8-2017

Modeling Inquiry-oriented Instruction of Beginning Secondary Science Teachers

Lyrica Lucas
University of Nebraska - Lincoln, lyricalucas@huskers.unl.edu

Elizabeth B. Lewis
University of Nebraska-Lincoln, elewis3@unl.edu

MODELING INQUIRY-ORIENTED INSTRUCTION OF BEGINNING SECONDARY SCIENCE TEACHERS

Lyrica Lucas and Elizabeth Lewis

University of Nebraska-Lincoln, Nebraska, USA

New national science education standards, the Next Generation Science Standards, in the United States (US) promote inquiry-based instruction through an integrated emphasis on scientific practices and disciplinary content. Thus, it is important for beginning science teachers to reach proficient implementation of reformed teaching practices by the end of their induction phase in order to become effective science teachers. Yet, extant science education research studies on development of beginning teachers' classroom practices is rare. In this study, we collected data from a longitudinal study of science teachers from two teacher preparation programs - a bachelor's program with teacher candidates who had less than a major in science and a 14-month master's degree program with candidates who had at least a major in science - in a large, Midwestern university in the US. These data were used to examine the impact of observation-level and teacher-level characteristics on the likelihood of an observed science lesson being at or below a proficient inquiry level on the Electronic Quality of Inquiry Protocol (EQUIP) instrument. Using observation-level and teacher-level data, two-level hierarchical generalized linear models were built to investigate the relationship between proficiency in inquiry-oriented instruction and the predictor variables at both levels. The parameters estimated in the best fitting model for the data indicate that observation-level variables do not significantly predict the likelihood of an observed science lesson being at or below a proficiency level on the EQUIP scale. Among the teacher-level characteristics, only the teacher preparation program was found to be statistically significant. Controlling for all other variables in the best-fitting model, the likelihood of an observed lesson being taught at the proficient inquiry level was significantly higher for teachers with a stronger science background who graduated from the master's program. Limitations of the study and future research directions are discussed.

Keywords: secondary science teachers, inquiry-oriented instruction, multilevel generalized linear models

INTRODUCTION

An inquiry-based approach to teaching and learning for science education reform has been promoted in science teacher preparation programs in response to science education policy, research literature, and standards frameworks for teaching science in the US since the early 1990s (NGSS Lead States, 2013; NRC, 1996). Supovitz, Mayer, and Kahle (2000) defined inquiry-oriented instruction as a "student-centered pedagogy that uses purposeful extended investigations set in the context of real-life problems as both a means for increasing student capacities and as a feedback loop for increasing teachers' insights into student thought processes" (p.332). Teachers need to be well-versed in inquiry-based instruction to promote student learning of science through experiential, active learning that emplys scientific practices, or thinking like a scientist (NRC, 2000). Yet, an examination of the literature on the preparation of science teachers reveals that little is known about new teachers' induction period; we need more research on how secondary science is taught by beginning science teachers (Bianchini, 2012). Unfortunately, even the existing research (e.g., Luft, Firestone, Wong, Ortega, Adams, & Bang, 2011) has failed to improve our understanding of the effectiveness of teacher preparation for the purpose of reformed-based science teaching.

This study sought to add to the knowledge base on teacher preparation and growth over time by modeling how beginning science teachers' use of inquiry-based science instruction develops throughout the first four years of in-service teaching. Using 455 coded classroom observations of 51 science teachers from two teacher education programs in a large, Midwestern university, the effects of observation-level variables and teacher-level variables on the level of reformed science instruction was examined. Since the data are hierarchically organized (i.e., class observations nested within teachers), multilevel models were used to properly account for the hierarchical (correlated) nesting of data (Hox, 2002; Raudenbush & Bryk, 2002; Snijders & Bosker, 2012).

