Characteristics of Teachers and Professional Development that Predict Growth in Life Science Content Knowledge

Thomas R. Tretter

Stephanie B. Philipp

Sherri L. Brown

University of Louisville

NARST annual conference, March 2012, Indianapolis, IN

The purpose of this study was to investigate the relationship between teacher characteristics (such as years experience, grade level certification, and content area certification), professional development (PD) characteristics (such as duration of PD, proportion of teachers at various grade bands, and proportion of participants with no science certification) and growth of scores attained on a science content assessment after a PD experience. The teachers are nested in PD groups, so hierarchical linear modeling (HLM) will be used to adjust for the non-independence of scores.

Literature Summary

Two central goals of American public education policy are preparing and retaining science teachers who are highly qualified to teach and educating students to be proficient in science. A report from the National Academies (2006) linked poor science student achievement with the lack of highly qualified science teachers and recommended improving math and science education by providing high quality summer PD programs for math and science teachers. Recently, the National Research Council (2011, July) released a draft framework for generating new science core content standards that recommended focusing on a “limited set of core ideas in order to avoid the coverage of multiple disconnected topics—the oft-mentioned mile wide and inch deep” (p. 30). In order to develop and retain well-prepared science teachers, districts will need to support teachers in exploring and securing a depth of content knowledge across core ideas. It is important for schools to develop and retain well-prepared science teachers because teacher preparation has a positive impact on student learning (Darling-Hammond, 2000; Monk & King, 1994). Although science content knowledge and the depth of that knowledge are not the only predictors of effective science teaching, they are a major concern for science teachers and are the focus of this study.

Teacher Certification

Student learning standards devised by each state and district identify the science content areas that teachers need to teach. Middle school science teachers, in particular, do not usually teach in only one or two content areas; they teach many topics in the content areas of physical, life, and earth/space sciences as a common expectation. According to the U.S. Department of Education, during the 1999-2000 school year, almost half the middle school science teachers, and nearly two thirds of mathematics teachers did not have a major in their subject. Approximately 20% did not have certification in their subject (U.S. Department of Education, 2004). Recently, Laczko-Kerr and Berliner (2002) compared student achievement in regularly certified teachers' classrooms and in under-certified teachers' classrooms on the Stanford Achievement Test-9 (SAT-9). They found that "students of under-certified teachers make about 20% less academic growth per year" when compared to students of teachers with regular certification (Abstract, ¶ 1).

Teacher Experience

Experts agree that experience is an important factor in teacher effectiveness. Roughly, five to eight years are needed to accumulate the amount of skills and experience necessary to effectively adapt instruction to the variety of student needs and be considered a master teacher (Darling-Hammond, 2000). Rockoff (2004) utilized a random effects meta-analysis approach in measuring differences in teacher effects on elementary students from two districts in a New Jersey county. He found "…evidence that teaching experience significantly raises student test scores…Reading test scores differ by approximately 0.17 standard deviations on average between beginning teachers and teachers with ten or more years of experience" (p. 248). Clotfelter, Ladd and Vigdor (2007) used data collected and maintained by the North Carolina

Education Research Center to determine the relationship between teacher characteristics and fifth grade student achievement. The researchers determined that teacher experience and licensure test scores were the only teacher characteristics that predicted student achievement. “For the typical student, the benefit from having a highly experienced teacher is approximately one-tenth of a standard deviation on reading and math test scores” (p. 27).

Measurement of Teacher Science Content Knowledge

Tretter, Brown, Bush, Saderholm, and Holmes (under review) have developed an assessment, the Diagnostic Teacher Assessment of Mathematics and Science (DTAMS), to measure middle school teachers’ science content knowledge. The purpose of this assessment is to help school administrators, researchers, PD providers, and teacher educators characterize the science content knowledge (Saderholm & Tretter, 2008) and depth of content knowledge of middle school science teachers. This assessment could show teacher strengths and weaknesses in knowledge of specific science content areas. This assessment could be used before and after content area PD to demonstrate growth and to direct future PD.

