the organisms or fewer or more of one type of organism. They didn't just look for one reason. They talked about all the different reasons and how maybe what they saw didn't explain it—because they were there just briefly. And I thought that was good."

Assessor—"And how did the teacher help support that?"

"Well—he affirmed it. He sort of was like yeah. I thought just sort of re-agreeing with the students."

Assessor—"I'm sorry I couldn't hear..."

"Agreeing or affirming with the students that made those comments."

Final Version

- (1) Do you think that the teacher has been effective in facilitating and supporting meaningful scientific discussion where students explore and have the opportunity to understand important scientific ideas? What does this teacher do or not do that supports your answer?**

"Yeah—for example the fact that there—changes in the organisms or fewer or more of one type of organism. They didn't just look for one reason. They talked about all the different reasons and how maybe what they saw didn't explain it—because they were there just briefly. And I thought that was good. Well—he affirmed it. He sort of was like yeah. I thought just sort of re-agreeing with the students—agreeing or affirming with the students that made those comments."

Appendix D

The NBPTS Standards for Accomplished Teaching in AYA Science

Source: National Board for Professional Teaching Standards, 2001c, pp. 5-6.

1. Preparing the Way for Productive Student Learning

I. Understanding Students—This scale pertains to Teacher Knowledge of students. More specifically, teachers know how students learn, actively get to know students as individuals, and determine students' understandings of science as well as their individual learning backgrounds.

II. Knowledge of Science—This scale pertains to teachers broad and current knowledge of science and science education, along with in-depth knowledge of one of the subfields of science, which they use to set important appropriate learning goals.

III. Instructional Resources—This scale pertains to teachers ability to select and adapt instructional resources, including technology and laboratory and community resources, and create their own to support active student explorations of science.

2. Advancing Student Learning

IV. Science Inquiry—This scale pertains to a teacher's ability to develop in students the mental operations, habits of mind, and attitudes that characterize the process of scientific inquiry.

V. Conceptual Understandings—This scale pertains to teacher's use of a variety of instructional strategies to expand students' understandings of the major ideas of science.

VI. Contexts of Science—This scale pertains to the ability of a teacher to create opportunities for students to examine the human contexts of science, including its history, reciprocal relationship with technology, ties to mathematics, and impacts on society so that students make connections across the disciplines of science and into other subject areas.

3. Establishing a Favorable Context for Student Learning

VII. Engagement—This scale pertains to teachers ability to stimulate interest in science and technology and elicit all their students' sustained participation in learning activities.

VIII. Equitable Participation—This scale pertains to ability of a teacher to take steps that ensure that all students, including those from groups which have historically not been encouraged to enter the world of science, participate in the study of science.

IX. Learning Environment—This scale pertains a teachers ability to create safe and supportive learning environments that foster high expectations for the success of all students and in which students experience the values inherent in the practice of science.

4. Supporting Teaching and Student Learning

X. Family and Community Outreach—This scale pertains to the teacher's ability to proactively work with families and communities to serve the best interests of each student.

XI. Assessment—This scale pertains to a teacher's ability to assess student learning through a variety of means that align with stated learning goals.

XII. Reflection—This scale pertains to a teacher's ability to constantly analyze, evaluate, and strengthen their practice in order to improve the quality of their students' learning experiences.

XIII. Collegiality and Leadership—This scale pertains to a teacher's willingness and ability to contribute to the quality of the practice of their colleagues, to the instructional program of the school, and to the work of the larger professional community.

About the Authors

David Lustick

University of Massachusetts Lowell

Gary Sykes

Michigan State University

Email: David.Lustick@uml.edu

David Lustick is a National Board certified secondary science teacher and an assistant professor of science and mathematics at the University of Massachusetts Lowell's Graduate School of Education where he focuses on issues of professional development, teacher learning, and science education. His research interests examine science education from a pedagogical, international, and policy perspective.

Gary Sykes is a professor of educational administration and teacher education at Michigan State University where he specializes in educational policy relating to teaching and teacher education. His research interests center on policy issues associated with the improvement of teaching and teacher education, on the development of leadership preparation programs, and on educational choice as an emerging policy issue.

EDUCATION POLICY ANALYSIS ARCHIVES <http://epaa.asu.edu>

Editor: Sherman Dorn, University of South Florida

Production Assistant: Chris Murrell, Arizona State University

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Sherman Dorn, epaa-editor@shermamdorn.com.