We specifically investigated the relationship between *observation-level variables* (i.e., time, level of observed lesson (HS vs. MS), length of observed lesson (block vs. regular), and mode of observation (video vs. real-time)) and *teacher-level characteristics* (i.e., teacher's sex and education program) on the likelihood of an observed science lesson being at or below proficient use of inquiry in an observation instrument used to measure the level of inquiry-based instruction. Using observation-level (Level 1) and teacher-level (Level 2) data, hierarchical generalized linear models were built to investigate the relationship between proficiency in inquiry-based instruction and the predictor variables at both levels. The following research questions were posed in this study: (1) What is the likelihood of a science lesson being at or below proficient inquiry instruction levels taught by a typical science teacher? (2) Does the likelihood of being at or below each proficiency level vary across science teachers? (3) What is the relationship between the time of observation and the likelihood of an observed lesson being at or below a proficiency level while controlling for observation- and teacher-level characteristics? and, (4) What is the relationship between the teacher education program and the likelihood of an observed lesson being at or below a proficiency level while controlling for observation- and teacher-level characteristics?

METHOD

We collected data as part of a longitudinal study of beginning science teachers' professional practice using four cohorts of students who completed an intensive, 14-month graduate teacher certification program at a large, Midwestern university (Lewis, Musson, Pedersen, 2013). The intensive program prepares science majors and professionals to become highly qualified K-12 science teachers. This study builds upon prior exploratory work (Lewis & Musson, 2013) and the specific teacher education program details shown in Table 1 are described and presented elsewhere (Lewis, McCarty, and Musson, 2014; Lewis, Rivero, Musson, Lu, & Lucas, 2016). Science teachers who completed a bachelor's degree in secondary science education from the same university were also recruited to serve as a comparison group.

Table 1. Comparison of bachelor's and master's degree secondary science teacher preparation coursework.

Program	Bachelor's Degree	Master's Degree
Content Prerequisites	*Pre-professional Education Coursework:* Foundations of Education; Adolescent Psychology + Practicum	*Prior to Acceptance:* Undergraduate major in one area of science; some MA students have graduate-level coursework or advanced degree *MA Coursework:* Reading in the Content Area (Cohort 3-7); History and Nature of Science (Cohorts 1-2 only); Teaching ELLs in the Content Area; Intro to Educational Research; Curriculum Theory; Teacher Action Research Project
Common Coursework	Accommodating Exceptional Learners; Adolescent Development / Human Cognition; Science Teaching Methods (two classes, each with a practicum experience); Multicultural Education / Pluralistic Society	
Resulting Degree	BA Secondary Science Education, with State Content Area Teaching Endorsement	MA with Emphasis in Science Teaching, with State Content Area Teaching Endorsement

Over four years, five researchers observed and coded lessons using the Electronic Quality of Inquiry Protocol (EQUIP) instrument (Marshall, Horton, Smart, & Llewellyn, 2008) to measure the level of inquiry-based instruction in middle and high school science classrooms. By design, every teacher participant was targeted to be observed up to six times within one academic year. The validated EQUIP instrument has 19 items; each item employs a scale of 1 to 4 to describe the level of inquiry-oriented instruction in an observed science lesson. Level 1, the lowest level in the scale, corresponds to "pre-inquiry" (a teacher-centered classroom, i.e., lecture-based) and Level 4, the highest level, to "exemplary inquiry" (an open-ended and engaging student-centered classroom). For instance, in terms of instructional strategies, a teacher may be observed to "predominantly lecture to cover content" (Level 1) or "occasionally lecture but used classroom activities that promoted strong conceptual understanding" (Level 4). In this study, the EQUIP score for an observed lesson corresponds to the median score for all the 19 items in the instrument. We assume that the four-item outcomes form an underlying latent variable that is inquiry-oriented instruction behavior.

The data has a two-level structure with a set of classroom observations conducted over time that are nested within teachers. The variation of outcomes within subjects over time is at the lowest level (Level 1) and the variation of the underlying mean outcomes between subjects is at level two (Singer, 1998). Since the data gathered via the EQUIP instrument are categorical, ordinal data, multilevel generalized linear models (GLM) were used in modeling the data. The data are multinomial, violating standard linear mixed model assumptions such as normality and homogeneity of variance (Hox, 2002). In contrast with hierarchical linear models (HLM) that has continuous, approximately normally distributed outcomes, GLMs are appropriate for many kinds of non-normally distributed outcomes (e.g., binary, unordered categorical, ordered categorical, counts, censored, zero-inflated, and continuous but skewed data). The models were estimated and interpreted using SAS PROC GLIMMIX. The variables included in the models are shown in Table 2.

Table 2. Frequency count distribution of science lesson observations and teachers.