The assessment measures breadth of knowledge, with a grain size designed to sample across the full spectrum of life sciences that middle school teachers are expected to know. Thus the instrument sensitivity to a PD program that focused on only one aspect of life science – e.g. evolution – may be limited. The assessment measures depth of knowledge by examining different types of understanding. The terms of knowledge types used for these assessments were synthesized from the work of Li and Shavelson (2001) and Shulman (1986). They include: (a) declarative (recall of science content facts), (b) inquiry (skills of doing science and process skills), (c) schematic (understanding of fundamental concepts, principles, theories and the relationships between ideas), and (d) pedagogical (pedagogical content knowledge (PCK) as defined by Shulman (1986) and focused on the teacher task of directly addressing student misconceptions.

Internal consistency was determined by computing Cronbach's alpha for all six versions of the assessment. All versions of the life science assessments had Cronbach's alpha coefficients greater than .80 (Tretter, et al., under review), which is desirable (Urbina, 2004). All assessments contained the same number of items (20 multiple choice and 5 open response questions).

Research Questions

1. What characteristics of individual teachers in a PD program predict growth in life science content knowledge?
2. What characteristics of PD groups predict growth in life science content knowledge?

Methodology

Design

A number of sites implementing a variety of professional development (PD) programs across the U.S. chose to use the life science DTAMS instrument as part of their measurement of teachers' content knowledge. Teachers were given the DTAMS as a pretest prior to participation in PD group activities, and at the conclusion of the PD an equivalent alternate version of the life science DTAMS was given as a posttest to assess science knowledge growth. Because each project independently designed and delivered their PD and we were not involved in this, it is likely that the variety of PD approaches included in our sample would be reasonably representative of the spectrum of PD programs that occur across the U.S. This study also used teacher demographics that had been reported by the teacher at the time of testing.

Sample

For this analysis, 315 teacher participants grouped in 19 independent PD groups served as the sample. On average, the life science content gain was less than one point (0.94 points) across all teachers in the study. The average post-PD score of 16.61 out of 25, with a minimum of 5 and a maximum of 24 (see Table 1), suggested that a ceiling effect had not occurred. Descriptive statistics for the sample are shown in Table 1.

Table 1. *Descriptive Statistics for Sample*

Variable	Mean	Standard Deviation	Min	Max
Post-PD score	16.58	3.83	5	24
Pre-PD score	15.64	5.82	6	24
Years of experience	10.7	8.8	0	42
PD Group-level Descriptives				
Teachers per PD group	16.3	6.9	5	33
Time between pre and post-PD scores (weeks)	17.5	14.5	1	52
Group average years of experience	10.3	3.0	5.7	15.7
Teacher-level Descriptives				
Gender	260 (83%) female	55 (17%) male		
Certification Level	42 (13%) elementary	203 (64%) middle	70 (23%) high	
Certification Content	210 (67%) Not science	64 (20%) General science	33 (10%) Life science	8 (3%) Other science

Participant data collected included the DTAMS pre-PD score and post-PD score on the Life Science assessment and self-reported demographic data such as gender, years of experience, content area certification, and grade level certification (elementary, middle school or high school). If the teacher held a certification in more than one grade level, the highest grade level was used. The science content certification areas were categorized as (a) no science, if the teacher reported no specific science or no general science certification; (b) general science, if teachers reported general science or indicated they were certified to teach the multiple content areas across middle school; (c) life science certification, if teachers reported life science certification and adequate number of life science classes to support that certification; and (d) other science, if teachers reported certification in a particular area or areas that did not include life science, such as chemistry or earth science, and had a concentration of class work in those areas of content certification. These teacher-level variables were shown in the literature above to be predictors of student achievement, therefore they may also be predictors of teacher content knowledge growth after experiencing a PD program, assuming that greater teacher content knowledge translates into greater student achievement.