Editorial Board

Michael W. Apple

University of Wisconsin

Robert Bickel

Marshall University

Casey Cobb

University of Connecticut

Gunapala Edirisooriya

Youngstown State University

Gustavo E. Fischman

Arizona State University

Gene V Glass

Arizona State University

Aimee Howley

Ohio University

William Hunter

University of Ontario Institute of Technology

Benjamin Levin

University of Manitoba

Les McLean

University of Toronto

Michele Moses

Arizona State University

Michael Scriven

Western Michigan University

John Willinsky

University of British Columbia

David C. Berliner

Arizona State University

Gregory Camilli

Rutgers University

Linda Darling-Hammond

Stanford University

Mark E. Fetler

California Commission on
Teacher Credentialing

Richard Garlikov

Birmingham, Alabama

Thomas F. Green

Syracuse University

Craig B. Howley

Ohio University

Daniel Kallós

Umeå University

Thomas Mauhs-Pugh

Green Mountain College

Heinrich Mintrop

University of California, Berkeley

Anthony G. Rud Jr.

Purdue University

Terrence G. Wiley

Arizona State University

EDUCATION POLICY ANALYSIS ARCHIVES
English-language Graduate-Student Editorial Board

Noga Admon
New York University

Jessica Allen
University of Colorado

Cheryl Aman
University of British Columbia

Anne Black
University of Connecticut

Marisa Cannata
Michigan State University

Chad d'Entremont
Teachers College Columbia University

Carol Da Silva
Harvard University

Tara Donahue
Michigan State University

Camille Farrington
University of Illinois Chicago

Chris Frey
Indiana University

Amy Garrett Dikkers
University of Minnesota

Misty Ginicola
Yale University

Jake Gross
Indiana University

Hee Kyung Hong
Loyola University Chicago

Jennifer Lloyd
University of British Columbia

Heather Lord
Yale University

Shereeza Mohammed
Florida Atlantic University

Ben Superfine
University of Michigan

John Weathers
University of Pennsylvania

Kyo Yamashiro
University of California Los Angeles

Archivos Analíticos de Políticas Educativas

Associate Editors

Gustavo E. Fischman & Pablo Gentili

Arizona State University & Universidade do Estado do Rio de Janeiro

Founding Associate Editor for Spanish Language (1998—2003)

Roberto Rodríguez Gómez

Editorial Board

Hugo Aboites

Universidad Autónoma
Metropolitana-Xochimilco

Dalila Andrade de Oliveira

Universidade Federal de Minas
Gerais, Belo Horizonte, Brasil

Alejandro Canales

Universidad Nacional Autónoma
de México

Erwin Epstein

Loyola University, Chicago,
Illinois

Rollin Kent

Universidad Autónoma de
Puebla. Puebla, México

Daniel C. Levy

University at Albany, SUNY,
Albany, New York

María Loreto Egaña

Programa Interdisciplinario de
Investigación en Educación

Grover Pango

Foro Latinoamericano de
Políticas Educativas, Perú

Angel Ignacio Pérez Gómez

Universidad de Málaga

Diana Rhoten

Social Science Research Council,
New York, New York

Susan Street

Centro de Investigaciones y
Estudios Superiores en
Antropología Social Occidente,
Guadalajara, México

Antonio Teodoro

Universidade Lusófona Lisboa,

Adrián Acosta

Universidad de Guadalajara
México

Alejandra Birgin

Ministerio de Educación,
Argentina

Ursula Casanova

Arizona State University,
Tempe, Arizona

Mariano Fernández

Enguita Universidad de
Salamanca. España

Walter Kohan

Universidade Estadual do Rio
de Janeiro, Brasil

Nilma Limo Gomes

Universidade Federal de
Minas Gerais, Belo Horizonte

Mariano Narodowski

Universidad Torcuato Di
Tella, Argentina

Vanilda Paiva

Universidade Estadual Do
Rio De Janeiro, Brasil

Mónica Pini

Universidad Nacional de San
Martín, Argentina

José Gimeno Sacristán

Universidad de Valencia,
España

Nelly P. Stromquist

University of Southern
California, Los Angeles,
California

Carlos A. Torres

UCLA

Claudio Almonacid Avila

Universidad Metropolitana de
Ciencias de la Educación, Chile

Teresa Bracho

Centro de Investigación y
Docencia Económica-CIDE

Sigfredo Chiroque

Instituto de Pedagogía Popular,
Perú

Gaudêncio Frigotto

Universidade Estadual do Rio
de Janeiro, Brasil

Roberto Leher

Universidade Estadual do Rio
de Janeiro, Brasil

Pia Lindquist Wong

California State University,
Sacramento, California

Iolanda de Oliveira

Universidade Federal
Fluminense, Brasil

Miguel Pereira

Catedrático Universidad de
Granada, España

Romualdo Portella do

Oliveira

Universidade de São Paulo

Daniel Schugurensky

Ontario Institute for Studies in
Education, Canada

Daniel Suarez

Laboratorio de Políticas
Públicas-Universidad de
Buenos Aires, Argentina

Jurjo Torres Santomé

Universidad de la Coruña,
España