Variables Included in the Models	Science Lesson Observations (*n*=455)	Teachers (*J*=51)
Observation-level (Level 1)		
Time		
Year 1	174 (38%)	
Year 2	149 (33%)	
Year 3	100 (22%)	
Year 4	32 (7%)	
Level		
High School	350 (77%)	
Middle School	105 (23%)	
Length of Observed Lesson		
Block (90 minutes)	111 (24%)	
Regular (50 minutes)	344 (76%)	
Mode of Observation		
Video	78 (17%)	
Real-time	377 (83%)	
Teacher-level (Level 2)		
Sex		
Female		31 (61%)
Male		20 (39%)
Teacher Education Program		
Bachelor’s program		13 (25%)
Master’s program		38 (75%)

At the observation level, the variables included in the model are time of the observation in years, the level of the observed lesson (i.e., high school vs. middle school), the length of the observed lesson (i.e., block vs. regular), and mode of the observation (video vs. real-time). In this study, the time of observation refers to the post-program year of teaching when the observation was done. A lesson could be observed in the high school level (Grades 9-12) or middle school (Grades 7-8). It could be designed for a block period (90 minutes) or a regular period (50 minute). The mode of observation could be through the use of a video sent by a participating teacher or via a real-time observation, which could be done in-person or through a teleconferencing software such as Skype or FaceTime. Program and sex were included in the models as teacher-level variables. The program refers to the teacher education program completed by the teacher (i.e., bachelor’s degree vs. master’s degree in science teaching). Both teacher education programs are offered in the same college of the university and graduates from both programs were endorsed to teach science.

The outcome variables from the EQUIP scale are polytomous, ordinal-type. In SAS PROC GLIMMIX, a multinomial distribution and a cumulative logit link were used to allow for the computation of the cumulative odds for each EQUIP category (i.e., 1=Pre-inquiry, 2=Developing inquiry, 3=Proficient inquiry, 4=Exemplary inquiry), or the odds that an outcome would be at most, in that category (O'Connell, Goldstein, Rogers, & Peng, 2008). In this study, we were interested in the probability of being at or below a proficiency level defined in the EQUIP scale and in the influence of observation (Level-1) and teacher (Level-2) characteristics on this probability for each category. The conceptualization of the models in a generalized linear framework is represented by a set of equations in the next section.

RESULTS

Three proportional odds logistic models were estimated with the EQUIP data. In all of the models, the default convergence criterion (GCONV=1E-8) was satisfied. Table 3 shows the distribution of EQUIP scores for all observations from the response profile generated by SAS PROC GLIMMIX. The scores were distributed in the first three categories of the EQUIP scale, but not the fourth.

Table 3. Distribution of EQUIP Scores of Observed Science Lessons (*n*=455)

EQUIP category	Frequency (*n* (%))
1 – Pre-inquiry	85 (19%)
2 – Developing inquiry	291 (64%)
3 – Proficient inquiry	79 (17%)
4 – Exemplary inquiry	0 (0%)

The ordinal empty means, random intercept only model, is represented by two logit-based model equations (1). When dealing with polytomous outcomes, multiple logits are simultaneously estimated (M-1 logits, where M=the number of outcome categories). For the case of three outcomes as shown in Table 3, two logits are simultaneously estimated by the model.

$$\eta_{1ij} = \log\left(\frac{P(R_{ij} \le 1)}{1 - P(R_{ij} \le 1)}\right) = \gamma_{001} + U_{0j};\eta_{2ij} = \log\left(\frac{P(R_{ij} \le 2)}{1 - P(R_{ij} \le 2)}\right) = \gamma_{002} + U_{0j} \quad (1)$$

The two intercepts in the model represent the log odds of an observation in a typical teacher being at or below the first two levels of inquiry-based instruction (i.e., pre-inquiry and developing inquiry) in the EQUIP scale. These log odds can be used to calculate the probabilities of being at or below each proficiency level by using the following equation (2) wherein φ_{ij} stands for cumulative probability.

$$\phi_{ij} = \frac{e^{\eta_{ij}}}{1 + e^{\eta_{ij}}} \quad (2)$$

Parameter estimates for Model 1 are shown in Table 4. Using the model equations, the log odds of being at the pre-inquiry level for an observed science lesson in a typical teacher is -1.58, resulting in a probability of 0.17. Similarly, the log odds of being at or below the developing inquiry level is 1.98, resulting in a cumulative probability of 0.88. Finally, the cumulative probability of being at or below the proficient inquiry level adds to 1. To calculate the actual probabilities of being at each level, cumulative probabilities of adjacent categories are subtracted from one another. As a result, the predicted probability of an observed lesson being at the pre-inquiry level for a typical teacher is 0.17, 0.71 at the developing inquiry level, and 0.12 at the proficient inquiry level.