The data that were collected about the PD experiences were the time between the pre-test and post-test, the average years of experience of the group, the percentage of the group that has no science certification, the percentage of the group that has life science certification, and the percentage of the group that was certified to teach high school. These variables may be indicative of the overall group experience since the PD would likely be designed differently for different groups of teachers. For example, PD for a group of teachers, the majority of whom do not have science content certification, should be a very different experience from that with a group of teachers all certified in a life science area. Moreover, the PD experience for a group of more experienced teachers may be different from a group of relatively novice teachers.

Analysis Procedures

Given the hierarchical structure of the data, teachers within PD groups (McCoach & Adelson, 2010), we used hierarchical linear modeling (HLM) to assess the relationships between both individual teacher and PD group characteristics and life science test score growth. HLM takes into account that outcome data from individuals in groups may not be independent (individuals in the same PD group may share similarities in outcomes based on some feature of the group), resulting in a more correct Type I error rate. Additionally, HLM allows one to model both teacher-level and PD group level data at the same time in order to investigate relationships and interactions among the variables at both levels.

Data Screening. Before starting to build the model using HLM Version 6 (Raudenbush, Bryk and Congdon, 2004), we analyzed the bivariate correlations between the outcome variable (post-PD Life Science Test score) and the teacher level predictors to assess linearity of the relationships (SPSS, 2010). The most positive correlation occurred, as expected, between pre-PD and post-PD Life Science Test scores ($r = .656$, $p < .001$). This indicates that controlling for pre-PD score will be important to parse out the differential effects (if any) of the other predictors.

The mean pre- and post-PD scores were calculated for each PD group to check for any outliers. No pre-PD group mean or post-PD group mean was more than one standard deviation above or below the grand mean, indicating no group outliers in this respect. Eight of the 19 PD groups had significant positive gain scores, and another eight of the 19 groups had a negative gain score, but only one was significant.

Hierarchical Modeling. The outcome variable in the model was the post-PD score. The level-1 (teacher level) variables of pre-PD score (a control variable entered as a fixed variable rather than a random variable) and years experience were centered around the grand-mean. Certification grade band (3 levels) was dummy coded. Content area certification was dichotomously coded as having any type of science certification or not. At level 2 (the PD group level), variables explored in the model included: (a) PD program duration in weeks; (b) the group mean years experience; (c) proportion of teachers in the group who had any type of science certification (dichotomous); and (d) proportion of teachers in the group certified at each grade level band (elementary, middle, high).

The first analysis performed using HLM was a fully unconditional model. This was used to establish a baseline and to determine potential explanatory magnitudes at each level by calculating the intraclass correlation coefficient (ICC) to identify the proportion of variance between the PD groups compared to variance within groups. In building subsequent models, all level 1 variables were first entered as a group, and then variables that were not statistically significant predictors were removed. Level 2 variables were then entered as a group. Random effects maximum likelihood estimation was used for all models, which adjusts for the uncertainty about the fixed effects.

Results

The unconditional model returned an ICC=0.40, indicating that 40% of the total variance in post-PD scores could be attributed to between PD groups (thereby indicating the remaining 60% of total variance was within groups). The first conditional model with only level 1 predictors returned two statistically significant predictors: pre-PD score (a control variable labeled PRE in model below) and high school certification (labeled HIGH in model below).

Individuals' years of experience, whether they had a science certification or not, and both elementary and middle-level certification were all nonsignificant predictors. The two significant predictors explained 34% of the within groups variance. The addition of these two level 1 predictors was also able to explain 21% of the between group variance.

Because there was between group variance left to explain (τ_{00} = 4.67, $p < .05$), level 2 PD group variables were next added to the model. To screen for variables that might potentially add explanatory value to the model, a p-value of 0.1 was chosen as the criterion by which to retain predictors during these exploratory stages. The only level 2 PD group variable that met the $p<0.1$ criterion was the mean number of years of experience of the group (labeled AVE_YEARS in model below) as a fixed effect on the intercept but not on the slopes of PRE or HIGH. This statistically significant level 2 predictor explained 17% of the between group variance. The resulting 2-level model was then:

Level-1 Model

$$POST_{ij} = \beta_{0j} + \beta_{1j}(PRE_{ij}) + \beta_{2j}(HIGH_{ij}) + r_{ij}$$

Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(AVE\_YEARS_{j}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$

Table 2 shows the results of this model.