Table 4. Estimates for two-level generalized linear models of inquiry-based instruction.

	Model 1 (Unconditional model)	Model 2 (Model 1 + Observation-level fixed effects)	Model 3[a] (Model 2 + Teacher-level fixed effects)
Fixed Effects			
Intercept 1 (Pre-Inquiry)	-1.58* (0.19)	-1.34* (0.41)	-0.21 (0.52)
Intercept 2 (Developing Inquiry)	1.98* (0.21)	2.24* (0.45)	3.37* (0.56)
Time (in years)		-0.18 (0.12)	-0.20 (0.11)
Level (HS=1, MS=0)		0.11 (0.39)	0.32 (0.35)
Length of Observed Lesson (Block=1, Regular=0)		-0.48 (0.35)	-0.42 (0.32)
Mode of Observation (Video=1, Real-time=0)		0.15 (0.36)	0.14 (0.33)
Sex (Female=1, Male=0)			-0.12 (0.30)
Teacher Education Program (MAst = 1, BSEd = 0)			-1.51* (0.38)
Error Variance			
Intercept	0.92* (0.34)	0.89* (0.36)	0.51* (0.24)
Model Fit			
-2 Log Likelihood	787.04	780.92	767.00**

Note:*p<.05; **=likelihood ratio test significant; Values based on SAS PROC GLIMMIX. Entries show parameter estimates with standard errors in parentheses; Estimation Method=Laplace.

[a]Best fitting model

The empty, unconditional model with no predictors provides an overall estimate of the intraclass correlation (i.e. ICC = $\tau_{00} / (\tau_{00} + 3.29) = 0.92 / (0.92 + 3.29) = 0.22$). In multilevel GLMs, there is assumed to be no error at Level-1, therefore, a modification was needed to calculate the ICC. This modification assumes that the outcome originates from an unknown latent continuous variable with a Level-1 residual that follows a logistic distribution with a mean of 0 and a variance of 3.29 (Snijders & Bosker, 2012). Therefore, 3.29 was used as the Level-1 error variance in calculating the ICC. The ICC indicates that approximately 22% of the variability of being at or below a proficiency level in the EQUIP scale is accounted for by the teachers in the study, leaving 78% of the variability to be accounted for by the observations or other unknown factors. However, it should be noted that the ICC is somewhat problematic to interpret due to non-constant residual variance. Model 1 also indicates that there is a statistically significant amount of variability in the log odds of being at or below a proficiency level between teachers [$\tau_{00} = 0.92$; $z(50) = 2.75$, $p<.05$].

Model 2 includes the fixed effect estimates for observation-level variables (i.e., time, level, the length of the observed lesson, and mode of observation). The fixed effect estimates illustrate the relationship between an observation-level characteristic and the log odds of being at or below a proficiency level in the EQUIP scale. The value of each fixed effect estimate remains constant across logits although there are two estimates for the intercepts. This means that the fixed effects are assumed to be the same for each cumulative odds ratio. Model 3 was similar to Model 2 with the addition of teacher-level fixed effects. Table 4 presents a summary of the results and estimates for all three models considered in the model-building process as well as model fit information.

We compared the three models in terms of fit in order to decide the best fitting model for these data. Based on the changes in the -2 Log Likelihood between nested models, Model 3 is the best fitting model for these data; it fits significantly better than Model 2 ($\chi^2(2) = 13.92$, $p<.001$) and also better than Model 1 ($\chi^2(6) = 6.12$, $p<.05$). The addition of teacher-level variables improved model fit.

DISCUSSION

To address our research questions, the parameter estimates from the best-fitting model (Model 3) were used. The first research question requires finding the likelihood of being at or below each proficiency level in inquiry-based instruction for an observed lesson taught by a typical science teacher. Using Model 3, we found that the probability of an observed lesson being at the pre-inquiry level for a typical teacher was 0.45; 0.52 at the developing inquiry level, and 0.03 at the proficient inquiry level. These predicted probabilities are interpreted based upon all variables in the model being equal to zero. As a follow-up, in our second research question, we were interested to know if the likelihood of being at or below each proficiency level varied across science teachers. Looking at the the error variance estimate for the random intercept, Model 3 indicates that there is a statistically significant amount of variability in the log odds of being at or below a proficiency level between teachers [$\tau_{00} = 0.51$; $z(48) = 2.08$, $p<.05$]. The probability of being at or below a proficiency level varies considerably across teachers.

For our third research question, we found that there was no statistically significant relationship between the time of observation and the likelihood of an observed lesson being at or below a proficiency level while controlling for observation- and teacher-level characteristics. The final, fourth research question refers to the relationship between teachers' education program and the likelihood of an observed lesson being at or below a proficiency level while controlling for observation- and teacher-level characteristics. To answer this question, we used the parameter estimate for teacher education program (b=-1.51, *p*<.05), which indicated a negative, statistically significant relationship between teachers' education program and the likelihood of an observed lesson being at or below a proficiency level. Specifically, as we move from a lesson taught by a science teacher with a bachelor's degree in secondary science education to a lesson taught by a teacher with a master's degree with an undergraduate degree in an area of science, the likelihood of an observed lesson being at the proficient level increases.

To make a more meaningful interpretation, we calculated the corresponding predicted probabilities for observed lessons taught by teachers in different preparation programs and controled for other observation- and teacher-level characteristics. Using Model 3 parameter estimates in Table 3, the log odds of an observed lesson taught by a teacher graduate of the master's degree program (program=1) being at or below the pre-inquiry level is 0.18, resulting in a probability of 0.15. Similarly, the log odds of being at or below the developing inquiry level is 6.43, resulting in a cumulative probability of 0.87. From these values, we found that the probability of an observed lesson being at the pre-inquiry level for a graduate of the master's program is 0.15; 0.71 at the developing inquiry level, and 0.13 at the proficient inquiry level. These predicted probabilities are interpreted for the case of program=1 and all other variables in the model being equal to zero. This means that the predicted probability of an observed lesson (at the beginning of Year 1, taught in middle school on a regular schedule by a male teacher with a master's degree, and observed in-person) to be at the pre-inquiry level is 0.15. For a teacher with a bachelor's degree, the predicted probability of an observed lesson being at the lowest proficiency level is 0.45. Thus, controlling for all other observation- and teacher-level characteristics, an observed lesson taught by a graduate of the bachelor's program has a higher probability of being at the lowest proficiency level in the EQUIP scale compared to a lesson taught by a graduate of the master's program. In other words, the master's level teachers enacted reformed-based science teaching more frequently.

Figure 1 compares teachers by teacher education program in terms of the change in probability of EQUIP score outcomes across years of teaching. For both groups, the likelihood of an observed science lesson to be teacher-centered or being in the lowest level of the EQUIP scale decreases as the teachers gain more experience. However, teachers from the master's program start at a higher level; they are more likely to create and implement more inquiry-based lessons and continue to improve as they gain teaching experience. Thus, teachers with a master's degree in science teaching appear to show accelerated growth in the in the used of inquiry-based teaching practices compared to teachers with only a bachelor's degree in secondary education with science endorsement.

These findings imply that differences in teacher education affect the long-term development of inquiry practices in the first four years of teaching. However, there are several limitations that need to be taken into account when interpreting these results. Adding new observation data from the fifth year of the longitudinal study could increase the precision of the models. It could also allow us to better understand and describe the growth of beginning teachers since the first 5 years are commonly considered to encompass the notion of beginning teaching (Loughran, 2014). The findings regarding the particular ramifications of the teacher education programs are also context-dependent; the results may only be transferable to similar program designs.

Also, several background variables such as age, science credits hours, and work experience could be contributing to differences in the performance of inquiry-oriented science teaching. Finally, there were several factors that were not considered in building the models that may have a significant impact on the enactment of inquiry-based instruction such as size and diversity (i.e., racial diversity and socioeconomic status) of the students in the observed lessons, the amount of in-service teacher professional development activities, subject matter knowledge, and teacher beliefs and self-efficacy in teaching.