Table 2. *Final Two-level Model*

Fixed Effects	Coefficient (SE)	t (df)	p
	Model for Post-PD Intercept (β_0)		
Intercept (γ_{00})	12.78 (1.48)	8.63 (17)	<.001
Mean of group years experience (γ_{01})	0.23 (0.13)	1.82 (17)	.087
	Model for Pre-PD slope (β_1)		
Intercept (γ_{10})	0.52 (0.04)	11.73 (292)	<.001
	Model for High school certification slope (β_2)		
Intercept (γ_{20})	0.57 (0.28)	2.07 (292)	.040
Random Effects	**Variance**	**Chi Square (df)**	**p**
Var. in intercepts (τ_{00})	2.07	114.17 (17)	<.001
Var. in schools (σ^2)	5.92		

Discussion

Overall Impact of PD Programs

This sample can be overall characterized as: generally experienced teachers, mostly middle school certified, and many not having any type of science certification. These characteristics suggest that perhaps many of these PD programs were working with these teachers to add on science certification to already existing certifications. The likely pressures that may lead to this approach could include NCLB requirement for ‘highly qualified’ teachers in every classroom coupled with a dearth of middle school science teachers in many places (Ingersoll, 2001).

Because the overall post-PD gain score was less than 1 point out of 25, this highlights the challenge faced by science PD providers when working with teachers having no science certification. Another possible interpretation of this overall impact may be one of questioning the sensitivity of the DTAMS measurement instrument to the PD. Because DTAMS measures broad life science knowledge, if many of the PD programs were targeting smaller subsets of content there may be weak alignment of intervention to the measure. Given the complexity of science content across the life sciences, it certainly is plausible that a reasonable approach would be to develop PD around smaller, more targeted concepts. However, if the overall goal of these PD programs were to support teachers without science certification to add on middle science certification, this highlights the challenge of PD programs to support teachers in mastering the breadth of life science content knowledge (in addition to physical science and earth/space science) that is part of the middle science curriculum standards.

Another possible contributor to the relatively low overall gains may be a question of validity of the posttest scores representing the full extent of teacher knowledge. From ongoing interactions with some of the providers of PD who use the DTAMS measure, there are a number of situations where the provider informally reported that the teachers may not have given their best effort on the posttest. It is easy to imagine scenarios where this may be a natural tendency. For example, if the assessment were given as the last task on the last day of an extended and tiring sequence of many day-long PD sessions, teachers may not have done their best in this situation. The fact that 8 out of the 19 groups had lower average post-PD scores than pre-PD scores supports this possibility.

Predictors of Science Knowledge Growth

After controlling for the pretest score, high school certified teachers outgained teachers with other grade band certifications. In general, the high school teachers may have developed a deeper knowledge of some content domain than that of elementary or middle school teachers (Ton de Jong & Ferguson-Hessler, 2010) who are usually more broadly educated to meet the requirements of their certification. A possible explanation for the stronger gains made by high school teachers may be that a deep knowledge of some field may develop a mental schema for being able to see relationships between concepts in a new domain more easily, thereby enabling them to more efficiently take advantage of the learning available in a PD program.

These data showed that the years experience of individual teachers was not a significant predictor of teacher content knowledge growth. So while Rockoff (2004) showed that more teacher experience had a positive impact on student test scores, it may be that experience strengthens the pedagogical skills of teachers to impact students (Darling-Hammond, 2000), but may not facilitate teacher growth in content knowledge. Thus, after approximately 8 years when, according to Darling-Hammond, teachers have accumulated skills and experiences necessary to effectively adapt instruction to the variety of student needs and be considered a master teacher, further development of teacher effectiveness may depend on deepening teachers' content knowledge so that they are better equipped to employ their pedagogical skills. These results suggest that this deeper content knowledge will not necessarily be facilitated at the individual level by experience teaching.