Although it appears that lessons taught by graduates of the bachelor's program have a higher likelihood of being in the lowest level of the EQUIP scale corresponding to a more teacher-centered approach, it should be noted that the features of the two teacher education programs were not systematically investigated.

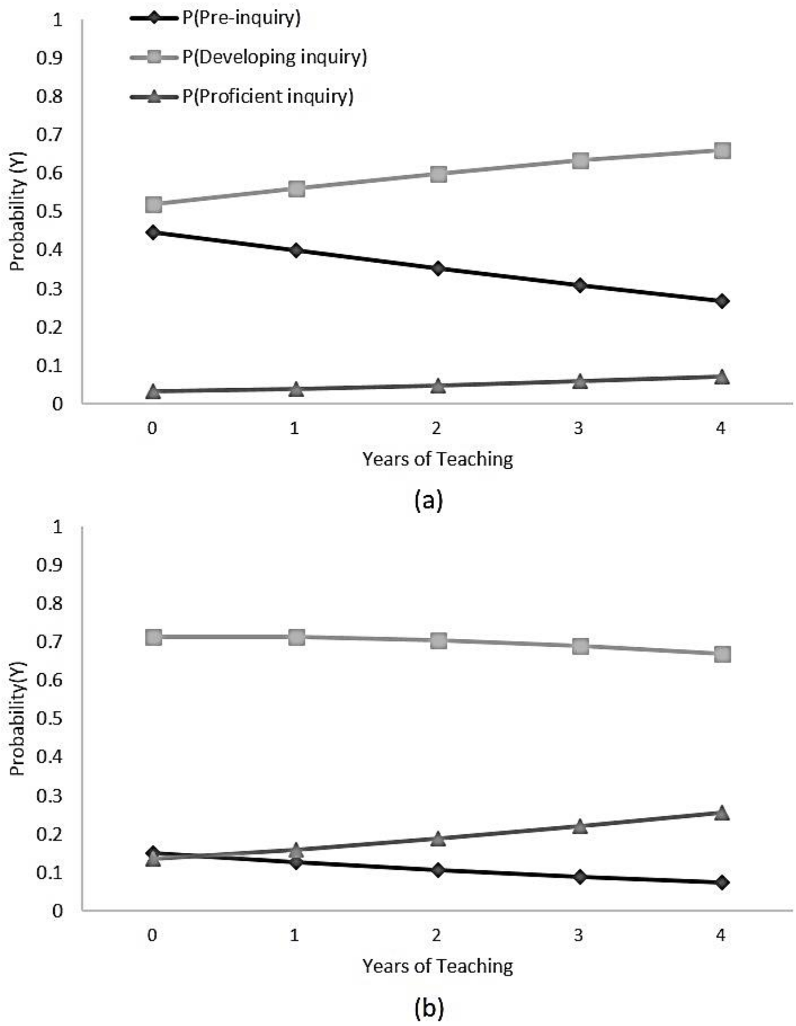

Figure 1. Change in probability across years of teaching. (a) Teacher has a bachelor's degree in secondary education (science endorsement). (b) Teacher has a master's degree in science teaching.

Teachers from the master's program could have a stronger science content knowledge due to their completed science degree prior to taking a graduate-level master's program on teaching. Also, they were older and may have worked as a science professional which may have led to the development and mastery of science process skills that are important in the teaching of science as well as their understanding of the nature of science. More studies that explore how the master's level program accelerates new science teachers' growth would be productive (Lewis, Rivero, Musson, Lu, & Lucas, 2015).

In this exploratory study, the variables were entered in aggregate into the models. Thus, the modeling process did not consider specific predictors alone and as a result there may be possible interactions within the models. More complex hierarchical models that use longitudinal data on teachers, schools, and school districts are needed to capture the intricacies of teacher change.

CONCLUSIONS

This study examined the effect of observation-level variables (i.e., time, lesson (HS vs. MS), length of observed lesson (block vs. regular) and mode of observation (video vs real-time)) and teacher-level characteristics (i.e., teacher's sex and education program) on the likelihood of an observed science lesson being at or below a proficiency level in the EQUIP scale. Using observation-level (Level 1) and teacher-level (Level 2) data, we built two-level hierarchical generalized models to investigate the relationship between proficiency in inquiry-based instruction and the predictor variables at both levels. The parameters estimated in the best fitting model for the data indicate that observation-level variables do not significantly impact the likelihood of an observed science lesson being at or below a proficiency level in the EQUIP scale. Among the teacher-level characteristics, only the teacher preparation program was found to be statistically significant. Controlling for all other variables in the full model, the likelihood of an observed lesson being at the lowest proficiency level is significantly lower for teachers who graduated from the master's program. These findings imply that differences in teacher education preparation determine the future development of reformed science teaching. Future research that identifies aspects of instruction (e.g., discourse, assessment, instructional strategies, curriculum design) that display the least growth during the induction period would be useful in designing and improving programs for teacher professional development. Finally, it is important to build other models to explore which variables account for the unexplained variance in the enacted teaching practices of beginning secondary science teachers.

ACKNOWLEDGEMENTS

This material is based upon work supported by two National Science Foundation (NSF) Noyce grants (DUE-1035358 and DUE-1540797). Ideas presented in this study are those of the authors and do not necessarily reflect the views of personnel affiliated with the NSF. We would also like to thank other members of our research group, Ana Rivero, Amy Tankersley and Aaron Musson who collected and coded data for this study.

REFERENCES

Bianchini, J. (2012). Teaching while still learning to teach: Beginning science teachers' views, experiences, and classroom practices. In B.J. Fraser, K. Tobin, & C.J. McRobbie (Eds.), *Second international handbook of science education* (pp. 389-399). The Netherlands: Springer.

Hox, J.J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Lewis, E. B., Rivero, A. Musson, A., Lu, Jia, & Lucas, L. (2016, April). *Building exemplary teaching practices: Following the paths of new science teachers*. Paper presented at the 2016 Annual

International Conference of the National Association for Research in Science Teaching, Baltimore, Maryland.

Lewis, E.B., McCarty, R., & Musson, A. (2014, January). *Evaluating a graduate-level science teacher education program.* Paper presented at the annual meeting of the Association for Science Teacher Education, San Antonio, Texas.

Lewis, E.B., & Musson, A. (2013, September). *Science teacher knowledge, beliefs, and enacted curricular practices from pre-teacher education program to first-year teaching.* Paper presented at the biannual conference of the European Science Education Research Association, University of Cyprus, Nicosia, Cyprus.

Lewis, E.B., Musson, A.A., & Pedersen, J.E. (2013, January). *Growing science teachers: From content to pedagogy.* Paper presented at the annual meeting of the Association for Science Teacher Education, Charleston, South Carolina.

Loughran, J. J. (2014). Developing understandings of practice. In S. K. Abell & N. G. Norman. *Handbook of Research on Science Education volume II* (pp.811-829). New York, NY: Routledge.

Luft, J. A., Firestone, J. B., Wong, S. S., Ortega, I., Adams, K., & Bang, E. (2011). Beginning secondary science teacher induction: A two-year mixed methods study. *Journal of Research in Science Teach*ing, *48*(10), 1199–1224. doi:10.1002/tea.20444

Marshall, J.C., Horton, B., Smart, J., & Llewellyn, D. (2008). *EQUIP: Electronic quality of inquiry protocol.* Retrieved from Clemson University's Inquiry in Motion Institute, www.clemson.edu/iim.

National Research Council (1996). *National science education standards: Observe, interact, change, learn.* Washington, DC: National Academy Press.

National Research Council (2000). *Inquiry and the national science education standards: A guide for teaching and learning.* Washington, DC: The National Academies Press.

NGSS Lead States (2013). *Next generation science standards: For states, By states*. Achieve, Inc. on behalf of the twenty-six states and partners that collaborated on the NGSS.

O'Connell, A.A., Goldstein, J., Rogers, H. J., & Peng, C.-Y. J. (2008). Multilevel logistic models for dichotomous and ordinal data. In A. A. O'Connell & B. McCoach (Eds.), *Multilevel analysis of educational data* (pp. 199-242). Charlotte, NC: Information Age Publishing Inc.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23*(4), 323-355.

Snijders, T.A.B. & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed). Thousand Oaks, CA: Sage.

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy*, *14*, 331-356.

Treagust, D. (2010). General instructional methods and strategies. In S. K. Abell & N. G. Norman. *Handbook of Research on Science Education* (pp.373-391). New York, NY: Routledge.