On the other hand, the overall mean years of experience of the group was predictive of higher content knowledge gains. This suggests that PD providers who are working with relatively experienced groups of teachers may be able to more effectively impact their growth in content knowledge if the PD takes advantage of that wealth of experience. It is quite reasonable to assume that PD pedagogical learning goals and instructional approaches would differ for groups of experienced teachers vs. relatively novice teachers, but this suggests that there may be interactions between group experience levels and teacher science content learning as well.

However, these results do not offer any guidance on the nature of how PD providers may take advantage if working with more experienced groups of teachers on science content learning.

Contrary to literature about best practices for PD (Garet, Porter, Desimone, Birman and Yoon, 2001) that indicates longer duration of PD positively impacts the effects that PD has on content knowledge and practice, these results do not show any differential impacts of PD duration. For this study, we have no information about what happened in the various PD programs between the pretest and posttest, which limits the ability to know why duration, surprisingly, did not have an impact. Further investigations into how to best use PD time for content knowledge growth is needed to strengthen impacts of PD initiatives that extend over longer durations of time.

These results offer a snapshot of a set of variables that do – and don't – predict life science content knowledge gains over the course of a PD program. Because the PD sites were developed and delivered independently from this study, and were geographically dispersed across the U.S., these results may be generalizable to a broad spectrum of PD offerings. These results may offer helpful indicators of where future research may be designed to strategically collect appropriate data to unpack details related to the predictive variables uncovered in this study. With appropriate lenses guided by these results, it may be possible to explain how PD programs can impact science content knowledge growth.

References

Bryk, A.S., & Raudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Clotfelter, C. T., Ladd, H. F. & Vigdor, J. L. (2007) Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26,* 673-682.

Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education, 51*, 166-173.

Garet, M.S, Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915-945.

Ingersoll, R.M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, 38(3), 499-534, DOI: 10.3102/00028312038003499

Laczko-Kerr, I., & Berliner, D. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Education Policy Analysis Archives, 10*(37).

Li, M., & Shavelson, R. J. (2001). *Examining the links between science achievement and assessment.* Paper presented at the American Educational Research Association Annual Meeting, Seattle, WA.

McCoach, D.B. & Adelson, J.L. (2010). Dealing with dependence (part I): Understanding the effects of clustered data. *The Gifted Child Quarterly, 54*, 152-155.

Monk, D. & King, J. (1994). Multi-level teacher resource effects on pupil performance in secondary mathematics and science: The role of teacher subject matter preparation. In R. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29-58). Ithaca, NY: ILR Press.

National Academies, Committee on Prospering in the Global Economy of the 21st Century. (2006). *Rising above the gathering storm: Energizing and employing America for a brighter future.* Washington, DC: National Academies Press.

National Research Council, (2011). A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press.

Raudenbush, S.W., Bryk, A.S, & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247-252.

Saderholm,J.C. & Tretter, T.R. (2008). Identification of the most critical content knowledge base for middle school science teachers. *Journal of Science Teacher Education*, *19*, 269-283.

Shulman, L. (1986). Those who understand, knowledge growth in teaching. *Educational Researcher, 15*(2), 4−14.

SPSS for Windows, Rel. 19. 2010. Chicago: SPSS Inc.

Ton de Jong, M., Ferguson-Hessler G.M. (2010). Types and qualities of knowledge. *Educational Psychologist , 31*(2), 105-113.

Tretter, T.R., Brown, S.L., Bush, W., Saderholm, J.C., & Holmes, V. (under review). *Valid and reliable science content assessments for science teachers*.

U.S. Department of Education, National Center for Education Statistics. (2004). *Qualifications of the public school teacher workforce: Prevalence of out-of-filed teaching, 1987-88 to 1999-2000.* NCES 2002-603 Revised. Washington D.C.

Urbina, S. (2004). *Essentials of Psychological Testing*. John Wiley & Sons: Hoboken, NJ.